idiap
RESEARCH INSTITUTE

SCIENTIFIC REPORT

APRIL 2019

Contents

List of Figures

1 Idiap Structure and Research Areas

1.1 Research Areas

Idiap's research activities span five broad areas, all directly related to our central themes and competencies of **Signal Processing**, **Machine Learning**, and **"Social Intelligence for Society"**: perceptual and cognitive systems, human and social behavior, multimodal/multimedia information processing and presentation interfaces, biometrics security and privacy, and machine learning, which are summarized, with a set of descriptive keywords in Figure 1 below.

In addition to these transverse research themes, new activities have been recently developed, towards bio-medical applications (through the new group in Computational Bioimaging) and environmental modeling (through the new group in Uncertainty Quantification and Optimal Design).

Idiap maintains a policy of continuous growth and diversification in new and complementary research directions, while building upon its long-standing institutional theme of "Human and Media Computing".

Research Areas	***Keywords***
Perceptual and cognitive systems	*Speech and audio processing, computer vision, document processing, robotics, natural language processing, machine translation, computational cognitive science*
Human and social behavior	*Social media, social sensing, verbal and nonverbal communication analysis, smartphone sensing, computational social science*
Information and presentation interfaces	*Multimedia information systems, user interfaces, personalization, system evaluation, mobile HCI using big data, data driven services*
Biometrics Security and Privacy	*Face recognition, speaker recognition, vein recognition, multimodal fusion, soft-biometrics, remore photoplethysmography, presentation attack detection (anti-spoofing), template protection, privacy preservation, mobile and wearable biometrics*
Machine learning	*Statistical and neural network based ML, computational efficiency, online learning, multiple sensor processing, big data*

Figure 1: Overview of Idiap research areas. These research areas are currently covered by eleven research groups, listed in Section 1.3 below, and are designed to be as pro-active as possible in covering the applications areas presented in Figure 2.

1.2 Application Areas

Idiap's application areas are listed in Figure 2, together with keywords and pointers to relevant projects. Idiap contributes to a broad range of applications areas, which reflect concrete, ongoing projects.

We would like to draw attention to the difference between our *research* and our *application* areas. While Idiap's research areas (Figure 1) are quite generic and have a strong transverse nature, the application areas are more targeted and take place along well-defined application axes. Although we do not necessarily address at Idiap all aspects within a given application area, we are often in the best position to provide the necessary enabling Information and Communication Technology components.

Application Areas	***Keywords***
Human-human & human-machine interaction	*Voice and gesture controlled devices and robots, hand-free control, spoken language systems, translation systems, social robotics, user profiling*
Exploitation of rich multimedia archives (audio, video, text)	*Hyper-events, semantic indexing, keyword spotting, object detection and recognition, image bank browsing, audio-video content filtering (summarization and recommendation), broadcast data analysis, scanned document analysis, analysis of cultural heritage media*
Collaborative and creative systems	*Remote meeting assistance, smart meeting room, video-conferencing, multimedia indexing and access, cross-lingual collaboration, interaction analysis, dynamics of negotiation*
Healthcare and bio-engineering	*Smart management of patient data, prosthesis (hearing aids, artificial body parts, voice banking), bio-systems modeling, interfaces for impaired users (speech and other inputs), behavior-based health diagnosis, bio-medical document and data processing, smartphone platforms for health, bio-signals and imaging, microscopy*
Entertainment	*Multilingual gaming, remote family games, togetherness*
Mobile computing	*Signal processing for mobile platforms, mobile social networks, participatory sensing*
Security and risk management	*Biometric security, access control, mobile biometry, multi-sensor fusion, speaker identification, video monitoring of areas/activities, natural risk modeling, intrusion detection, crowd management*
Home automation (domotics)	*Multi-sensor activity analysis, adaptation to users' behavior, efficient use of energy, home safety and security*
Energy	*Energy grids, multiple sensor and smart meter networks, large-scale sensor data integration, modeling of behaviors to anticipate demand, safer, cheaper, and cleaner energy production*
Smart cities	*Ecology, environment management, reduction in pollution, traffic and noise, better use of roads*

Figure 2: Idiap application areas with several examples for each of them.

1.3 Internal Structure

1.3.1 Overview

The main research themes, briefly presented in Section 1.1 above, are currently covered by eleven research groups, presented in the following section. The resulting structure is presented in Figure 3 along with the overall structure of Idiap. The leaders of the groups, who are all PIs of research projects, are explicitly indicated.

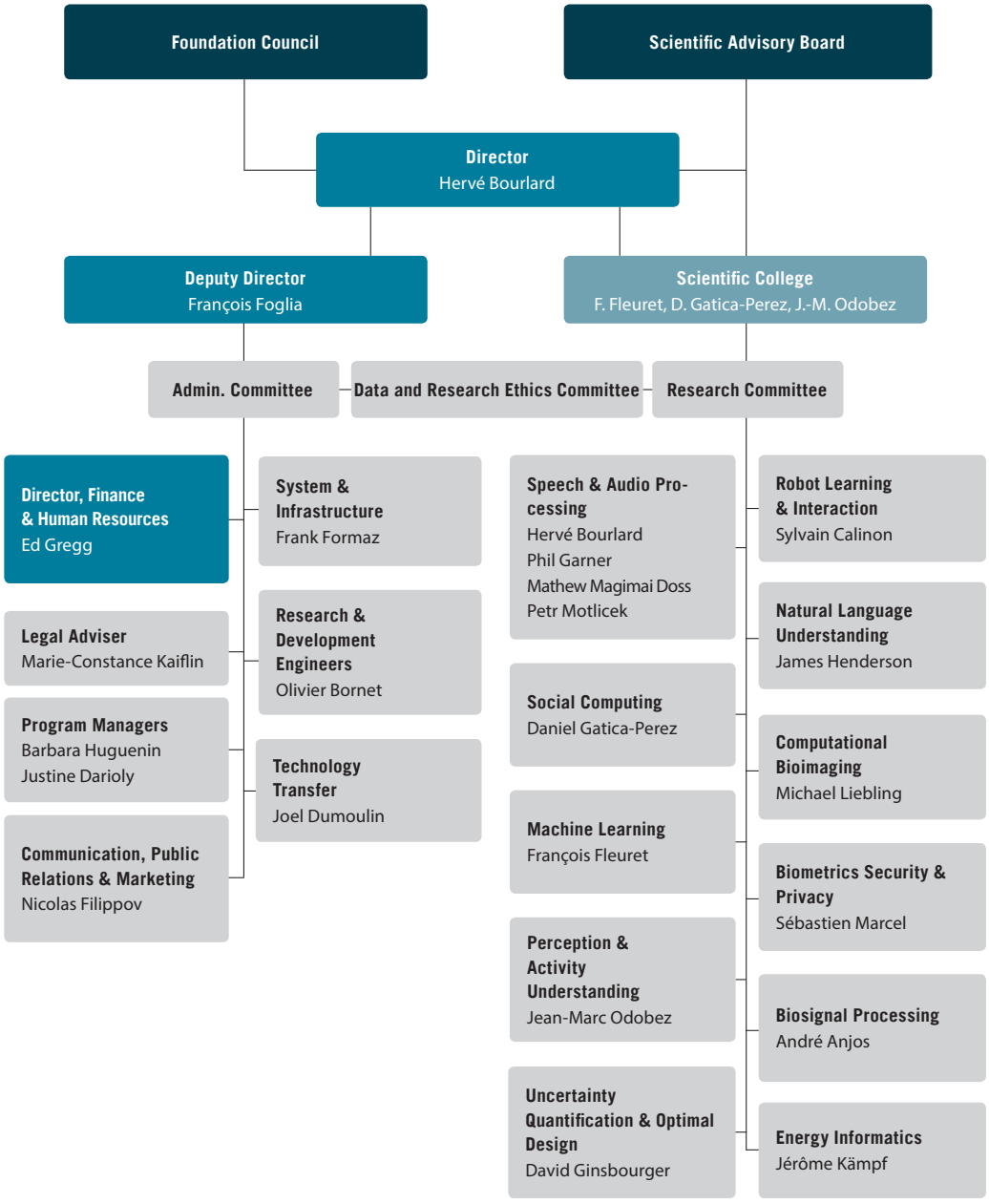

Figure 3: Idiap management and operational structure, including the main research and administrative responsibilities. For the research block, all names in white are key researchers, with clear supervision roles, and are all project PIs.

1.3.2 Research Groups

In 2018, the research areas presented in Figure 1 were covered by eleven research groups. The activity of each research group can be summarized as follows:

1. **Speech and Audio Processing (Prof. Hervé Bourlard, Dr. Philip N. Garner, Dr. Mathew Magimai-Doss, & Dr. Petr Motlicek)**

   Speech processing has been one of the mainstays of Idiap's research portfolio for many years. Today it is still the largest group within the institute, and Idiap continues to be recognized as a leading proponent in the field. The expertise of the group encompasses multilingual speech recognition, multilingual text-to-speech conversion, 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.

   See Section 2.1 for the 2018 progress report.

2. **Machine Learning (Dr. François Fleuret)**

   The goal of the Machine Learning group is the development of new statistical learning techniques with a particular interest in their computational properties. Our application domain encompasses any processing of real-world signals, such as image understanding, detection of persons and biological structures, signal synthesis, or decision from low-level industrial sensors.

   See Section 2.2 for the 2018 progress report.

3. **Social Computing (Prof. Daniel Gatica-Perez)**

   Social computing is an interdisciplinary domain that integrates theories and models from ubiquitous computing, multimedia, machine learning, and social sciences, to sense, analyze, and interpret human and social behavior in everyday life, and to create systems that support social interaction. Current lines of research include social media analytics and mobile crowdsourcing for cities and health, analysis of ubiquitous social interaction, and analysis of cultural heritage media.

   See Section 2.3 for the 2018 progress report.

4. **Perception and Activity Understanding (Dr. Jean-Marc Odobez)**

   The group investigates models from machine learning, computer vision, multimodal signal processing, or social sciences, to address the understanding of activities from real-world signals, with an emphasis on those related to humans. Detection, tracking, pose estimation, recognition and analysis of non-verbal behaviors or the temporal interpretation of all this information in forms of gestures, activities behavior or social relationships are examples of studied tasks. Application domains encompasses surveillance, traffic and human behavior analysis, human-robot interactions, and multimedia content analysis.

   See Section 2.4 for the 2018 progress report.

5. **Robot Learning and Interaction (Dr. Sylvain Calinon)**

   This group focuses on human-centric 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 the development of intuitive active learning interfaces to acquire meaningful demonstrations, the development of models that can exploit the structure and geometry of the acquired data in an efficient way, and the development of adaptive control techniques that can exploit the learned task variations and coordination patterns.

   See Section 2.5 for the 2018 progress report.

6. **Uncertainty Quantification and Optimal Design (Dr. David Ginsbourger)**

   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.

   See Section 2.6 for the 2018 progress report.

7. **Computational Bioimaging (Dr. Michael Liebling)**

   This group focuses on research in computational imaging and 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 for the observation and quantification of complex and live biological systems.

   See Section 2.7 for the 2018 progress report.

8. **Biometrics Security and Privacy (Dr. Sébastien Marcel)**

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

   See Section 2.8 for the 2018 progress report.

9. **Natural Language Understanding (Dr. James Henderson)**

   The Natural Language Understanding group (created September 2017) studies deep learning for natural language processing tasks, focusing on models with learned representations of the meaning of text. It continues the NLP group's work on neural machine translation and information retrieval, and extends to neural network structured prediction and representation learning for modelling the syntax and semantics of text and speech, including modelling abstraction (textual entailment) and summarisation.

   See Section 2.9 for the 2018 progress report.

10. **Biosignal Processing (Dr. André Anjos)**

    This group focuses on biomedical-related areas such as the analysis of e-Health records, human-signal and imaging sensing for healthcare and similar applications. Current trends in the field show refreshed interest on the use of machine learning techniques, complementing basic signal and sequence processing, all of which are key domains of research at Idiap. It leverages on Idiap's expertise on human subject handling, data acquisition, open science and data processing.

    See Section 2.10 for the 2018 progress report.

11. **Energy Informatics (Dr. Jérôme Kämpf)**

    The Energy Informatics Group (created September 2018) studies the exploitation of state-of-the-art Information and Communication Technologies to tackle global warming and climate change challenges, to increase integration of renewable and distributed energy sources by making energy systems smarter, and to increase energy efficiency beyond what improvements at component level can achieve. In that vein, the group researches into ways of simulating energy transition pathways with intelligent control and adjustment mechanisms of evolving buildings with retrofitting and use, renewable energy production and energy storage in a changing climate.

    See Section 2.11 for the 2018 progress report.

1.3.3 Administration and Services

At the administration and services level, there are seven groups: finances and human resources; communication, public relations and marketing; system and infrastructure; technology transfer; development engineers; program managers; and legal advisor. The resulting admin organization is presented in Figure 3, page 3. The main responsibilities of the seven admin groups can be summarized as follows:

1. **Finance and Human Resources (Ed Gregg)**: The Human Resources Department (HR) is integrated within the financial and accounting activities of Idiap and has taken on a greater importance in the past year. With employees from over 30 different countries, the finances and HR department is continually growing to meet the needs of each employee.
2. **Communication, Public Relations & Marketing (Nicolas Filippov)**: The mission of the communication, public relations and marketing department is to use all forms of media and communication to build, maintain, manage the reputation of the Institute, and to promote the Idiap services available for external institutions, such as EU project management, submission proposal tools, etc.
3. **System and Infrastructure (Frank Formaz)**:The main mission of the system and infrastructure group is to provide an optimal and efficient work environment for the Idiap collaborators. The tasks can be split into three main activities covering (1) centralized IT services for the whole Institute (network, storage, servers, workstations, high performance computing, identity management, data distribution), (2) support for collaborators (helpdesk, project specific tasks, web presence), and (3) Infrastructure (building, offices, equipments, central purchasing office).
4. **Technology Transfer and Industrial Relations (Dr Joël Dumoulin)**: Technology transfer is one of the Idiap Research Institute's three core missions. One of the fundamental challenges is to facilitate the interface between the knowledge and the skills of the researcher and the needs of the industrial partners. Idiap resolves this by providing a dedicated multi-disciplinary team of developers and programmers which transfers pieces of software, algorithms, knowledge and expertise. This transfer of technology is usually done by granting rights on the commercial exploitation of this technology (through license). Maintaining the Technology Portfolio and helping the researchers in the process of patenting are among the other activities of the Technology Transfer Officer (TTO). Also in charge of the industrial relations, the TTO is the contact person for companies interested in collaborating with Idiap and presents Idiap and Idiap's activities to them. His role is to identify collaboration opportunities with potentially new partners, initiate the discussion and involve the related researcher at the right time in order to improve the efficiency of the process of preparing industrial projects.
5. **Program Managers (Justine Darioly & Barbara Huguenin)**: The work of the program management team is divided into two types of activities. The first is the provision of services to researchers within the framework of European and Swiss projects. The second category includes activities ranging from event organization to database management, which are not directly linked to the management of research projects but facilitate the work of Idiap researchers.
6. **Development Engineers (Olivier Bornet)**: The mission of the development team is to provide support to Idiap researchers in the software development tasks. This is done in three areas. The first is to help on Idiap research by building prototypes, implement algorithms, design and run experiments, and manage legacy code. The second area is for all the technology transfer tasks. In the third activity, development engineers give daily support to Idiap researchers (software disclosures, showroom and internal demonstrators, development tools).
7. **Legal Adviser (Marie-Constance Kaiflin)**: The main missions of the legal adviser are to write, analyse, and negotiate project contracts (research, consortium, consultant agreements, NDA, Memorandum of Understanding, etc) or technology transfer contracts (patents, knowhow, licenses) with industries, universities or research institutions. The legal adviser deals also with all the legal aspects related to human resources (work contracts, staff regulations, rules) and data protection (ethics, databases collection and distribution).

2 Research Groups

2.1 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 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, automatic speech intelligibility evaluation, or speech processing for classification of motor speech disorders. The expertise and activities of the group encompass statistical automatic speech recognition (based on Hidden Markov Models-HMM, or hybrid systems exploiting Deep Neural Networks-DNN and new deep learning architectures), text-to-speech (TTS), speaker recognition (with extensions towards text-dependent and forensics scenarios) 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 2018 was composed of 1 head of group, 3 principal investigators, 1 research associate, 11 postdocs, 13 PhD students, and 9 interns.

Key scientific outputs: Idiap has always significantly contributed 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 new architectures of deep learning, studied in particular through *PyTorch* and other open source frameworks. Advances in Automatic Speech Recognition (ASR) are usually researched through *Kaldi* toolkit, now used by most of the international speech community.

In 2018, several key research contributions were achieved by the group, including: (1) multilingual speech recognition, especially in cross-lingual adaptation, and speech recognition in low-resourced language conditions, (2) speaker recognition, through both text-independent and particularly text-dependent speaker verification scenarios and information fusion for large-scale speaker identification, (3) large scale media processing, including multilingual broadcast news recognition, and spoken query for spoken term detection, and (4) detection impairments in speech signal to uncover motor speech disorders.

The group has 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. The group is also involved in deployment of speech and speaker recognition algorithms for industrial applications (e.g. keyword spotting detection and speaker identification for embedded devices, automatic speech recognition jointly trained with natural language modeling task, etc.).

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

Automatic speech recognition

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

- **Robust parametrisation and acoustic modelling**: 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**: The EC SUMMA project (Scalable Understanding of Multilingual Media), as illustrated in Figure 4, aims at integrating stream-based media processing tools (including speech recognition and machine translation) with deep language understanding capabilities (including named entity relation extraction and semantic parsing), for open-source applications and implemented in use cases at the BBC and Deutsche Welle. From 2017, Idiap collaborates on the US IARPA SARAL project (Summarization and domain-Adaptive Retrieval of Information Across Languages). The project aims at developing cross-lingual retrieval and summarization techniques that will work for any language in the world, given minimal resources to work with. In those contexts, 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 Deep Neural Network (DNN) acoustic model is boot-strapped using well-resourced data and adapted to the target language.

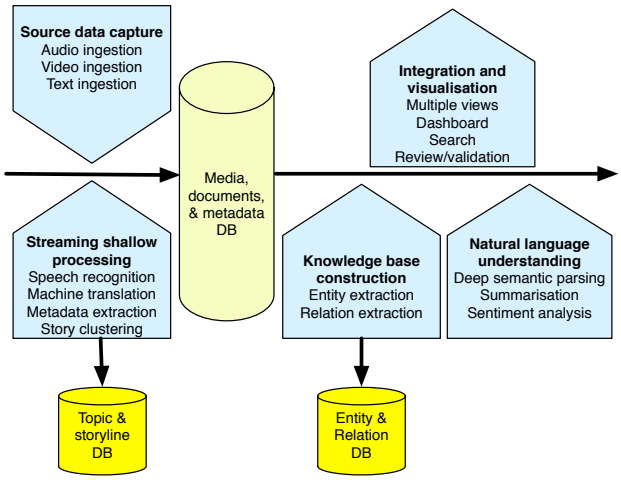

Figure 4: An overview of the EC H2020 SUMMA project, a current focus for multilingual speech recognition at Idiap.

- **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. 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. In 2018, new Innosuise project SM2 ("Extracting semantic meaning from spoken documents") has started, fostering collaboration of Idiap with *recapp IT AG* and *Crealogix AG*, with aim to develop a customisable technology for "semantic keyword and concept detection" allowing bank institutes to meet MIFID (business and organisational requirements for investment firms).

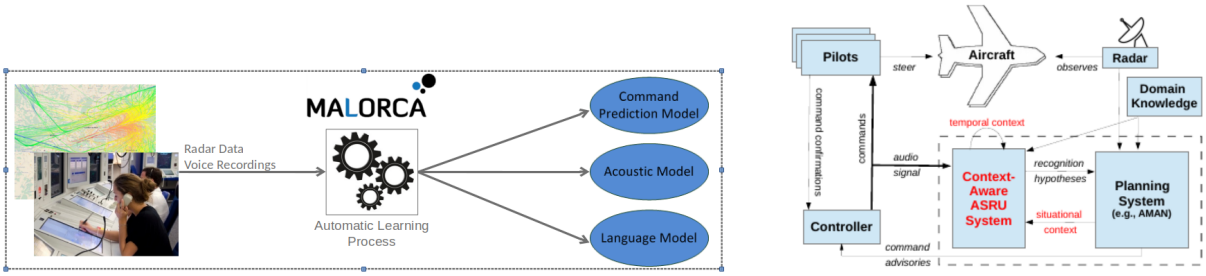

Figure 5: Left: Overview of the EC MALORCA project; semi-supervised adaptation of three models on the left side. Right: graphical representation of building blocks.

- **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[1]. 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.
- **Lexicon development**: Over the past seven 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 grapheme-to-phoneme relationship through the acoustic speech signal (funded through the SNSF-project FlexASR and HASLER Foundation project AddG2SU).
- **Semi-supervised learning of acoustic and language model parameters**: This problem has been conducted at Idiap over several past years, as it aims to develop a technology to (1) automatically select speech data which is acoustically highly informative, or (2) to automatically assess recognition output with high confidence. This type of technology is essential to allow semi-supervised learning of speech recognition models using unlabelled data. One of the specific architecture has been developed within the EC H2020 MALORCA project[2], combining both acoustic (i.e. speech) and situational (i.e. radar) information to build automatic speech recognition system for air-traffic controllers (i.e. controller-pilot communication). The MALORCA project, graphically represented in Figure 5 further focused on technologies to automatically adapt the developed system for new airports, while minimising human effort to manually transcribe speech data. The project has ended in spring 2018, with high impact recognized by SESAR EC H2020 programme.
- **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. More recently, this is influenced by the encoder-decoder approaches used in machine translation.
- **Failure analysis of ASR and HSR decoding channels, uncertainty analysis**: This is a novel strategy further developed in 2018 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

[1] https://www.idiap.ch/scientific-research/projects/multiveo
[2] http://www.malorca-project.de

characterization of the sub-word 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. In 2018, interesting research was done, exploiting uncertainty to estimate word-error rates without the need for explicit transcriptions. A novel DNN framework was proposed which uses 'dropout' at the test time to model uncertainty in prediction word recognition hypotheses.

- **Effective processing of speech using embedded devices**: ASR or other tasks such as keyword spotting and speaker identification are among interest of industrial partners to operate on voice-enabled devices. The devices are usually special-purpose computing systems with limited computing resources and strict power requirements, but have extensive applications in consumer, industrial or healthcare markets. From 2018, Idiap collaborates with *Logitech* through a CTI project to develop a speech hybrid analytics platform for dedicated consumer and enterprise devices. Idiap also runs a self-funded project with *CSEM* to integrate speech processing algorithms on specialised embedded devices for healthcare market.

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[3] (Spoken Interaction With Interpretation in Switzerland), D-BOX[4] and SP2[5] (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. Current work under MASS (Multilingual Affective Speech Synthesis) brings the concept of emotion into the speech synthesis, particularly via modelling of prosody. The research focusses on how to integrate emotional indicators such as pitch, energy and formant position into state of the art deep learning solutions.
- **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.

Speaker recognition and speech data analytics

In the context of the European SIIP project[6], illustrated by Figure 6, 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 three proof-of-concept and field-test events (2016 - 2018), 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. SIIP has also developed a framework allowing

[3] http://www.idiap.ch/project/siwis/
[4] http://www.idiap.ch/scientific-research/projects/dbox
[5] http://www.idiap.ch/scientific-research/projects/sp2
[6] http://www.siip.eu

to combine evidences extracted by different types of engines (i.e. inter-task engines such as language/accent/gender/age identification, or keyword-spotting) to eventually improve speaker identification.

As discussed in Section 2.8, and as part of the SNSF Project UniTS,[7], the group is also contributing to the Biometrics Security and Privacy group, including the development of countermeasures to detect attacks on speaker verification systems through forged speech samples.

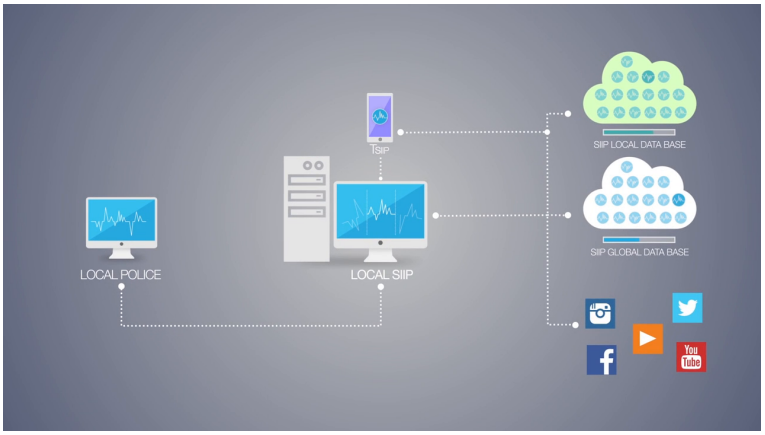

Figure 6: Illustration of SIIP speaker identification system exploiting lawfully intercepted calls, including multiple types of social-media information streams.

Forensic speaker verification and audio analysis

In 2018, Idiap continued offering its work in audio forensic sciences, aiming to extract 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

- **Query-by-example spoken term detection (QbE-STD) based on subspace modelling and detection**: Mainly exploiting its know-how in Deep Neural Networks (DNN) and sparse recovery modeling, Idiap continues its research efforts towards keyword spotting and spoken term detection with a focus on searching large audio archives using spoken queries. In 2018, we further explored retrieval solutions relying on the characterization and detection of the low-dimensional subspace of the DNN phonetic posteriors.
- **Large scale spoken query indexing**: In 2018, Idiap further developed a powerful methodology for large-scale 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 three years, Idiap has been actively developing novel methods to objectively assess intelligibility of speech signal. This work is now exploited in the context of the pathological speech processing activities discussed below.

[7] https://www.idiap.ch/scientific-research/projects/units

Pathological speech processing

In 2018, we expanded our research track focusing on the processing of pathological speech, mainly through two projects:

- MoSpeeDi (Motor Speech Disorder) SNSF Sinergia project: aiming at characterizing phonetic speech planning and motor speech programming/execution and their impairments, in collaboration with Marina Laganaro, Faculty of Psychology and Educational Science, University of Geneva, Cécile Fougeron, Laboratoire de Phonétique et Phonologie, Paris 3-Sorbonne Nouvelle, and Frédéric Assal, Neurology, Geneva University Hospitals and Faculty of Medicine, University of Geneva.
- EC TAPAS (Training Network on Automatic Processing of PAthological Speech): targeting three key research problems, (1) detection (develop speech processing techniques for early detection of conditions that impact on speech production), (2) therapy (using newly-emerging speech processing techniques to produce automated speech therapy tools), and (3) assisted Living (redesigning current speech technology so that it works well for people with speech impairments and also helps in making informed clinical choices.

Detection of impairments from speech signal

Motor speech disorders (MSDs) are a common symptom of several neurological diseases and refer to a broad range of impaired speech dimensions such as articulation, speech rate, voice, and prosody. MSDs can dramatically worsen the intelligibility, communication skills, and quality of life of patients. To assist clinicians in monitoring and managing MSDs, the multi-disciplinary SNSF-Sinergia "MoSpeeDi" project between the University of Geneva, University of Paris 3-Sorbonne Nouvelle, Geneva University Hospitals, and Idiao aims at i) developing accurate models to characterize non-impaired and impaired phonetic speech planning and motor speech programming, and ii) developing automatic techniques to detect and classify several speech impairments as well as to assess the intelligibility of patients. While our project partners are addressing the first objective, Idiap is addressing the second objective, summarized in Figures 7 and 8.

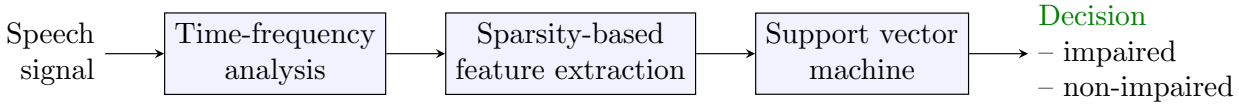

Figure 7: Automatic detection of impaired speech.

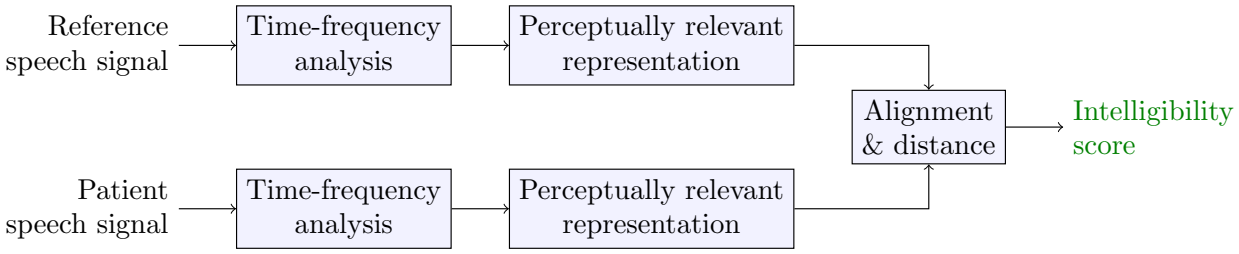

Figure 8: Automatic assessment of impaired speech intelligibility.

Sign language recognition and assessment

In the context of SNSF Sinergia project SMILE,[8] 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).

Sound localization and microphone array

In 2018, our research activities in the area of microphone array based speech processing and speaker diarization were mainly addressing the problem of source localization and speech detection and also reconstruction through binary sparse coding framework. In 2018, the group has also contributed to the Perception and Activity Understanding group through EC H2020 MuMMER project, focusing on audio source localization, speech detection and speaker re-identification applied in robotics.

Deep learning of Charisma

A small project on inference of charisma from textual data was completed in 2018. The concept of charisma arises from the field of leadership studies, where it has been correlated with success in leadership and politics. Our research work has cast the automatic classification of charisma as a text classification problem with a solution in the recent advances in deep learning. The automatic method opens the possibility for research on larger datasets in more diverse fields; in particular in conflict resolution.

Key publications

[1] P.-E. Honnet, B. Gerazov, A. Gjoreski, *et al.*, “Intonation modelling using a muscle model and perceptually weighted matching pursuit,” *Speech Communication*, 2018

[2] D. Ram, A. Asaei, and H. Bourlard, “Sparse subspace modeling for query by example spoken term detection,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2018

[3] S. Tong, P. N. Garner, and H. Bourlard, “Cross-lingual adaptation of a ctc-based multilingual acoustic model,” *Speech Communication*, vol. 104, pp. 39–46, Nov. 2018

[4] M. Razavi, R. Rasipuram, and M. Magimai.-Doss, “Towards weakly supervised acoustic subword unit discovery and lexicon development using hidden markov models,” *Speech Communication*, vol. 96, pp. 168–183, Feb. 2018

[5] S. Dey, S. Madikeri, and P. Motlicek, “End-to-end text-dependent speaker verification using novel distance measures,” in *Proceedings of Interspeech 2018*, 2018

[8] https://www.idiap.ch/scientific-research/projects/smile

2.2 Machine Learning

Overview

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

Group overview: Machine learning encompasses computer techniques that aim at learning automatically representations and decision rules from data. It lies at the intersection between statistics, algorithmic, and signal processing. The main objective of the Machine Learning group is the development of novel machine-learning and statistical methods, with a strong emphasis on their algorithmic efficiency.

The application domain for the methods we develop includes image and video processing, but also industrial sensor analysis and sequence modeling.

The group was composed in 2018 of one head of group, one post-doctoral researcher, four PhD students, one intern, and it had strong interactions with the software development team. François Fleuret is also the supervisor of one PhD student from the EPFL Space Engineering Center, and was the co-supervisor of two PhD students from the EPFL Computer Vision lab.

Key scientific outputs: In 2018, our work has resulted in contributions that improved planetary surface reconstruction from stereo images with large-size signal, knowledge transfer between large neural networks, speed-up of training and inference with deep models, training of generative adversarial models, and modeling of sequences with recurrent neural networks.

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

Efficient Machine learning

- **Bootstrapping for back-propagation**
  The computational requirement for training deep models is one of their most problematic aspects, even though most of the computation is spent on samples that are properly handled, and could be ignored.
  In the framework of the SNSF project ISUL, we proposed to mitigate this phenomenon with a principled importance-sampling scheme that focuses computation on “informative” examples, and reduces the variance of the stochastic gradients during training. We derived a tractable upper bound to the per-sample gradient norm, and an estimator of the variance reduction achieved with importance sampling, which enables us to switch it on when it will result in an actual speedup.
  The resulting scheme can be used by changing a few lines of code in a standard training procedure, and we demonstrated experimentally, on image classification, CNN fine-tuning, and RNN training, that for a fixed wall-clock time budget, it provides a reduction of the train losses of up to an order of magnitude and a substantial improvement of test errors.
- **Kronecker recurrent units:** In spite of their wide use, recurrent neural networks suffer from two main issues: they are over-parametrized, and recurrent weight matrices are often ill-conditioned. The former increases the sample complexity of learning and the training time. The latter causes the vanishing and exploding gradient problem, making the training unstable. We have developed a novel approach that relies on a Kronecker factorization of the weight matrix to achieve parameter efficiency.
  It overcomes the ill-conditioning of the recurrent matrix by enforcing soft unitary constraints on the factors. Thanks to the small dimensionality of the factors, maintaining these constraints is computationally efficient. Our experimental results on seven standard data-sets reveal that KRU can reduce the number of parameters by three orders of magnitude in the recurrent weight matrix

compared to the existing recurrent models, without trading the statistical performance. These results in particular show that while there are advantages in having a high dimensional recurrent space, the capacity of the recurrent part of the model can be dramatically reduced.

- **Jacobian matching**
  We started last year to develop a new line of work about transfer learning. This category of methods aims at taking advantage of pre-existing models to facilitate the training of new models, either by speeding it up, or by allowing it with very small amount of training data. The key notion is to "transfer" structures learned by the existing network, which capture information relevant to a new task.
  Or approach consists of a novel penalty that not only forces the model to be trained to mimic the response of the existing one, but to also mimic the dynamic of change of the output, given changes of the input. This can be envisioned geometrically as forcing curves not only to match in certain points of passage, but also to have the same slope there (see Fig. 9).

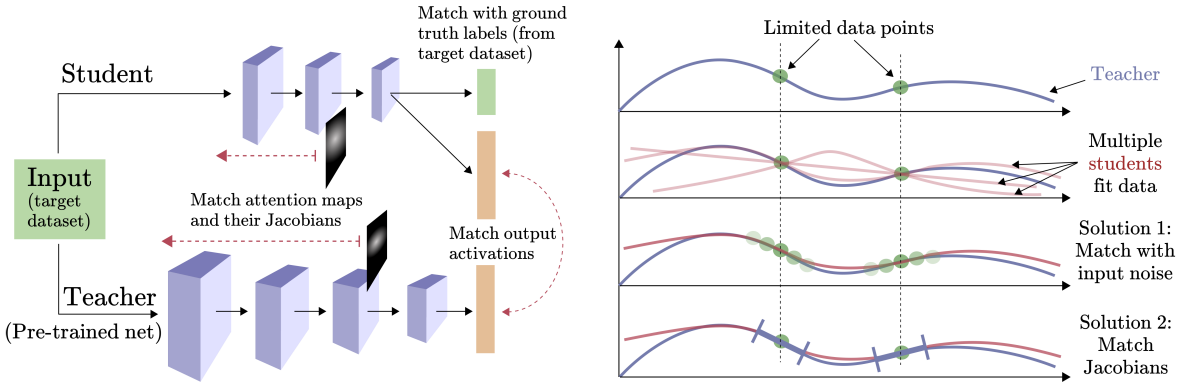

Figure 9: We have proposed a new strategy to use an existing neural network trained on a large corpus to improve the training of a new network on a small data-set (left). Our method aims not only at matching the responses of the models on the data points, but also to match how these responses change when the input to the model changes (right).

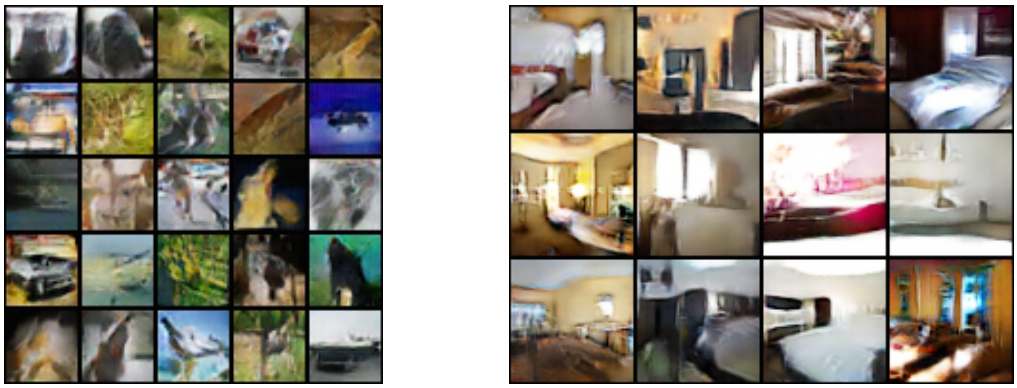

Figure 10: Synthetic images generated with a deep adversarial architecture trained with our SGAN optimization strategy.

Deep learning

- **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.
  End-to-end deep-learning networks demonstrate extremely good performance for stereo matching. However, existing networks are difficult to use for practical applications since they are

memory-hungry and unable to process even modest-size images, and they have to be trained for a specific disparity range.
The Deep Stereo network that we developed addresses both issues: First, its architecture relies on novel bottleneck modules that drastically reduce the memory footprint in inference, and additional design choices allow to handle greater image size during training. This results in a model that leverages large image context to resolve matching ambiguities. Second, a novel sub-pixel cross- entropy loss combined with a MAP estimator make this network less sensitive to ambiguous matches, and applicable to any disparity range without re-training.
This novel architecture demonstrates state-of-the-art performance on standard benchmark datasets.

- **Deep learning for multi-camera detection:** We have developed in the MEMUDE project funded by the Hasler Foundation a new approach to adapt a monocular deep-learning detector to a multi-camera context. We first 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.

- **Generative Adversarial Networks:** A very promising family of methods are the so called "Generative Adversarial Networks", which rely on training jointly two models, one synthesizing realistic signals (images, sound, text) and another trying to discriminate the synthetic signal from the genuine one. Such techniques have demonstrated striking performance in many application domains, but involve a complex and unstable optimization problem.
We have developed a new method that consists of training several such pairs in parallel, and maintaining carefully their statistical independence. This insures that their joint behavior has a good "covering" property, and we show experimentally that the resulting synthesis is less likely to miss sub-families of samples. (see Fig. 10).

- **Low-computation Drone Detection:** In a collaboration with ArmaSuisse, we have investigated the development of efficient neural networks for drone detection in video feeds captured by in-flight drones. The main objective of this project was to adapt object detectors to the limited computation capabilities of embedded hardware, usually one order of magnitude less powerful than standard hardware for machine learning. Our approach consist in penalizing during training the use of large number of parameters, pushing the process toward the removal of certain sub-structures of the model identified as less useful in the inference.

Key publications

[1] S. Tulyakov, A. Ivanov, and F. Fleuret. Practical Deep Stereo (PDS): Toward applications-friendly deep stereo matching. In Proceedings of the international conference on Neural Information Processing Systems (NeurIPS), pages 5874–5884, 2018.

[2] C. Jose, M. Cissé, and F. Fleuret. Kronecker Recurrent Units. In Proceedings of the International Conference on Machine Learning (ICML), pages 2380–2389, 2018.

[3] S. Srinivas and F. Fleuret. Knowledge Transfer with Jacobian Matching. In Proceedings of the International Conference on Machine Learning (ICML), pages 4723–4731, 2018.

[4] A. Katharopoulos and F. Fleuret. Not All Samples Are Created Equal: Deep Learning with Importance Sampling. In Proceedings of the International Conference on Machine Learning (ICML), pages 2525–2534, 2018.

[5] P. Baqué, E. Remelli, F. Fleuret, and P. Fua. Geodesic Convolutional Shape Optimization. In Proceedings of the International Conference on Machine Learning (ICML), pages 472–481, 2018.

[6] T. Chavdarova and F. Fleuret. SGAN: An Alternative Training of Generative Adversarial Networks. In Proceedings of the IEEE international conference on Computer Vision and Pattern Recognition (CVPR), pages 9407–9415, 2018.

2.3 Social Computing

Overview

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

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 2018 was composed of one group head, two postdoctoral researchers, two PhD students, one scientific collaborator, and one EPFL master student. The main research lines investigated in 2018 included social media analytics and mobile crowdsourcing for cities and health, and ubiquitous conversational interaction analysis.

Key scientific outputs: Publications on (1) social media analytics and mobile and online crowdsourcing to understand youth nightlife patterns and urban perception in cities; (2) multimodal analysis of soft skills in face-to-face employment interviews and hospitality service encounters. 13 EPFL PhD students have graduated from the group since 2002.

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

Social media analytics and mobile crowdsourcing for cities and health

Our work in this domain spans several research lines. First, in the context of the Dusk2Dawn project[9] (Characterizing Youth Nightlife Spaces, Activities, and Drinks, in collaboration with La Trobe University and the University of Zurich) supported by the Swiss National Science Foundation (SNSF), we investigated the use of mobile crowdsourcing and social media analytics to characterize urban phenomena related to nightlife, both in Switzerland and other countries. This included the automatic recognition of night drinking activity from smartphone sensor data (location, motion, bluetooth, wiki, and app logs), for a population of 200 young volunteers in Switzerland [1] [10], and the automatic recognition of place ambiance from both social media images and crowdsourced videos using deep learning (Figure 11). This research line was also extended to study the phenomenon of place ambiance perception on place-centric online platforms like Airbnb, as well as deep-learning based automatic ambiance recognition [2], thus extending the research on social perception of places from public spaces to personal ones [11].

Regarding mobile crowdsourcing and health, in the context of the Bites-n-Bits project [12] (Understanding Eating Routines in Context, in collaboration with Nestle Research Center), we continued our work on analysis of rich multimedia data collected through smartphone sensing to capture everyday life eating and drinking patterns for a population of 120 college students. Furthermore, using visual social media (Instagram) as data source, we analyzed drinking patterns combining hashtags, other textual sources, and image data, to distinguish nuanced uses of social media posts with respect to alcohol consumption.

Regarding mobile crowdsourcing for social innovation, we continued our work using the Civique platform[13] (Mobile Data for Local Causes), which has shown its versatility of use, ranging from supporting cities to collect information related to urban issues like street harassment, to teaching students about humanitarian technologies, and to collecting audio and visual data from Swiss heritage speakers living abroad. Finally, our work in Latin America on mobile crowdsourcing for youth engagement on local

[9] http://www.idiap.ch/project/dusk2dawn
[10] https://www.idiap.ch/en/allnews/using-mobile-data-to-model-the-drinking-habits-of-swiss-youth
[11] https://www.idiap.ch/en/allnews/modeling-human-behavior-with-airbnb
[12] https://www.bitesnbits.org/
[13] https://www.civique.org

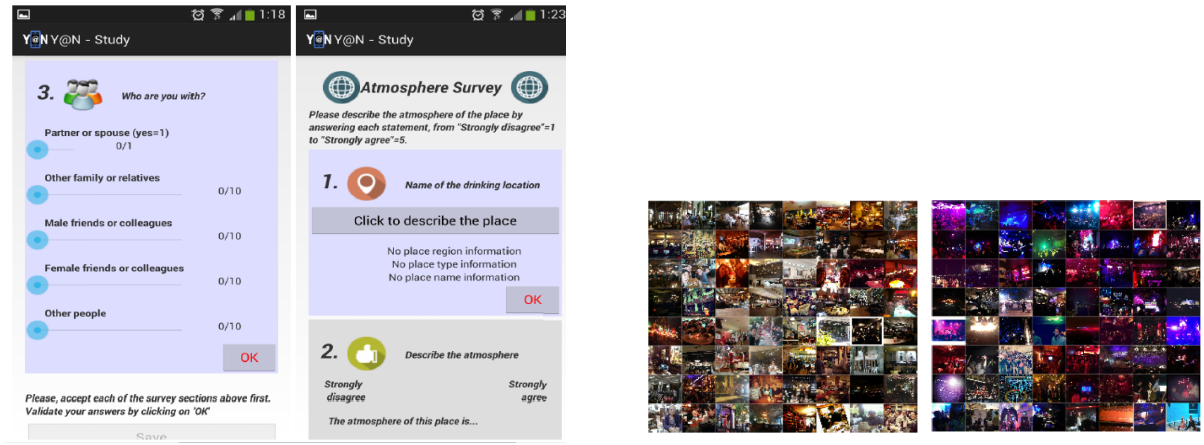

Figure 11: Left: Youth@Night survey logger app. Right: Social media images classified as Restaurant and Stage by convolutional neural network.

urban issues [3], integrating youth-driven data collection, machine-learning based data analyses, and community feedback, continued through collaborations with institutions in Mexico [14] and Colombia [15]. One of the key motivations of our international work is the advocacy for diversity in data and algorithms to improve the representation of non-western cities and citizens [16]. Other collaborations using the Civique platform are under development.

Ubiquitous interaction analytics

In the context of the SNSF UBImpressed project[17] (Ubiquitous First Impressions and Ubiquitous Awareness, in collaboration with University of Lausanne and Cornell University), we developed methods to analyze dyadic interactions in workplace situations (job interviews, hospitality service encounters, and online resumes and video interviews) using multiple sensors (cameras, Kinect, microphone arrays, and wearables), and to infer variables like perceived soft skills (Figure 12). We found connections between certain nonverbal behavior (including speaking activity, prosody, head activity, and body motion) and impressions in job interviews and hospitality service encounters, some of which hold across situations [4]. Some of this work was done in collaboration with Idiap's Perception and Activity Understanding Group [5]. Recognizing this research as innovative, D. Gatica-Perez was named in 2018 as one of the "Romandes Who Invent the Future" by the economic magazine Bilan [18].

Some of the above work has been transfered to industry in the context of the Innosuisse RISE project (collaboration with VimaLink SA, a Valais-based startup, and University of Lausanne). Furthermore, a new SNSF interdisciplinary project under the Digital Lives program was launched (in collaboration with University of Neuchatel and University of Lausanne about machine learning, behavior computing, and feedback in video selection interviews. The project is inserted within the current federal interest to support research addressing questions about the Swiss digital transformation.

[14] https://www.idiap.ch/en/allnews/collaboration-with-mexico-at-first-national-forum-on-science-te
[15] https://www.idiap.ch/en/allnews/children-education-project-supported-by-sensecityvity-at-the-wo
[16] https://www.idiap.ch/en/allnews/ai-and-diversity-the-whole-world-is-not-yet-represented
[17] http://www.idiap.ch/project/ubimpressed
[18] https://www.bilan.ch/techno/ces-romands-qui-inventent-le-futur

Figure 12: Sensors available in the social sensing lab. Kinect V2, Microcone, Q-Sensor, Android smartphone, Google Glass.

Key publications

[1] D. Santani, T. Do, F. Labhart, S. Landolt, E. Kunstche, and D. Gatica-Perez, DrinkSense: Characterizing Youth Drinking Behavior using Smartphones, IEEE Transactions on Mobile Computing, Vol. 17, No. 10, pp. 2279-2292, Oct. 2018.

[2] L. S. Nguyen, S. Ruiz-Correa, M. Schmid Mast, and D. Gatica-Perez, Check Out This Place: Inferring Ambiance from Airbnb Photos, IEEE Transactions on Multimedia, Vol. 20, No. 6, pp. 1499-1511, Jun. 2018.

[3] D. Santani, S. Ruiz-Correa, and D. Gatica-Perez, Looking South: Learning Urban Perception in Developing Cities, ACM Transactions on Social Computing, Vol. 1, No. 3, Article 13, Dec. 2018.

[4] S. Muralidhar, M. Schmid Mast, and D. Gatica-Perez, A Tale of Two Interactions: Inferring Performance in Hospitality Encounters from Cross-Situation Social Sensing, PACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT), Vol. 2, No. 3, Sep. 2018

[5] S. Muralidhar, R. Siegfried, J.-M. Odobez, and D. Gatica-Perez, Facing Employers and Customers: What Do Gaze and Expressions Tell About Soft Skills? in Int. Conf. on Mobile and Ubiquitous Multimedia (MUM), Cairo, Nov. 2018

2.4 Perception and Activity Understanding

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 2018, the group was composed of one group head, three post-doctoral researchers, seven PhD students, one research engineer from the development team.

Key scientific outputs: The group is known for its work on probabilistic tracking, non-verbal behavior extraction (in particular attention modeling), and temporal motif discovery. In 2015 and 2016, the PAU team ranked first at the MediaEval Person discovery challenge. Its continuing and patended work on 3D face and gaze tracking from cheap RGB-Depth sensors has led to the creation of the Eyeware SA start-up company. In recent years, the group has investigated deep learning methods for several tasks like gesture recognition, audio-visual speaking activity modeling, gaze, joint audio localization and speech/non-speech detection, body landmark detection, and multimedia processing (cross-modal transfer learning, shape recognition, text localization and semantic extraction). It is also working on the integration of its sensing technology and algorithms into real-time perceptual systems used in collaboration projects (human-robot interaction Pepper platform, EU MuMMER project), or by companies (e.g. anti-tailgating detection system). During the period 2014-2018, the group published 16 journal papers, above 40 conference papers, and filled 3 patents.

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

Human activity analysis, non-nerbal behavior (NVB) extraction, human-robot interactions

Our long-standing work in this domain spans several research lines.

Gesture Recognition. In [5] we studied multimodal Deep Dynamic Neural Networks (DDNN) for the segmentation and recognition of short spontaneous communicative gestures, as illustrated in Figure 13, left. We adopted a semi-supervised hierarchical approach relying on high-level spatio-temporal gesture representations trained using DNN suited to the input modalities (dynamics of skeletal joint information,

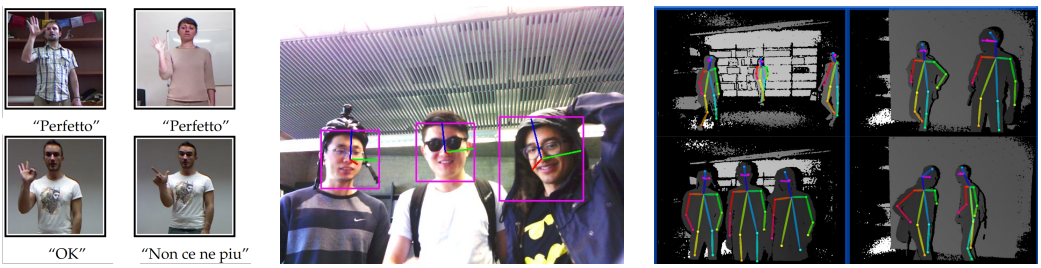

Figure 13: Human activity analysis. Left: Example of multimodal spontaneous communicative gesture recognition from RGB-D data, using deep-learning methods. Middle: real-time head pose estimation from color images. Right: real-time body landmark detection from depth images.

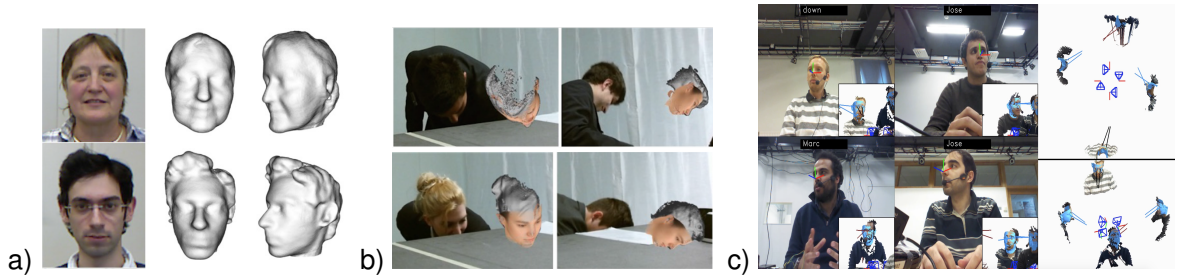

Figure 14: Attention analysis - Accurate and robust 3D head and gaze tracking. a) Example of online reconstructed heads. b) Tracking results in adverse conditions (UBImpressed registration desk set-up). c) Gaze coding in multi-party interactions: Heads are tracked, eye localized and aligned before inferring gaze and attention towards people or other semantic labels (all information analysed in 3D).

fusion of batches of depth and RGB images).

Efficient head pose analysis from color images. We have developed a realtime head pose estimation algorithm, improving accuracy and efficiency by leveraging the output of a 2D body landmark localization (Figure 13).

Head and body pose inference from RGB-D (color and depth) data. A recent achievement has been the development of a robust and accurate head pose tracking (below 2 degreess of errors) framework from RGB-D data [1], making head tracking a commodity for situations up to 1.5m. Combining the benefits of the online fitting of a 3D face morphable model with the online 3D reconstruction of the full head (Figure 14a), the approach provides more support when handling extreme head poses (Figure 14b). Research on fast people detection, 2D and 3D body pose estimation and tracking is also conducted. For the 2D pose estimation task, we proposed a new Resnet DNN architecture and leveraged the synthesis of a large training dataset, and obtained good results as demonstrated by experiments with three different depth sensors and as illustrated in Figure 13. Investigations on adversarial training for domain adaptation (synthetic to real images), and on distillation for learning even faster architectures showed that the former does not work well, while the latter is indeed useful.

Gaze inference in the 3D space. We designed a differential approach (patented) working much better than state-of-the-art methods on all benchmark datasets. A DNN learns to predict gaze differences from the same eye, better eliminating in this way all nuisance parameters (eye shape, illumination conditions, inaccurate cropping), which can impact gaze prediction. Other works include robust eye segmentation using DNN and leveraging colorisation, as well as online adaptation techniques leveraging social situations for weak target labeling. Figure 14 illustrates the result of applying our tools for gaze coding in multiparty situations.

Perception system. We continue improving our real-time multi-person perception system (Pepper robot in the EU MuMMER project), with features including tracking, pose estimation, re-identification, extraction of non-verbal cues (attention, head gestures), as well as speech sound localization.

Multimedia and multimodal analysis

Semantic text recognition (OCR). In the context of the Innoswiss VIEW project, we investigated deep learning methods for the detection, segmentation, categorization and recognition of text content in slides, allowing further semantic tagging (see Figure 15). Models and systems have been transfered to the Klewel company, improving the internet referencing and visibiliy of the recorded events of its client.

Audio analysis. In the context of the EU MuMMER project on social robotics, we are investigating different DNN architectures for sound processing. We have proposed an efficient multi-task approach for the joint localization and categorization (speech vs non-speech) of multiple sources from a microphone array, a frequent situation for robots placed in public spaces. Our methods relies on raw frequency rep-

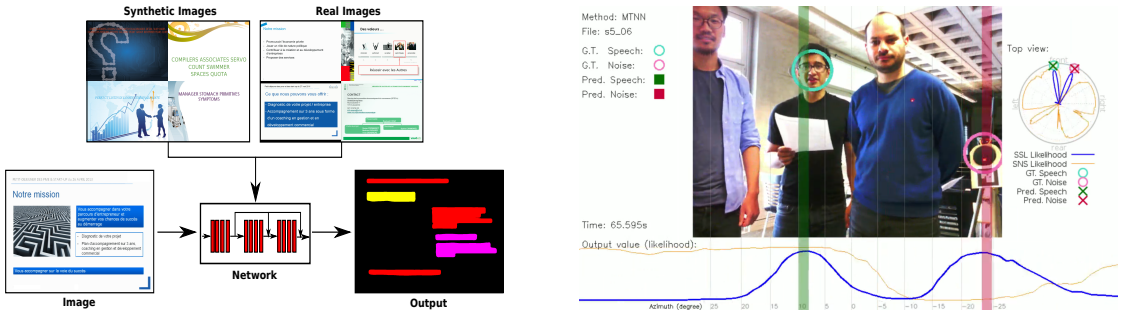

Figure 15: Left: Semantic Text Segmentation Network (STSN) trained with synthetic and real data, that apart from localizing text in images with high accuracy also classifies text regions in different semantic categories (title, bullets, standard text), and then further recognizes text (OCR). Right: Multiple sound source localization and discrimination for HRI. A multi-task network learns to predict for each direction (circle on the right) the likelihood that there is a sound source (blue curve) and whether it is a speech sound (yellow curve). The sound sources can then be mapped to the visual data.

resentations of the audio signals, keeping both the phase information (important for localization) and spectrum (essential for catgorization), and on a mixture of training data. It is illustrated in Figure 15, and was nominated for the best student paper award at Interspeech. The method was demonstrated to perform much better than state-of-the-art deep learning approaches (e.g. based on localisation, beam-forming, and categorization), both on benchmark data and when using our realtime implementation on the Pepper robot during public events.

Multimodal face and person diarization and naming. Identifying people appearing and speaking in multimedia data as we have done 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 or broadcast programs. On addition to face tracking, clustering, audio-visual association, we also investigated deep-temporal representations for voice representation like the modeling of audio-visual speaking activities, e.g. to differentiate dubbing from genuine talking situations, or cross-modal domain adaptation and transfer learning methods for improving the estimation of short utterance speaker embedding leveraging face embeddings [3].

Multiple object tracking

Our previous work resulted in an enhanced Conditional Random Field tracking-by-detection framework with important characteristics making it successful for person or face tracking 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.

Key publications

[1] Y. Yu, K. Funes and J.-M Odobez. HeadFusion: 360^{o}Head Pose tracking combining 3D Morphable Model and 3D Reconstruction *IEEE Trans. Pattern Anal. Mach. Intell.*, 40(11): 2653-2667 (2018).

[2] A. Martínez-Gonzàlez, M. Villamizar, O. Canévet, J.-M. Odobez. Real-time Convolutional Networks for Depth-based Human Pose Estimation. *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, sept. 2018.

[3] N. Le and J.-M. Odobez. Improving speech embedding using crossmodal transfer learning with audio-visual data, In *Multimedia Tools and Applications*, published online Dec. 2018.

[4] G. Liu, Y, Yu, K. Funes and J.-M Odobez. A Differential Approach for Gaze Estimation with Calibration. *British Machine Vision Conference (BMVC)*, sept. 2018.

[5] 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.

2.5 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 human-centered 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 the development of intuitive active learning interfaces to acquire meaningful demonstrations, the development of models that can exploit the structure and geometry of the acquired data in an efficient way, and the development of adaptive control techniques that can exploit the learned task variations and coordination patterns.

The Robot Learning and Interaction group in 2018 was composed of two postdoctoral fellows, six PhD students and two visiting PhD/MSc students.

Key scientific outputs: Development of robot learning and adaptive 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 in DexROV). Attentive to reproducible research, the group regularly releases open source codes accompanying its 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

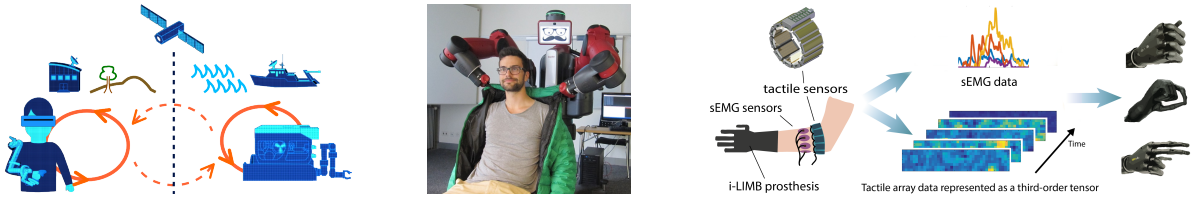

Figure 16: Human-centered robotic applications. ***Left:*** *Teleoperation assistance for the control of a bimanual robot (DexROV project), where the robot assists the teleoperator when repetitive or structured tasks are detected, as a form of human-robot collaboration. For example, if the robot observes that the task of drilling requires the drill to be perpendicular to a surface, the same tool will later be automatically oriented when the robot approaches a surface to drill, letting the teleoperator concentrate on the location to drill while letting the robot maintain the correct orientation.* ***Center:*** *Personalized assistance in dressing (I-DRESS project), where the robot learns by demonstration the preferences and type of assistance required by each user. This is achieved by extending movement primitives representations 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 coordinate systems that could potentially be relevant for the movement/task to transfer. A statistical learning approach is then used to determine the variability and coordination patterns in the movement by considering different coordinate systems simultaneously. This approach allows the orchestration of multiple coordinate systems to reproduce movements in new situations (typically, to adapt a 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.

A distinctive usage of optimal control techniques

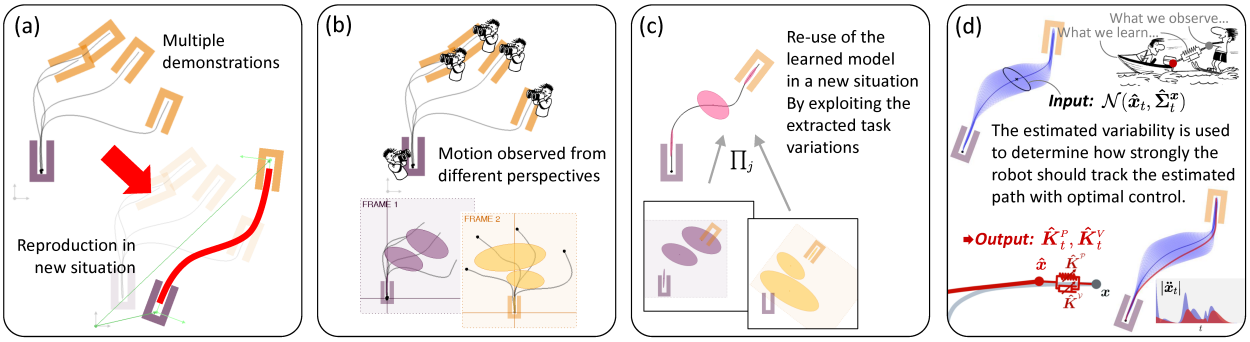

Figure 17: Task-parameterized model to synthesize movements in new situations. (a) *Observation in different situations to extract the underlying structure of the task.* (b) *Probabilistic encoding of the task in multiple coordinate systems.* (c) *The cross-situational observations are used to adapt the motion to new situations.* (d) *Model predictive control strategy to reproduce the movement by exploiting the retrieved variability and correlation information.*

Model predictive control (MPC) and linear quadratic tracking (LQT) are ubiquitous in robot control, but the core formulation of such optimal control problem and its 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 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 analytically, by taking into account variation and coordination constraints. Instead of learning trajectories directly, the approach allows the learning of the underlying controllers to move 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 modify autonomously the tracking gains and compliance required to reproduce a task.

Geometry-aware statistical learning and control

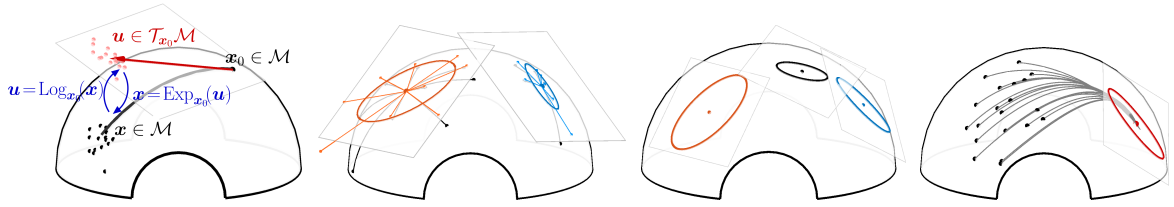

Figure 18: Statistics and control on Riemannian manifold. From left to right: Average of datapoints computed by iterative mappings between tangent spaces and manifold, encoding as Gaussian mixture model, fusion of information with product of Gaussians, linear quadratic tracking.

The data encountered in robotics are characterized by simple but varied geometries, which are often underexploited when developing learning and control algorithms. Such data range from joint angles in revolving articulations, rigid body motions, orientations represented as unit quaternions, sensory data processed as spatial covariances, or other forms of symmetric positive definite matrices such as inertia or manipulability ellipsoids. Moreover, many applications require these data to be handled altogether.

We exploit Riemannian manifold techniques to extend algorithms initially developed for Euclidean data, by efficiently taking into account prior knowledge about these manifolds and by modeling joint distributions among these heterogeneous data. The use of these differential geometry techniques allow us to treat data of various forms in a unified manner. It can typically be used to revisit common optimization problems in robotics formulated in standard Euclidean spaces, by treating them as unconstrained problems inherently taking into account the geometry of the data.

Key publications

[1] Tanwani, A.K. and Calinon, S. (2019). Small Variance Asymptotics for Non-Parametric Online Robot Learning. International Journal of Robotics Research (IJRR), 38:1, 3–22.

[2] Silvério, J., Calinon, S., Rozo, L. and Caldwell, D.G. (2019). Learning Task Priorities from Demonstrations. IEEE Transactions on Robotics, 35:1, 78–94.

[3] Calinon, S. (2018). Robot Learning with Task-Parameterized Generative Models. Bicchi, A. and Burgard, W. (eds). Robotics Research, pp. 111–126. Springer International Publishing.

[4] Ficuciello, F., Falco, P. and Calinon, S. (2018). A Brief Survey on the Role of Dimensionality Reduction in Manipulation Learning and Control. IEEE Robotics and Automation Letters (RA-L), 3:3, 2608–2615.

[5] Jaquier, N., Rozo, L., Caldwell, D.G and Calinon, S. (2018). Geometry-aware Tracking of Manipulability Ellipsoids. In Proc. of Robotics: Science and Systems (RSS).

2.6 Uncertainty Quantification and Optimal Design

Overview

Head: Prof. Dr. David Ginsbourger (Ph.D. Mines Saint-Etienne 2009, Habilitation Univ. Bern 2014, Titularprofessor Univ. Bern 2018)

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 2018, the UQOD group has been composed of a permanent senior researcher, two PhD students, an intern within a collaboration with the Centre de Recherches Énergétiques et Municipales (CREM), and three visiting students (one PhD student at the University of Neuchâtel and two UniBE master students).

Key scientific outputs: Current contributions include efficient algorithms for Bayesian optimization and 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, a collaboration with hydrogeologists towards flow simulation-based contaminant source localization, and a new Swiss National Science Foundation project on uncertainty quantification and efficient design of experiments for data- and simulation-driven inverse problem solving with applications in computational cosmology and geophysics.

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 processes (GPs) has become a standard for optimizing prohibitively expensive to evaluate systems, e.g. with response(s) of interest stemming from heavy numerical simulations. More generally, sequential design of (computer) experiments based on GPs have flourished in the last few decades to efficiently addess a variety of goals (See for instance [2] for a brief overview). This constitutes one of the core domains of expertise of the UQOD group, with recent contributions ranging from theoretical to methodological questions (such as parallelization, handling large data sets, coping with high-dimensional inputs, etc.) and 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 [5].

The UQOD group also investigates GP emulation per se, with a 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. Recent work includes non-stationary GP modelling using warping approaches, see for instance [1], where the Warped Multi-Index (WaMI) GP model is introduced (Fig. 19). Also, several new aquisition functions using GP derivatives are considered for sequential design in the context of functions with heterogeneous variations.

UQOD has also been involved in a collaboration with colleagues from CentraleSupélec and Toulouse University towards establishing convergence results for some important classes of Sequential Uncer-

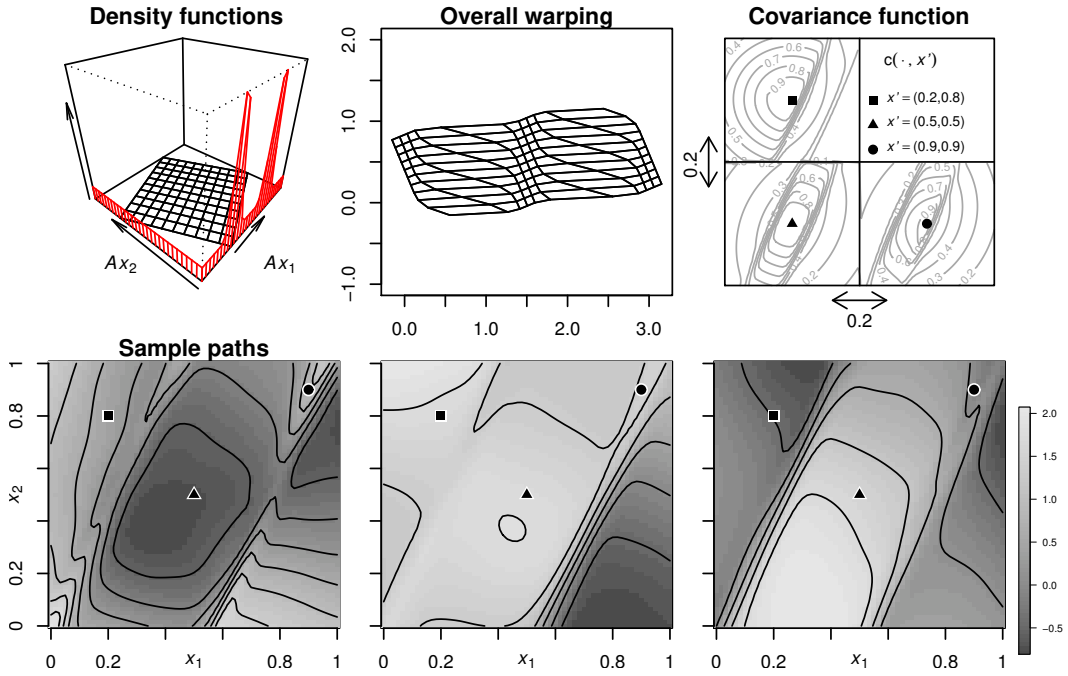

Figure 19: WaMI-GP example with two high-variation regions. From left to right: density functions of the deformations in each direction after linear transformation, warping of a grid, the covariance function $c(\cdot, \mathbf{x}'))$ for different values of $\mathbf{x}'$, and three corresponding WaMI-GP realizations.

tainty Reduction strategies, with links to Bayesian optimization. The article [3] "A supermartingale approach to Gaussian process based sequential design of experiments" is now accepted to Bernoulli. The main theoretical results of this paper actually not only apply to the Expected Improvement but to a whole class of criteria deriving from a variety of "uncertainty functionals" as illustrated by three other examples.

Computer experiments for the quantification and the reduction of uncertainties

Besides global optimization, UQOD has indeed also been focusing on strategies dedicated to other goals such as locating parameter regions leading to a response exceeding a given threshold, corresponding e.g. to an abnormal behaviour of the considered system.

In the previous project "Learning and visualizing dangerous regions in multivariate parameter spaces" (January-August 2017, funded by the Hasler foundation), novel approaches based on profile extrema have been developed for visualizing and quantifying uncertainties on excursion regions with application to a coastal flooding test case provided by colleagues from BRGM (Bureau de Recherches Géologiques et Minières, Orléans, France). Figure 20 represents the kind of output delivered by the approach developed in this project and in the recently accepted paper [4]: for each individual parameter of interest, estimates of maximum/minimum response with respect to all remaining parameters are estimated in function the individual parameter in question. The article presents an approach to estimate such profiles and to deliver associated uncertainty assessments in cases where the functions are expensive to evaluate and are modelled using GPs.

More generally, the combination of GP modelling and stochastic simulation techniques have been investigated within UQOD for quantifying and reducing uncertainties on sets, with example application in safety engineering with goals such as identifying sets of dangerous and/or safe configurations of a complex system and if possible providing some measures of confidence along with the estimate(s).

Previously developed approaches on "asymmetric nested Monte Carlo" algorithms and their use for ef-

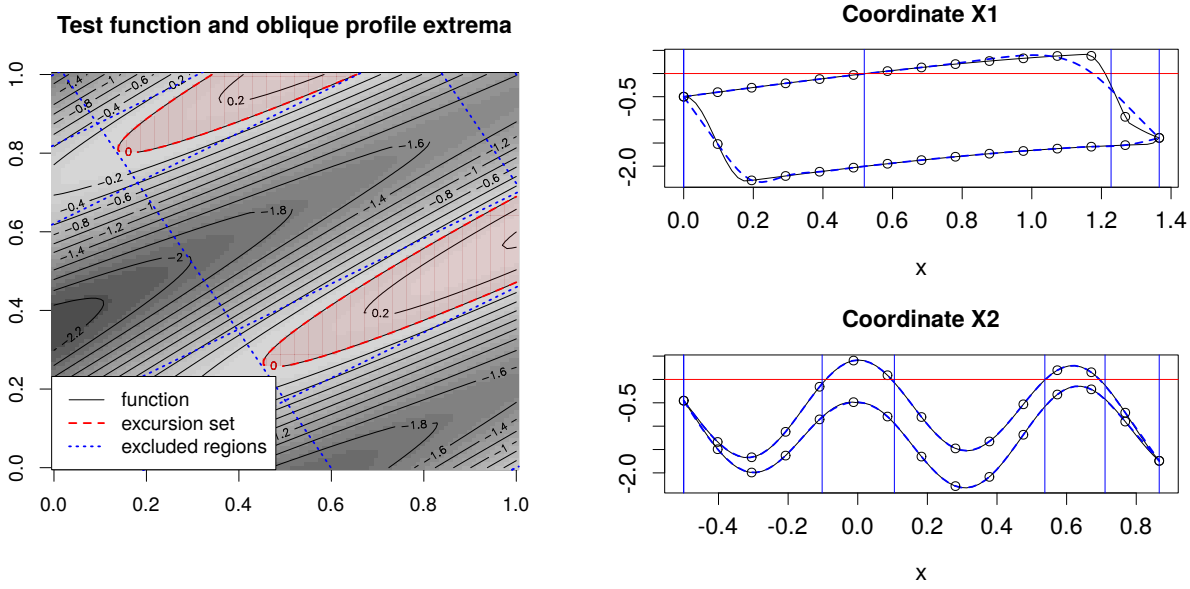

Figure 20: Left: Regions excluded by oblique profile extrema delimited by dotted lines on a 2-dimensional example test case. Right: Oblique profile extrema functions for γ (black solid lines) and their approximation (blue dashed lines) from $k = 15$ points.

ficiently estimating orthant probabilities of high-dimensional Gaussian vectors have been instrumental to derive conservative set estimates on a neutronic criticality safety test case (from IRSN, the French Institut de Radioprotection et de Sûreté Nucléaire) starting from a initial experiments, and also to derive sequential design strategies dedicated to this class of conservative set estimation problem, that were applied to this IRSN test case and also to another BRGM test case. This collaboration with colleagues from Neuchâtel, CentraleSupélec and IRSN, summarized in the preprint “Adaptive Design of Experiments for Conservative Estimation of Excursion Sets” (arXiv:1611.07256), has kept active in 2018.

As for the collaboration between UQOD and CREM, the internship of Athénaïs Gautier has taken place from April to September 2018, and has lead to several outcomes of both practical and theoretical nature regarding the sensitivity of the total energetic cost of a holiday resort in Valais to uncertain unit costs coming into play within the mixed-integer programming problem formulated by CREM.

Key publications

[1] Warped Gaussian processes and derivative-based sequential design for functions with heterogeneous variations, Sébastien Marmin, David Ginsbourger, Jean Baccou and Jacques Liandrat, in: *SIAM/ASA Journal on Uncertainty Quantification*, 6(3):991-1018, 2018.

[2] Sequential Design of Computer Experiments, David Ginsbourger, in: *Wiley StatsRef: Statistics Reference Online*, Wiley, 2018.

[3] A supermartingale approach to Gaussian process based sequential design of experiments, Julien Bect, François Bachoc and David Ginsbourger, in: *Bernoulli*, forthcoming (Accepted in 2018).

[4] Profile extrema for visualizing and quantifying uncertainties on excursion regions. Application to coastal flooding, Dario Azzimonti, David Ginsbourger, Jérémy Rohmer and Déborah Idier, in: *Technometrics*, forthcoming (Accepted in 2018).

[5] Contaminant source localization via Bayesian global optimization, Guillaume Pirot, Tipaluck Krityakierne, David Ginsbourger and Philippe Renard, in: *Hydrology and Earth System Sciences*, 23:351-369, 2019.

2.7 Computational Bioimaging

Overview

Head: Dr. Michael Liebling (MS, EPFL, 2000; PhD, EPFL 2004; postdoc, Caltech, 2004–2007; Assistant Prof (2007–2013), Associate Prof (2013–2017), Adjunct Prof (2017–) University of California Santa Barbara (UCSB)

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 multi-modality images.

In 2018, the Computational Bioimaging Group was composed of the head of the group and four PhD students.

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.

Spatially-variant deconvolution with local point spread function parameter estimation via a convolutional neural network

Optical microscopy allows biologists to acquire both qualitative and quantitative data about cellular function, organ development, or diseases, even within live organisms. Light diffraction (both within the sample and the microscope) limits achievable resolution. When the diffraction process can be accurately described mathematically, images can be improved computationally. In practice, it is often difficult to measure or infer the source of the degradation, in part because it varies in different locations of the image. To address this issue, we developed a semi-blind, spatially-variant deconvolution technique aimed at optical microscopy. The method combines a local estimation step of the point spread function (PSF; which fully characterizes the optical properties of the imaging system that lead to image degradation) followed by a deconvolution step based on a spatially-variant, regularized Richardson-Lucy algorithm. To find the local PSF map in a computationally tractable way, we devised a convolutional neural network (CNN) that performs a regression of a parametric model of the PSF. We trained the CNN by synthetically blurring a library of image patches, using point spread functions of known parameters. This approach has the following features: (1) It does not require the experimental measurement of a PSF, only synthetic training data is necessary; (2) compared to non-parametric blind deconvolution techniques, the problem complexity remains low and therefore is more easily amenable to optimization; (3) parameters with a physical meaning (Zernike polynomials) are inferred from the image itself; (4) the algorithm is computationally efficient, resulting in a near real-time kernel regression and mapping (see Figure 21). This research is part of the SNSF funded project “COMP-BIO Computational biomicroscopy: advanced image processing methods to quantify live biological systems” and was carried out on Idiap’s imaging platform, SNSF R’Equip project “Platform for Reproducible Acquisition, Processing, and Sharing of Dynamic, Multi-Modal Data.”

Imaging cardiac development at the microscopic scale

Fluorescence microscopy is a widely-used modality to study developing organs in animal models, *in vivo*. Such studies are a key step to understanding many diseases, including congenital heart defects,

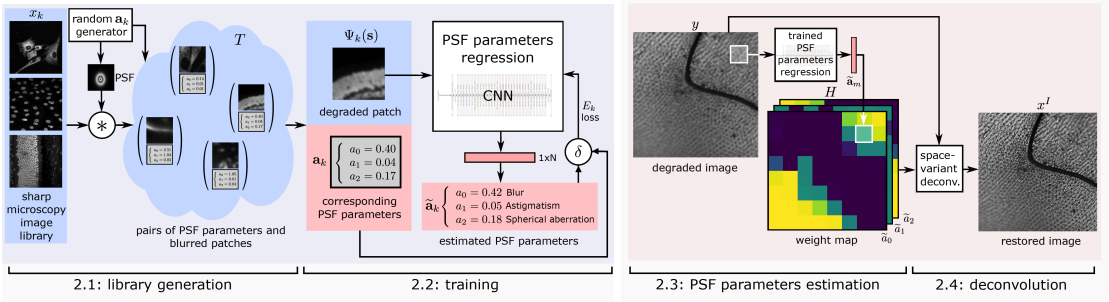

Figure 21: Left: method to estimate the parameters of the PSF that degraded an image patch. Right: method to estimate a map of the local PSF that degraded an image that can be fed into a spatially variant deconvolution algorithm. Adapted from [1].

a leading cause of birth defect-associated infant illness and death. The small scale and fast motions within the beating heart are a challenging imaging environment. Our group is addressing the problem of reconstructing dynamic images via multiple paths.

As part of the SNSF project COMET-S "Computational Methods for Temporal Super-resolution Microscopy", we developed a multi-spectral widefield microscopy method to image the beating heart in zebrafish. The method calls for acquiring a set of movies of the beating heart over multiple heartbeats, with each movie recording a signal filtered in a different (and possibly overlapping) range of wavelengths. From these multi-channel data, we build a single multi-spectral movie of the beating heart through temporal registration of the movies followed by color unmixing to yield non-overlapping wavelength ranges. We collaborated with the Institut de Recherche en Opthalmologie in Sion [2].

Within the joint SNSF and French ANR project "liveheart: The cellular basis of cardiac development revealed by live imaging," our group developed methods for virtual high-framerate microscopy of the beating heart, via a computational sorting approach of still images. Since achieving a sufficiently high framerate is difficult with conventional cameras, yet the latter are often the only available instruments to biologists, we developed a method to reconstruct an image sequence covering one full heartbeat from images acquired over multiple cardiac cycles, with each image triggered at an arbitrary time by sorting them according to their similarity. We formulated this task as a traveling salesman problem for which efficient solutions are available. We characterized our approach by evaluating its accuracy on synthetically generated data and sub-sampled high-speed movies of the beating heart in zebrafish larvae. We found that reconstructions are reliable when each phase produces a distinct image and when there are no abrupt cardiac motions, which amounts to collecting at least 100 images in a typical microscopy imaging scenario. We demonstrated that our method can be applied on data acquired with a fast confocal microscope (in collaboration with our partners at Uni-Bern), increasing its limited frame-rate by a factor 8.

Key publications

[1] A. Shajkofci and M. Liebling, "Semi-blind spatially-variant deconvolution in optical microscopy with local point spread function estimation by use of convolutional neural networks," 2018 25th IEEE International Conference on Image Processing (ICIP), Oct. 2018, pp. 2381-8549.

[2] O. Mariani, A. Ernst, N. Mercader, M. Liebling "Virtual High-Framerate Microscopy of the Beating Heart via Sorting of Still Images," IEEE International Symposium on Biomedical Imaging ISBI'19, *in press*.

[3] C. Jaques, L. Bapst-Wicht, D. Schorderet, M. Liebling, "Multi-Spectral Widefield Microscopy of the Beating Heart through Post-Acquisition Synchronization and Unmixing," IEEE International Symposium on Biomedical Imaging ISBI'19, *in press*.

2.8 Biometrics Security and Privacy

Overview

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

Group overview:

Biometrics refers to the automatic recognition of individuals based on their physiological and/or behavioral characteristics. The Biometrics Security and Privacy (BSP) group at Idiap focuses on three main areas of research:

- Biometric recognition: We investigate and develop new biometrics based recognition algorithms, notably for face, voice, and vein biometric modalities.
- Presentation attack detection: We look for new and better ways of detecting presentation attacks on face, voice, and vein biometric recognition systems.
- Biometric template protection: We research effective methods of preserving both the security of biometric recognition systems and the privacy of their users by protecting the biometric models ("templates") that are employed by the system for recognition purposes.

The BSP group prioritizes reproducibility in research. This is important for ensuring that our work can be both verified and built upon by the wider research community. To enable reproducibility, we mainly make use of our Python-based signal-processing and machine-learning toolbox, Bob (http://www.idiap.ch/software/bob/), which we make freely available for academic purposes. The group also develops and maintains the BEAT platform (https://www.beat-eu.org/platform/) – a MLaaS platform compliant with Swiss and European data-security norms.

The group participates in several large-scale biometrics projects at Swiss (SNSF), European (H2020) or world-wide levels (eg. IARPA/DARPA) but also conducts projects directly with companies.

The BSP group provides also expertise to the Swiss Center for Biometrics Research and Testing, which, among other things, carries out tests and evaluations on commercial products related to biometrics.

The BSP group in 2018 was composed of 1 head of group, 4 research associates, 5 postdocs, 3 PhD students, 1 intern and 1 engineer.

Key scientific outputs: The BSP 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 BSP group also investigated approaches for heterogeneous face recognition, vein recognition and more recently Deepfakes to assess its threat to face recognition.

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

Deepfakes: a New Threat to Face Recognition ?

It is becoming increasingly easy to automatically replace a face of one person in a video with the face of another person by using a pre-trained generative adversarial network (GAN). Recent public scandals call for automated ways to detect these Deepfake videos. To help developing such methods, we produced and analysed Deepfake videos. We showed that the state of the art face recognition systems based on Neural Networks are vulnerable to Deepfake videos, with about 90% false acceptance rates , which means methods for detecting Deepfake videos are necessary. By considering several baseline approaches, we found that audio-visual approach based on lip-sync inconsistency detection was not able to distinguish Deepfake videos. The best performing method, which is based on visual quality metrics and is often used in presentation attack detection domain, resulted in 8.97% equal error rate on high quality Deepfakes. Our experiments demonstrate that GAN-generated Deepfake videos are challenging for both face recognition systems and existing detection methods, and the further development of face swapping technology will make it even more so.

(a) Original Donor (b) Original Target (c) Face Swapped

Figure 22: Illustration of Deepfakes

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.

Figure 23: 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).

Heterogeneous face recognition

The task of Heterogeneous Face Recognition (Figure 23) 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 demonstrated that high level features of Deep Convolutional Neural Networks trained on visual spectra images are domain independent and can be used to encode faces sensed in different image domains.

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 24). 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.

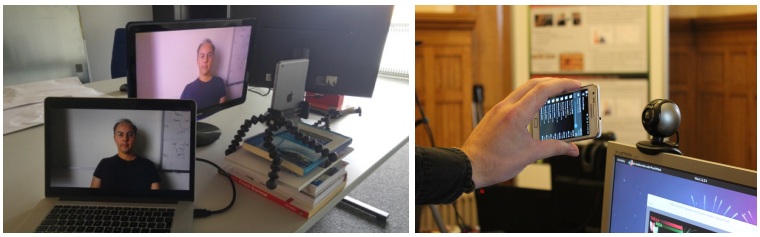

Figure 24: Illustration of video and audio presentation attacks.

Speaker recognition (SR) systems are highly vulnerable to presentation attacks (Figure 25) 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.

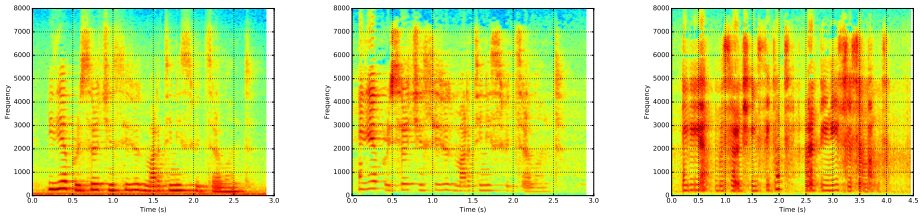

Figure 25: Spectrograms of genuine samples (left) vs. attack (middle and right) samples.

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 26).

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.

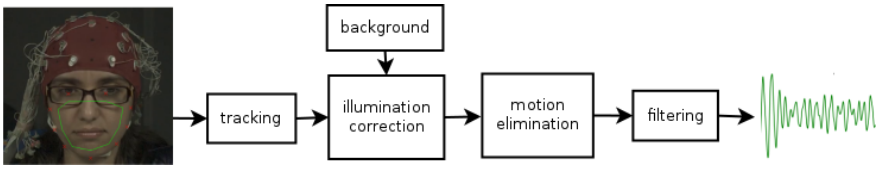

Figure 26: Illustration of remote photoplethysmography: colors from the video signal are filtered to produce an estimation of the heart beat signal.

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, wavelength between 700 and 1000 nm) illumination. The vein pattern can then be extracted with image pre-processing techniques (Figure 27) and used for recognition by any pattern recognition method.

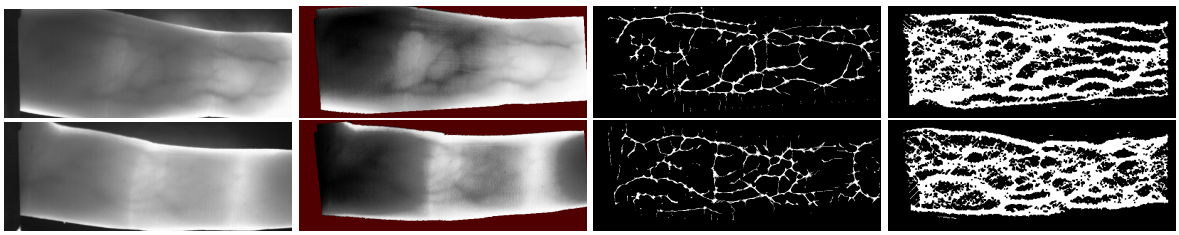

Figure 27: Illustration of image processing for vein biometrics (first row for data subject 1 and second row for data subject 2): the original image (left) is geometrically normalised (middle left) and binarised with maximum curvature (middle right) or repeated line tracking (right).

Swiss Centre for Biometrics Research and Testing

In 2014, the Idiap Research Institute launched the "Swiss Centre for Biometrics Research and Testing" (www.biometrics-center.ch), a competence centre within the Institute following recent successes in coordinating International research projects in Biometrics (MOBIO, TABULA RASA and BEAT). The aim of this centre is to serve as a legacy for these projects and to push for industry-driven research and testing in biometrics.

The centre attracted the attention of large companies (license, research and testing agreements) and led to many new projects (DARPA, IARPA, CTI). In 2018, the centre has developed over 3 directions:

- Improving the security of the BEAT platform: following an external IT security audit our engineers strengthened the platform. The platform is now used in multiple research projects such as ALLIES, LEARN-REAL and the major H2020 AI4EU project.
- Simplifying the BEAT platform: with the aim to be more attactive our engineers developed a new feature to easily edit not only biometric experiments but also more general pattern recognition and AI-based experiments.
- Training: we are teaching the UniDistance Certificate in Biometrics and Privacy (https://distanceuniversity.ch/cas-biometrics/).
- Testing: we engaged with the FIDO Alliance (https://fidoalliance.org), focused on providing open and free authentication standards to help reduce the world's reliance on passwords, to become an accredited FIDO lab able to perform certification of biometrics products. We equipped a testing enviroment and we are currently going through the accreditation.

Key publications

[1] P. Korshunov and S. Marcel, "DeepFakes: a New Threat to Face Recognition? Assessment and Detection.", *arXiv and Idiap Research Report 18-2018*, (http://publications.idiap.ch/index.php/publications/show/3988), 2018.

[2] O. Nikisins, A. Mohammadi, A. Anjos and S. Marcel, "On Effectiveness of Anomaly Detection Approaches against Unseen Presentation Attacks in Face Anti-Spoofing", *IAPR International Conference on Biometrics (ICB)*, 2018.

[3] G. Heusch and S. Marcel "Pulse-based Features for Face Presentation Attack Detection", *IEEE International Conference on Biometrics: Theory, Applications, and Systems (BTAS)*, 2018.

[4] S. Bhattacharjee, A. Mohammadi and S. Marcel, "Spoofing Deep Face Recognition With Custom Silicone Masks", *IEEE International Conference on Biometrics: Theory, Applications, and Systems (BTAS)*, 2018.

[5] V. Krivokuca and S. Marcel, "Towards Quantifying the Entropy of Fingervein Patterns across Different Feature Extractors". *IEEE International Conference on Identity, Security and Behavior Analysis (ISBA)*, 2018.

2.9 Natural Language Understanding

Overview

Head: Dr. James Henderson (BSc, Massachusetts Inst. Technology, USA, 1987; MSE & PhD, Univ. Pennsylvania, USA, 1991,1994; MER & Chargé de Cours, Univ. Geneva, 2008–2012,2012–2018)

Group overview: The Natural Language Understanding group was created in September 2017, in part as a continuation of the previous Natural Language Processing group. The NLU group studies deep learning for natural language processing tasks, focusing on models with learned representations of the meaning of text. These tasks include machine translation, information retrieval, language modelling, syntactic and semantic parsing, text classification, summarisation and natural language inference, applied to both text and speech. We focus on recurrent and attention-based neural network models, vector-space representations for entailment (rather than similarity), unsupervised learning of semantic representations, and multi-task learning.

During 2018, the NLU group had the following members: the head of the group, two postdoctoral researchers, and 5 PhD students. One PhD student graduated and three PhD students arrived.

Key scientific outputs: During 2018, the work of the NLU group has included neural machine translation, notably incorporating word sense disambiguation, using hierarchical attention to condition on previous sentences, and learning more effective representations of predicted words jointly with their input representations. In language modelling, work demonstrated how training on multiple languages can improve perplexity on low resource languages. In two new SNSF-funded projects, promising initial results were achieved in training entailment-vector sentence representations for the natural language inference task. Also, the head-of-group helped submit an SNSF NCCR grant proposal on the evolution of language, proposing biologically inspired deep learning models of syntactic processing.

Neural Network Architectures for NLP Tasks

The impressive initial results from using deep learning architectures in NLP tasks, such as for neural machine translation, are now being improved by modifying the architectures to better embody inductive biases that are important for NLP tasks. This is manifested in our work on self-attention in the decoder for neural machine translation. Instead of generating the translation with the standard LSTM decoder, we add an attention mechanism over the prefix of generated words. This better models non-local correlations in the output sentence. We find that this mechanism works best if the attention applies directly to the word embeddings, rather than the LSTM states at each word.

Another way to import inductive biases into a deep learning architecture is to explicitly model the similarity between output classes. This can be done with output embedding models, where the text of the output class label is used to compute a vector for each output class, such that similarity between classes is represented as similarity between their vectors. Previously such models have generalised well in “zero-shot” learning, where the output classes at test time do not occur at all in the training data. With our proposed use of cross-entropy loss in our attention-based output-embedding architecture for text classification, we showed for the first time that such models can also show improvement for the output classes which do occur in the training data, where previously non-output-embedding classification has performed better. This effect is particularly marked with very large label sets, where our model also trains faster, in time independent of the label set size. We have also generalised this architecture to the case of decoding in machine translation, where there are a large number of possible next output words.

Vector-Space Models of Entailment

Entailment is the relation which reflects adding information or abstracting away from information, and is fundamental to many theories of semantics. But typically vector-space models, such as word embeddings and deep learning architectures, are based on similarity between vectors, not entailment. Prior to joining Idiap, the head of group has developed a vector-space framework for modelling entailment. Unprecedented results have been achieved on the unsupervised learning of word embeddings that predict entailment between words (hyponymy), using a novel entailment-based distributional semantic model. This framework extends to entailment between larger texts (textual entailment), and has particular relevance to large-scale summarisation tasks. Textual entailment is important for summarisation because the summary must abstract away from unimportant information but still contain as much information as possible from the text.

This line of work is the basis of two grants which started in Autumn 2018. The first grant includes exploiting the entailment-vectors framework and the word embedding results to build models of the fundamental problem of textual entailment. In parallel, the project will apply textual entailment to developing models and resources for large-scale opinion summarisation. Opinion summarisation is where a large collection of texts must be summarised, including the popularity of different opinions. The second grant is for an inter-disciplinary project investigating the interpretation of official announcements. The interpretation text is a form of summary of the announcement text, so textual entailment is again relevant. But typically interpretation involves more complex reasoning and background knowledge, making it a challenging testbed.

Key publications

[1] Nikolaos Pappas & James Henderson (accepted) – GILE: A Generalized Input-Label Embedding for Text Classification. *Transactions of the Association for Computational Linguistics (TACL)*.

[2] Xiao Pu, Nikolaos Pappas, James Henderson & Andrei Popescu-Belis (2019) – Integrating Weakly Supervised Word Sense Disambiguation into Neural Machine Translation. *Transactions of the Association for Computational Linguistics (TACL)*, 6:635-649.

[3] Lesly Miculicich, Dhananjay Ram, Nikolaos Pappas & James Henderson (2018) – Document-Level Neural Machine Translation with Hierarchical Attention Networks. *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Brussels, Belgium.

[4] Nikolaos Pappas, Lesly Miculicich & James Henderson (2018) – Beyond Weight Tying: Learning Joint Input-Output Embeddings for Neural Machine Translation. *Proceedings of the Third Conference on Machine Translation (WMT)*, Brussels, Belgium.

[5] Lesly Miculicich Werlen, Nikolaos Pappas, Dhananjay Ram & Andrei Popescu-Belis (2018) – Self-Attentive Residual Decoder for Neural Machine Translation. *Proceedings of the 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL 2018)*, New Orleans, Louisiana, USA.

2.10 Biosignal Processing

Overview

Head: Dr. André Anjos (BSc, MSc & PhD, Federal University of Rio de Janeiro, Brazil, 1999, 2001, 2006; Visiting Professor, State University of Bauru, 2015; Lecturer, EPFL, 2013-)

Group overview: Biosignals are signals from living beings and their analysis to support medical or related research. This group focuses on biomedical-related areas such as the analysis of e-Health records, human-signal and imaging sensing for healthcare and similar applications. Current trends in the field show refreshed interest on the use of machine learning techniques, complementing basic signal and sequence processing, all of which are key domains of research at Idiap. It leverages on Idiap's expertise on human subject handling, data acquisition, open science and data processing.

In 2018, the Biosignal Processing Group was kick-started by the head of group.

Key scientific outputs: The head of the group transitioned from his role in the Biometrics and Security Group to this permanent position during 2018, submitting for projects via usual Swiss and European channels. The milestones for 2019 include the hiring of one intern for research in vessel segmentation for eye-fundus imaging, and extensive participation in Idiap's master AI program. Additional funds have been request for the Swiss National Science Foundation for hiring two doctoral students. The head of the group continues to lead work at Idiap for improving Reproducibility in Pattern Recognition and Machine Learning.

Reproducible Research

We've been actively looking at the reproducibility of published work and how to lower the entrance barrier of publication readers, converting them into engaged users of methods we create. We argue it is insufficient, in most cases, to only publish software leading to results if original data remains inaccessible. In particular, we note that reproducibility should imply in the following characteristics: repeatability, shareability, extensibility and stability, which is not guaranteed by most published material to date. We proposed a software suite called Bob[19] that possesses such characteristics, demonstrating its flexibility to various tasks including Biometric Person Recognition, Presentation Attack Detection, Remote Photoplethysmography, and Speech Processing. All contributions are

From another perspective, there are legitimate cases in which raw data leading to research conclusions cannot be published. Furthermore, in a growing number of use-cases, the availability of both software does not translate to an accessible reproducibility scenario. The user, for example, may not have the necessary equipment to perform the analysis. To bridge this gap, we built an open platform for research[20] in computational sciences related to pattern recognition and machine learning, to help on the development, reproducibility and certification of results obtained in the field.

Bob and BEAT are still active and support past and future work at Idiap and beyond. We conducted (and will continue doing) lectures to both master and graduate students about reproducibility in data science (refer to the applicant's CV under the section "Teaching" for details).

[19] https://www.idiap.ch/software/bob
[20] https://www.idiap.ch/software/beat

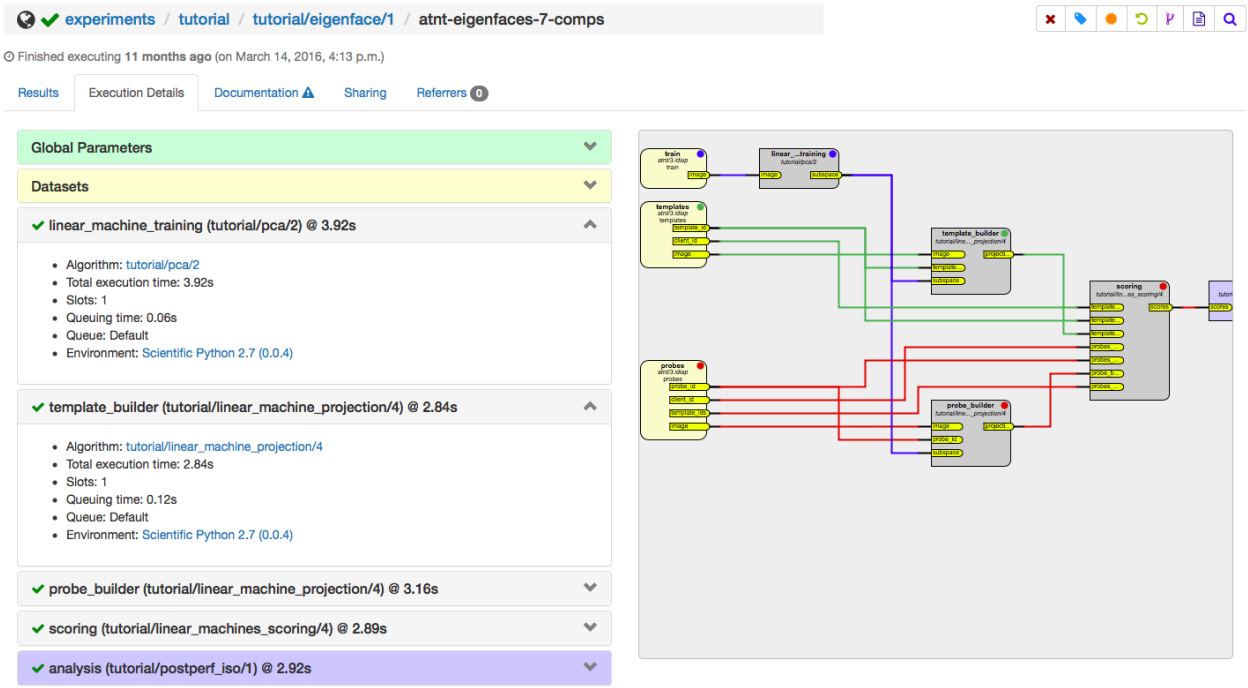

Figure 28: BEAT web user interface. Experiment view: after experiment execution, the user can check experimental details such as the results, but equally computing performance indicators. Each experiment may be certified by the platform guaranteeing its reproducibility, effectively making it read-only; Experimental results from different experiments may be combined into powerful reports that can be exported into publications. This experiment may be reviewed online at https://www.beat-eu.org/platform/experiments/tutorial/tutorial/eigenface/1/atnt-eigenfaces-7-comps/.

Key publications

[1] Olegs Nikisins, Teodors Eglitis, André Anjos and Sébastien Marcel, “Fast cross-correlation based wrist vein recognition algorithm with rotation and translation compensation", in: Sixth International Workshop on Biometrics and Forensics, 2018

[2] Olegs Nikisins, Amir Mohammadi, André Anjos, and Sébastien Marcel. “On Effectiveness of Anomaly Detection Approaches against Unseen Presentation Attacks in Face Anti-Spoofing". In: The 11th IAPR International Conference on Biometrics (ICB 2018). Feb. 2018

2.11 Energy Informatics

Overview

Head: Dr. Jérôme Kämpf (PhD, Ecole Polytechnique Fédérale de Lausanne, 2009)

Group overview: The Swiss Energy Strategy 2050 resides in three pillars: increasing the energy efficiency (including the building sector), increasing the use of renewable energy (by their promotion) and withdrawal from nuclear energy. These objectives are perfectly in-line with the Energy Informatics concepts: to exploit state-of-the-art Information and Communication Technologies to tackle global warming and climate change challenges, to increase integration of renewable and distributed energy sources by making energy systems smarter, and to increase energy efficiency beyond what improvements at component level can achieve. In that vein, the Energy Informatics Group at Idiap researches into ways of simulating energy transition pathways with intelligent control and adjustment mechanisms of evolving buildings with retrofitting and use, renewable energy production and energy storage in a changing climate.

In 2018, the Energy Informatics Group was kick-started by the head of the group.

Key scientific outputs: In September 2018, the head of the group initiated the transition between his position of Professor of Building Energy Efficiency at the HES-SO Fribourg to this permanent Senior Researcher position. The transition took the form of 20% of the time at the Idiap Research Institute initiating international collaborations for a rapid launch of the research group. Successful contacts were achieved in 2018 with the University of Pisa (Italy), University of Aveiro (Portugal) and the National University of Singapore (Singapore) for an exchange of respectively a Master and two PhD students in early 2019. Meanwhile, the head of the goup's PhD student at EPFL formed the jury for his thesis defense on "Automated Daylighting Control System based on Sky Luminance Monitoring and Lighting Computing" in April 2019. A first journal paper was published under the auspices of the Idiap Research Institute in the thematic of Machine Learning applied to Energy Savings.

Key publications

[1] Fusing TensorFlow with building energy simulation for intelligent energy management in smart cities, José Vázquez-Canteli, Stepan Ulyanin, Jérôme Kämpf and Zoltán Nagy, in: Sustainable Cities and Society, 2018

3 Researchers

3.1 Research Associates

Lastname	Firstname	Nationality	Supervisor	Start	Estimated End
BHATTACHARJEE	Sushil	CH	Sébastien Marcel	2016-01-01	2019-12-31
GEISSBUHLER	David	CH	Sébastien Marcel	2018-06-01	2019-05-31
HEUSCH	Guillaume	CH	Sébastien Marcel	2015-06-08	
IMSENG	David	CH	Hervé Bourlard	2016-04-01	2019-10-31
KORSHUNOV	Pavel	EE	Sébastien Marcel	2018-01-01	2018-12-31
LABHART	Florian	CH	Daniel Gatica-Perez	2017-11-01	2019-12-31
MADIKERI	Srikanth	IN	Petr Motlicek	2018-03-01	2019-08-01

3.2 Post-doctoral Scholars

Lastname	Firstname	Nationality	Supervisor	Start	Estimated End
ABROL	Vinayak	IN	Mathew Magimai Doss	2018-01-01	2018-10-31
AICHINGER	Ida	AT	Hervé Bourlard	2018-03-01	2018-05-31
CAO	Yuanzhouhan	CN	Jean-Marc Odobez	2017-11-15	2019-02-15
DEY	Subhadeep	IN	Petr Motlicek	2018-11-12	2019-11-12
GEORGE	Anjith	IN	Sébastien Marcel	2017-08-21	2019-03-31
HALSTEAD	Michael	AU	Sébastien Marcel	2018-09-01	2019-08-31
KHONGLAH	Banriskhem	IN	Hervé Bourlard	2018-03-01	2020-02-28
KODRASI	Ina	AL	Hervé Bourlard	2017-12-01	2019-11-30
KOMATY	Alain	FR	Sébastien Marcel	2017-05-01	2018-04-30
KOTWAL	Ketan	IN	Sébastien Marcel	2018-01-01	2018-12-31
KRIVOKUCA	Vedrana	HR	Sébastien Marcel	2017-01-15	2019-01-15
KUPCSIK	Andras	HU	Sylvain Calinon	2017-04-03	2018-08-31
LEPOITTEVIN	Yann	FR	François Fleuret	2016-08-01	2018-04-30
LIU	Gang	CN	Jean-Marc Odobez	2017-09-18	2018-12-31
MURALIDHAR	Skanda	IN	Daniel Gatica-Perez	2018-12-01	2020-05-31
NGUYEN	Laurent	CH	Daniel Gatica-Perez	2015-04-01	2019-07-31
NIKISINS	Olegs	LV	Sébastien Marcel	2014-07-01	2019-03-01
PAOLILLO	Antonio	IT	Sylvain Calinon	2018-01-15	2019-01-15
PAPPAS	Nikolaos	GR	James Henderson	2016-02-01	2020-01-31
PRASAD	Ravi	IN	Mathew Magimai Doss	2018-10-01	2019-09-30
RAZAVI	Marzieh	IR	Mathew Magimai Doss	2017-08-01	2018-07-31
REKABSAZ	Navid	IR	James Henderson	2018-03-01	2019-02-28
SAMUI	Suman	IN	Petr Motlicek	2018-09-01	2019-08-31
SARFJOO	Saeed	IR	Sébastien Marcel	2018-01-01	2018-12-31
SHAHNAWAZUDDIN	Syed	IN	Petr Motlicek	2018-06-08	2018-08-31
SHAKAS	Alexis	CY	Philip Garner	2018-01-22	2018-06-29
SRINIVASAMURTHY	Ajay	IN	Petr Motlicek	2017-09-01	2018-02-28
VILLAMIZAR	Michael	ES	Jean-Marc Odobez	2016-11-01	2019-10-31
VLASENKO	Bogdan	DE	Daniel Gatica-Perez	2017-07-01	2018-11-25

3.3 PhD Students

Lastname	Firstname	Nationality	Supervisor	Start	Estimated End
AMINIAN	Bozorgmehr	CH	Jean-Marc Odobez	2018-04-01	2022-03-31
CAN	Gulcan	TR	Daniel Gatica-Perez	2013-09-15	2018-01-31
CHAVDAROVA	Tatjana	MK	François Fleuret	2014-12-01	2019-11-30
DE FREITAS PEREIRA	Tiago	BR	Sébastien Marcel	2014-07-01	2019-02-28
DESPRÈS	Nicolas	FR	Jean-Marc Odobez	2018-02-01	2019-04-30
DEY	Subhadeep	IN	Petr Motlicek	2014-04-01	2018-11-11
DIGHE	Pranay	IN	Hervé Bourlard	2014-08-01	2018-11-30
DUBAGUNTA	Pavankumar	IN	Mathew Magimai Doss	2017-04-01	2021-03-31
FRITSCH	Julian	DE	Mathew Magimai Doss	2018-07-01	2022-06-30
GAUTIER	Athenais	FR	David Ginsbourger	2018-11-01	2022-10-31
GIRGIN	Hakan	TR	Sylvain Calinon	2018-09-01	2022-08-31
HE	Weipeng	CN	Jean-Marc Odobez	2016-06-01	2020-05-31
HERMANN	Enno	DE	Mathew Magimai Doss	2018-07-01	2022-06-30
JANBAKHSHI	Parvaneh	IR	Hervé Bourlard	2018-02-01	2022-01-31
JAQUES	Christian	CH	Michael Liebling	2016-04-01	2020-03-31
JAQUIER	Noémie	CH	Sylvain Calinon	2016-08-02	2020-08-01
JOSE	Cijo	IN	François Fleuret	2018-01-01	2018-05-31
KABIL	Selen	TR	Hervé Bourlard	2018-05-01	2022-04-30
KARIMI MAHABADI	Rabeeh	IR	James Henderson	2018-10-01	2022-09-30
KATHAROPOULOS	Angelos	GR	François Fleuret	2017-03-01	2021-02-28
KULAK	Thibaut	FR	Sylvain Calinon	2017-12-01	2021-11-30
LE	Nam	VN	Jean-Marc Odobez	2015-01-15	2019-01-15
LEMBONO	Teguh	ID	Sylvain Calinon	2018-07-23	2022-07-23
MAI	Florian	DE	James Henderson	2018-10-01	2022-09-30
MARELLI	François	BE	Michael Liebling	2018-10-01	2022-09-30
MARFURT	Andreas	CH	James Henderson	2018-11-01	2022-10-31
MARIANI	Olivia	CH	Michael Liebling	2016-06-01	2020-05-31
MARTINEZ GONZALEZ	Angel	MX	Jean-Marc Odobez	2016-11-15	2020-11-15
MICULICICH	Lesly	PE	James Henderson	2016-03-01	2020-02-29
MOHAMMADI	Amir	IR	Sébastien Marcel	2016-01-15	2020-01-15
MUCKENHIRN	Hannah	FR	Mathew Magimai Doss	2015-07-13	2019-07-12
MURALIDHAR	Skanda	IN	Daniel Gatica-Perez	2014-12-01	2018-11-30
NEWLING	James	GB	François Fleuret	2017-09-16	2018-02-28
PHAN	Trung	VN	Daniel Gatica-Perez	2015-09-01	2019-12-31
PIGNAT	Emmanuel	CH	Sylvain Calinon	2016-03-01	2020-02-29
PU	Xiao	CN	James Henderson	2014-08-01	2018-07-31
RAM	Dhananjay	IN	Hervé Bourlard	2015-01-01	2018-12-31
SCHNELL	Bastian	DE	Philip Garner	2017-05-01	2021-04-30
SEBASTIAN	Jilt	IN	Mathew Magimai Doss	2017-09-01	2018-08-31
SHAJKOFCI	Adrian	CH	Michael Liebling	2016-09-20	2020-09-20
SIEGFRIED	Rémy	CH	Jean-Marc Odobez	2017-01-01	2020-12-31
SIVAPRASAD	Prabhu	IN	François Fleuret	2018-11-01	2022-10-31
SRINIVAS	Suraj	IN	François Fleuret	2017-03-01	2021-02-28

TANWANI	Ajay	PK	Sylvain Calinon	2015-01-12	2018-04-13
TONG	Sibo	CN	Philip Garner	2016-05-01	2020-04-30
TORNAY	Sandrine	CH	Mathew Magimai Doss	2016-08-01	2020-07-31
TRAVELLETTI	Cédric	CH	David Ginsbourger	2018-11-01	2022-10-31
VYAS	Apoorv	IN	Hervé Bourlard	2018-07-15	2022-07-15
YU	Yu	CN	Jean-Marc Odobez	2015-07-15	2019-07-15
ZHAN	Qingran	CN	Petr Motlicek	2018-10-01	2020-03-31

3.4 Interns

Lastname	Firstname	Nationality	Supervisor	Start	Estimated End
BOURGEOIS	Dylan	FR	Sylvain Calinon	2017-12-23	2018-01-26
CANDY	Romain	FR	Hervé Bourlard	2018-03-01	2018-08-31
COURDIER	Evann	FR	François Fleuret	2018-10-22	2019-02-28
GIRGIN	Hakan	TR	Sylvain Calinon	2018-03-01	2018-08-31
JEANNINGROS	Loïc	FR	Philip Garner	2018-11-01	2019-04-30
KABIL	Selen	TR	Hervé Bourlard	2017-09-19	2018-04-30
LOUPI	Dimitra	GR	Philip Garner	2018-02-26	2018-08-24
MA	Wei	CN	James Henderson	2017-09-19	2018-01-19
MEDINA RIOS	Luis Emmanuel	MX	Daniel Gatica-Perez	2018-07-23	2018-08-31
MOSTAANI	Zohreh	IR	Sébastien Marcel	2017-10-02	2019-04-09
PATEL	Kumar	IN	James Henderson	2018-10-12	2019-02-28
ROSSIER	Alain	CH	François Fleuret	2017-07-17	2018-01-17
SHARMA	Shivam	IN	Mathew Magimai Doss	2018-10-01	2019-03-31
STANDAERT	Florian	CH	Sylvain Calinon	2017-12-23	2018-01-26
STERPU	George	RO	Petr Motlicek	2018-02-05	2018-08-04
SUOMALAINEN	Markku	FI	Sylvain Calinon	2018-01-29	2018-06-15
THOMAS	Chinchu	IN	Daniel Gatica-Perez	2018-08-15	2018-11-15
VUILLEUMIIER	Noémie	CH	Daniel Gatica-Perez	2018-04-01	2018-05-31

3.5 Visitors

Lastname	Firstname	Nationality	Supervisor	Start	Estimated End
CARRERA SOTO	Laura	CU	David Ginsbourger	2017-11-10	2018-05-31
HU	Rui	CN	Daniel Gatica-Perez	2017-12-01	2018-05-31
MAALOUF	Eliane	LB	David Ginsbourger	2018-11-01	2021-11-30
SCHÄRER	Cedric	CH	David Ginsbourger	2018-03-08	2018-08-31

4 Active and Granted Projects in 2018

An overview of the projects that have been active during the year 2018 is presented in Section 4.1. The projects are grouped by funding agencies, starting with European and International Agencies, then Swiss Agencies to finish with the projects funded by industrial partners.

Section 4.2 presents the list of projects accepted during 2018 but starting in the following year.

4.1 Projects in Progress during 2018

4.1.1 Projects Funded by Swiss Agencies

1. Name **3DFINGERVEIN** (3D FingerVein Biometrics)
   Funding Innosuisse
   Coordinator Hes-So Valais
   Duration 2016.06.01 - 2018.03.31
   Partner(s) Idiap Research Institute, Itservices Sonna Sàrl

2. Name **ALLIES** (Autonomous Lifelong learnIng intelLigent Systems)
   Funding SNF -ERA-NET
   Coordinator Idiap Research Institute
   Duration 2017.12.01 - 2020.11.30
   Partner(s) Laboratoire national de métrologie et d'essais, Université du Maine, Universitat Politecnica de Catalunya

3. Name **COMETS-M** (Computational Methods for Temporal Super-resolution Microscopy)
   Funding SNF - Division II
   Coordinator Idiap Research Institute
   Duration 2016.04.01 - 2019.03.30
   Partner(s) -

4. Name **COMPBIO** (Computational biomicroscopy: advanced image processing methods to quantify live biological systems)
   Funding SNF - Division II
   Coordinator Idiap Research Institute
   Duration 2018.04.01 - 2022.03.31
   Partner(s) -

5. Name **DCROWDLENS** (Crowdsourcing the Assessment of Deep Visual Explanations for Subjective Variables)
   Funding Hasler Foundation
   Coordinator Idiap Research Institute
   Duration 2017.05.15 - 2018.01.15
   Partner(s) -

6. Name **DEEPCHARISMA** (Deep Learning Charisma)
   Funding UNIL Collaboration

Coordinator	Idiap Research Institute
Duration	2017.04.01 - 2018.08.31
Partner(s)	-

7. Name **DEVEL-IA** (Formation « Développeurs spécialisés en Intelligence Artificielle » selon le modèle de formation continue duale postgrade)
   Funding SECO
   Coordinator Idiap Research Institute
   Duration 2018.05.01 - 2019.06.30
   Partner(s) Formation universitaire à Distance

8. Name **DOMAT** (On-demand Knowledge for Document-level Machine Translation)
   Funding SNF - Division II
   Coordinator Idiap Research Institute
   Duration 2018.01.01 - 2022.09.30
   Partner(s) HES-SO Vaud

9. Name **DUSK2DAWN** (Characterizing Youth Nightlife Spaces, Activities, and Drinks)
   Funding SNF - Sinergia
   Coordinator Idiap Research Institute
   Duration 2017.07.01 - 2019.12.31
   Partner(s) University of Zurich, La Trobe University

10. Name **ESGEM** (Enhanced Swiss German mEdia Monitoring)
    Funding Innosuisse
    Coordinator Idiap Research Institute
    Duration 2016.06.01 - 2018.01.31
    Partner(s) Recapp, Argus der Presse AG

11. Name **FARGO** (Convenient and Secure 3D Face Recognition based on RGB-D Cameras)
    Funding Innosuisse
    Coordinator Idiap Research Institute
    Duration 2016.05.01 - 2018.02.28
    Partner(s) Keylemon SA

12. Name **FLOSS** (Flexible Linguistically-guided Objective Speech aSessment)
    Funding Hasler Foundation
    Coordinator Idiap Research Institute
    Duration 2017.03.01 - 2020.02.29
    Partner(s) -

13. Name **GPUS** (Acquisition d'un cluster de calcul dédié à l'Intelligence Artificielle)
    Funding Loterie Romande
    Coordinator Idiap Research Institute
    Duration 2017.09.01 - 2018.08.31
    Partner(s) -

14. Name **HFACE** (Heterogeneous Face Recognition)
    Funding SNF - Division II
    Coordinator Idiap Research Institute
    Duration 2014.07.01 - 2018.06.30
    Partner(s) -

15. Name **I-DRESS** (Assistive Interactive robotic system for support in DRESSing)
    Funding SNF - ERA NET
    Coordinator Idiap Research Institute
    Duration 2015.12.01 - 2019.02.28
    Partner(s) -

16. Name **IMIM** (Intelligent Monitoring for In-line Manufacturing)
    Funding Innosuisse
    Coordinator AiSA Automation Industrielle SA
    Duration 2016.07.01 - 2018.05.31
    Partner(s) Idiap Research Institute

17. Name **IPEQ** (Uncertainty quantification and efficient design of experiments for data- and simulation-driven inverse problem solving)
    Funding SNF - Division II
    Coordinator Idiap Research Institute
    Duration 2018.11.01 - 2022.10.31
    Partner(s) University of California at Davis

18. Name **ISUL** (Importance sampling for large-scale unsupervised learning)
    Funding SNF - Division II
    Coordinator Idiap Research Institute
    Duration 2017.03.01 - 2020.02.28
    Partner(s) -

19. Name **LAB_FIDO** (Création d'un laboratoire de test biométrique agréé)
    Funding Loterie Romande
    Coordinator Idiap Research Institute
    Duration 2018.12.01 - 2019.11.30
    Partner(s) -

20. Name **LAOS** (Learning Representations of Abstraction in Text)
    Funding SNF - Division II
    Coordinator Idiap Research Institute
    Duration 2018.11.01 - 2022.10.31
    Partner(s) -

21. Name **LIVEHEART** (The Cellular Basis of Cardiac Development Revealead by Live Imaging)
    Funding SNF - ANR
    Coordinator Institut de Génétique et de Biologie Moléculaire et Cellulaire
    Duration 2016.06.01 - 2019.05.31
    Partner(s) Idiap Research Institute, University of Bern, École Polytechnique Paris

22. Name **MASS** (Multilingual Affective Speech Synthesis)
    Funding SNF - Division II
    Coordinator Idiap Research Institute
    Duration 2017.05.01 - 2020.04.30
    Partner(s) -

23. Name **MEMUDE** (Multi-view Detection with Metric-learning for Deep Network Fusion)
    Funding Hasler Foundation
    Coordinator Idiap Research Institute
    Duration 2017.06.01 - 2019.05.31
    Partner(s) -

24. Name **MOSPEEDI** (Motor Speech Disorders: characterizing phonetic speech planning and motor speech programming/execution and their impairments)
    Funding SNF - Sinergia
    Coordinator University of Geneva
    Duration 2017.10.01 - 2020.09.30
    Partner(s) Idiap Research Institute, University Hospitals of Geneva, Université Paris 3

25. Name **MPM** (Multimodal People Monitoring)
    Funding Idiap-CSEM Program
    Coordinator Idiap Research Institute
    Duration 2018.02.01 - 2019.01.31
    Partner(s) Centre Suisse d'Electronique et de Microtechnique

26. Name **MULTI_CROWD** (Tracing Swiss Heritage Speakers' Identities in North America)
    Funding EPFL Collaboration
    Coordinator University of Lausanne
    Duration 2017.02.01 - 2018.02.28
    Partner(s) Idiap Research Institute

27. Name **MULTIVEO** (High Accuracy Speaker-Independent Multilingual Automatic Speech Recognition System)
    Funding Innosuisse
    Coordinator Idiap Research Institute
    Duration 2013.11.01 - 2018.03.31
    Partner(s) Veovox

28. Name **ODESSA** (Online Diarization Enhanced by recent Speaker identification and Sequential learning Approaches)
    Funding SNF - ANR
    Coordinator Centre National de La Recherche Scientifique
    Duration 2016.03.01 - 2019.10.31
    Partner(s) Idiap Research Institute, Eurecom

29. Name **PHASER-QUAD** (Parsimonious Hierarchical Automatic Speech Recognition and Query Detection)
    Funding SNF - Division II
    Coordinator Idiap Research Institute
    Duration 2016.10.01 - 2019.09.30
    Partner(s) -

30. Name **REAPPS** (Reinforced audio processing via physiological signals)
    Funding Idiap-CSEM Program
    Coordinator Idiap Research Institute
    Duration 2018.03.01 - 2019.10.31
    Partner(s) Centre Suisse d'Electronique et de Microtechnique

31. Name **REGENN** (Robust Eye-Gaze Estimation Deep Neural Network)
    Funding Innosuisse
    Coordinator Idiap Research Institute
    Duration 2017.09.01 - 2018.12.31
    Partner(s) Eyeware

32. Name **RISE** (Rich Interpersonal Skill analytics for rEcruitment)
    Funding Innosuisse
    Coordinator Idiap Research Institute
    Duration 2018.05.01 - 2020.04.30
    Partner(s) University of Lausanne

33. Name **ROSALIS** (Robot skills acquisition through active learning and social interaction strategies)
    Funding SNF - Division II
    Coordinator Idiap Research Institute
    Duration 2018.04.01 - 2022.03.31
    Partner(s) -

34. Name **SHAPED** (Speech Hybrid Analytics Platform for consumer and Enterprise Devices)
    Funding Innosuisse
    Coordinator Idiap Research Institute
    Duration 2018.03.01 - 2020.02.29
    Partner(s) Logitech S.A.

35. Name **SHISSM** (Sparse and hierarchical Structures for Speech Modeling)
    Funding SNF - Division II
    Coordinator Idiap Research Institute
    Duration 2018.01.01 - 2021.12.31
    Partner(s) -

36. Name **SM2** (Extracting Semantic Meaning from Spoken Material)
    Funding Innosuisse
    Coordinator Idiap Research Institute
    Duration 2018.11.01 - 2020.04.30
    Partner(s) Recapp

37. Name **SMILE** (Scalable Multimodal sign language Technology for sIgn language Learning and assessmEnt)
    Funding SNF - Sinergia
    Coordinator Idiap Research Institute
    Duration 2016.03.01 - 2020.02.28
    Partner(s) University of Surrey, University of Applied Sciences of Special Needs Education

38. Name **SWISKO** (Swiss-Korean project to develop and integrate new wearable sensors into the existing DomoSafety ambient sensor system.)
    Funding Innosuisse
    Coordinator Idiap Research Institute
    Duration 2016.01.01 - 2018.02.28
    Partner(s) DomoSafety S.A., University of Bern, Hes-So Vaud

39. Name **TACT-HAND** (Improving control of prosthetic hands using tactile sensors and realistic machine learning)
    Funding SNF - DACH
    Coordinator Deutsches Zentrum Fuer Luft - Und Raumfahrt Ev
    Duration 2016.04.01 - 2020.01.31
    Partner(s) Idiap Research Institute, Universitaet Bielefeld

40. Name **UNICITY** (3D scene understanding through machine learning to secure entrance zones)
    Funding Innosuisse
    Coordinator Idiap Research Institute
    Duration 2017.03.01 - 2018.11.30
    Partner(s) Hes-So Fribourg, Fastcom Technology SA

41. Name **UNITS** (Unified Speech Processing Framework for Trustworthy Speaker Recognition)
    Funding SNF - Division II
    Coordinator Idiap Research Institute
    Duration 2015.07.01 - 2018.06.30
    Partner(s) -

42. Name **VIEW-2** (Visibility Improvement for Events Webcasting)
    Funding Innosuisse
    Coordinator Hes-So Valais
    Duration 2016.08.01 - 2018.03.31
    Partner(s) Idiap Research Institute, Klewel

43. Name **VNV** (Automatic analysis of verbal and non-verbal behavior and provision of feedback in video selection interviews)
    Funding SNF - Division I
    Coordinator University of Neuchâtel
    Duration 2018.10.01 - 2020.03.31
    Partner(s) Idiap Research Institute, University of Lausanne

4.1.2 Projects Funded by European and International Agencies

1. Name **4DHEART** (4D analysis of heart development and regeneration using advanced light microscopy)
   Funding H2020-MSCA
   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 Europeen de Recherche En Biologie et Medecine

2. Name **BATL** (Biometric Authentification with Timeless Learner)
   Funding USA IARPA
   Coordinator University of Southern California
   Duration 2017.03.01 - 2021.02.28
   Partner(s) Idiap Research Institute

3. Name **COLLABORATE** (Co-production CeLL performing Human-Robot Collaborative AssEmbly)
   Funding H2020-RIA-DT
   Coordinator Aristotle University Of Thessaloniki
   Duration 2018.10.01 - 2021.09.30
   Partner(s) Idiap Research Institute, Arcelik A.S., Association pour la Recherche et le Développement des méthodes et processes, ASTI Mobile Robotics, Blue Ocean Robotics APS, Centre for Research and Technology Hellas, Centro Ricerche Fiat SCPA, Jozef Stefan Institute, Katholieke Universiteit Leuven, Kolektor, University of Patras, Pratt & Whitney Rzeszów, Universita Degli Studi di Genova

4. 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.

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

6. Name **MALORCA** (Machine Learning of Speech Recognition Models for Controller Assistance)
   Funding H2020-RIA-SESAR
   Coordinator Deutsches Zentrum Fuer 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 Osterreichische Gesellschaft Fur Zivilluftfahrt Mbh

7. Name **MEMMO** (Memory of Motion)
   Funding H2020-RIA-ICT
   Coordinator Centre national de la recherche scientifique
   Duration 2018.01.01 - 2021.12.31
   Partner(s) Idiap Research Institute, University of Edinburgh, Max Planck Society for the Advancement of Sciences, University of Oxford, PAL ROBOTICS SL, AIRBUS SAS, Wandercraft, Centre de médecine physique et de réadaptation, Costain Group PLC

8. 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, Aldebaran Robotics, Teknologian Tutkimuskeskus Vtt, Kiinteistö Oy Ideapark Ab

9. Name **SARAL** (Summarization and domain-Adaptive Retrieval of Information Across Languages)
   Funding USA IARPA
   Coordinator University of Southern California
   Duration 2017.10.01 - 2020.07.31
   Partner(s) Idiap Research Institute, Massachusetts Institute of Technology, Raytheon Company, Reenselaer Polytechnic Institute, University of Massachusetts Amherst, Northeastern University

10. Name **SAVI** (Spotting Audio-Visual Inconsistencies)
    Funding USA DARPA
    Coordinator Sri International
    Duration 2016.05.19 - 2020.05.18
    Partner(s) Idiap Research Institute

11. Name **SIIP** (Speaker Identification Integrated Project)
    Funding FP7
    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 of Warwick, Laboratorio Di Scienze della Cittadinanza, The International Criminal Police Organization, Police Service of Northern Ireland, Ministério Da Justiça, Lisboa

12. 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

13. Name **SWAN** (Secure Access Control over Wide Area Network)
    Funding Research Council of Norway
    Coordinator Hogskolen I Gjovik
    Duration 2016.01.01 - 2019.12.31
    Partner(s) Idiap Research Institute, Morpho, Bankenverband, Universitetet I Oslo, Zwipe As

14. Name **TAPAS** (Training Network on Automatic Processing of PAthological Speech)
    Funding H2020-MSCA
    Coordinator Idiap Research Institute
    Duration 2017.11.01 - 2021.10.31
    Partner(s) University of Sheffield, Philips, Radboud University Nijmegen - Stichting Katholieke Universiteit, Ludwig-Maximilians-Universität München, Institut de Recherche En Informatique de Toulouse, Antwerpen University Hospital, Friedrich-Alexander-Universität Erlangen Nuernberg, Instituto de Engenharia de Sistemas E Computadores, Investigacao E desenvolvimento Em Lisboa, Interuniversitair Micro-Electronica Centrum Imec Vzw, Stichting Het Nederlands Kanker Instituut - Antoni Van Leeuwenhoek Ziekenhuis, Universitaet Augsburg

15. 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 - 2019.03.31
    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

4.1.3 Projects Funded by Industrial Partners

1. Name **AADES** (Adaptive and Asynchronous Detection and Segmentation)
   Funding Armasuisse
   Coordinator Idiap Research Institute
   Duration 2018.10.01 - 2019.09.30
   Partner(s) -

2. Name **ADATIS** (Adatis - Optimisation du Contrôle Qualité)
   Funding Fondation The Ark
   Coordinator Adatis
   Duration 2018.10.01 - 2019.05.31
   Partner(s) Idiap Research Institute

3. Name **CAMPRO** (Camera de profondeur)
   Funding Fondation The Ark
   Coordinator Idiap Research Institute
   Duration 2017.08.01 - 2018.01.31
   Partner(s) Digit Arena

4. Name **COBALT** (Content Based Call Filtering)
   Funding Fondation The Ark
   Coordinator Idiap Research Institute
   Duration 2017.07.01 - 2018.06.30
   Partner(s) Katia SA

5. 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

6. Name **RECOMEDIND** (RecoMed Industrialisation)
   Funding Fondation The Ark
   Coordinator Recapp
   Duration 2017.04.01 - 2018.12.31
   Partner(s) Idiap Research Institute, Clinique Romande de Réadaptation

7. Name **SNACK** (Bites'n'Bits: Understanding Eating Routines in Context)
   Funding Industrial Project
   Coordinator École Polytechnique Fédérale de Lausanne
   Duration 2016.03.01 - 2018.05.31
   Partner(s) Idiap Research Institute

4.2 Projects Awarded in 2018 and Starting in the following year

1. Name **ADVANCE** (Augmented dialogue tool based on verbal and non-verbal behavior computing)
   Funding Innosuisse
   Coordinator Idiap Research Institute
   Duration 2019.03.01 - 2020.08.31
   Partner(s) CM Profiling, HES-SO Fribourg

2. Name **AI4EU** (A European AI On Demand Platform and Ecosystem)
   Funding H2020-RIA-ICT
   Coordinator Thales Services SA
   Duration 2019.01.01 - 2021.12.31
   Partner(s) Idiap Research Institute, ABB AS, AGI Research SRO, Allianz SE, Atos Spain SA, Aristotle University of Thessaloniki, Blumorpho SAS, Budapest University of Technology and Economics, Bureau de Recherches Géologiques et Minières, Barcelona Supercomputing Center, CARTIF Foundation, Commissariat à l'énergie atomique et aux énergies alternatives, CINECA - Consorzio Interuniversitario, Consiglio Nationale delle Ricerche, Centre National de la Recherche Scientifique, Agencia Estatal Consejo Superior De Investigaciones Cientifícas, National Center for Sicentific Research "Demokritos", German Research Center for Artificial Intelligence, German Aerospace Center, EIT Digital, Eötvös Loránd University, European Organisation for Security, FundingBox Research, Fondazione Bruno Kessler, Fraunhofer Gesellschaft, France Digitale, Foundation for Research and Technology – Hellas, Forum Virium Helsinki, Grassroots Arts and Research, France Hub IA, Industrial Data Space e. V., Institut Mines-Télécom-IMT, Institut National de Recherche en Informatique et Automatique, Associacao Do Instituto Superior Tecnico Para A Investigacao E Desenvolvimento, Centre for Research and Technology Hellas, Jožef Stefan Institute, Karlsruhe Institute of Technology, Know-Center Gmbh Research Center for Datadriven Business & Big data Analytics, University of Leeds, Loupe 16, Università degli studi di Roma "La Sapienza", Norwegian University of Science and Technology, National University of Ireland Galway, Office National d'Etudes et Recherches Aérospatiales, Orange SA, Örebro University, PG WConsulting, Université Paris I Panthéon-Sorbonne, QWANT, Siemens AG, SAP SE, Smile, Smart Rural, Simula Research Laboratory, Thales Alenia Space, Thomson Licensing, Telenor ASA, Tilde SIA, Thales Research & Technology France, Technische Universität Berlin, Delft University of Technology, Centre for Intelligent Technologies, Technical University of Munich, Technical University of Vienna, Alma Mater Studiorum – University of Bologna, University College Cork, University of Coimbra, Université Grenoble Alpes, Unilever UK, National and Kapodistrian University of Athens, Universitat Politecnica de Catalunya, Technical University of Madrid, Universita Degli Studi Di Siena, Sorbonne University, Ca' Foscari University of Venice, Vrije Universiteit Brussel, Women in AI, Wavestone

3. Name **BIOWAVE-2** (BIOWAVE 2.0, the new Biowave generation)
   Funding Innosuisse
   Coordinator Idiap Research Institute
   Duration 2019.01.01 - 2020.12.31
   Partner(s) Biowatch

4. Name **BOAT** (Automated Braces generation for Orthopaedic Anatomical Treatment of fractures)
   Funding Innosuisse
   Coordinator Idiap Research Institute
   Duration 2019.04.05 - 2020.08.04
   Partner(s) HES-SO Fribourg, Swibrace SA

5. Name **HEAP** (Human-Guided Learning and Benchmarking of Robotic Heap Sorting)
   Funding SNF - ERA NET
   Coordinator University of Lincoln
   Duration 2019.04.01 - 2022.03.31
   Partner(s) Idiap Research Institute, Istituto Italiano di Tecnologia, Institut de Recherche en Informatique et en Automatique, Technische Universitaet Wien

6. Name **INTREPID** (Automated interpretation of political and economic policy documents: Machine learning using semantic and syntactic information)
   Funding SNF - Sinergia
   Coordinator Graduate Institute of International and Development Studies
   Duration 2019.01.01 - 2022.12.31
   Partner(s) Idiap Research Institute

7. Name **LEARN-REAL** (LEARNing physical manipulation skills with simulators using REAListic variations)
   Funding SNF - ERA NET
   Coordinator Idiap Research Institute
   Duration 2019.04.01 - 2022.03.31
   Partner(s) Ecole Centrale de Lyon, Istituto Italiano di Tecnologia

8. Name **LOIS** (Leveraging on-device smartphone inference to address resistance to participate in social surveys)
   Funding UNIL/EPFL
   Coordinator University of Lausanne
   Duration 2019.01.01 - 2019.12.31
   Partner(s) Idiap Research Institute, École Polytechnique Fédérale de Lausanne

9. Name **VERIFAKE** (Vulnerability assessment and detection of Deepfake videos)
   Funding Hasler Foundation
   Coordinator Idiap Research Institute
   Duration 2019.06.01 - 2020.05.31
   Partner(s) -

10. Name **WENET** (The Internet of US)
    Funding H2020-RIA-FETPROACT
    Coordinator University of Trento
    Duration 2019.01.01 - 2022.12.31
    Partner(s) Idiap Research Institute, Aalborg University, Amrita Vishwa Vidyapeetham, Ben-Gurion University of the Negev, University of Tübingen, Instituto Potosino de Investigacion Cientifica y Tecnologica, Jilin University, London School of Economics and Political Science, Martel GmbH, National University of Mongolia, Open University of Cyprus, U-Hopper SRL, Universidad Catolica Nuestra Senora de La Ascuncion

5 List of Publications in 2018

5.1 Books and Book Chapters

[1] S. Calinon, "Robot learning with task-parameterized generative models," in *Robotics Research*, A. Bicchi and W. Burgard, Eds., vol. 2, Springer, 2018, pp. 111–126.

[2] S. Calinon and D. Lee, "Learning control," in *Humanoid Robotics: A Reference*, P. Vadakkepat and A. Goswami, Eds., Springer, 2018.

[3] 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.), 2018.

[4] D. Ginsbourger, "Sequential design of computer experiments," in *Wiley StatsRef: Statistics Reference Online*, Accepted, Wiley, 2018.

[5] P. Korshunov and S. Marcel, "A cross-database study of voice presentation attack detection," in *Handbook of Biometric Anti-Spoofing: Presentation Attack Detection, 2nd Edition*, S. Marcel, M. Nixon, J. Fierrez, and N. Evans, Eds., Springer, Nov. 2018, ch. 19.

5.2 Articles in Scientific Journals

[1] D. Azzimonti and D. Ginsbourger, "Estimating orthant probabilities of high dimensional gaussian vectors with an application to set estimation," *Journal of Computational and Graphical Statistics*, vol. 27, no. 2, pp. 255–267, 2018.

[2] D. Azzimonti, D. Ginsbourger, J. Rohmer, and D. Idier, "Profile extrema for visualizing and quantifying uncertainties on excursion regions. application to coastal flooding," *Technometrics*, 2018.

[3] J. Bect, F. Bachoc, and D. Ginsbourger, "A supermartingale approach to gaussian process based sequential design of experiments," *Bernoulli*, 2018.

[4] A. Birk, T. Fromm, C. A. Mueller, T. Luczynski, A. Gomez Chavez, D. Koehntopp, A. Kupcsik, S. Calinon, A. K. Tanwani, G. Antonelli, P. Di Lillo, E. Simetti, G. Casalino, G. Indiveri, L. Ostuni, A. Turetta, A. Caffaz, P. Weiss, T. Gobert, B. Chemisky, J. Gancet, T. Siedel, S. Govindaraj, X. Martinez, and P. Letier, "Dexterous underwater manipulation from distant onshore locations," *IEEE Robotics and Automation Magazine*, 2018.

[5] G. Can, J.-M. Odobez, and D. Gatica-Perez, "How to tell ancient signs apart? recognizing and visualizing maya glyphs with cnns," *ACM Journal on Computing and Cultural Heritage (JOCCH)*, May 2018.

[6] G. Can, J.-M. Odobez, and D. Gatica-Perez, "Maya codical glyph segmentation: A crowdsourcing approach," *IEEE Transactions on Multimedia*, vol. 20, no. 3, pp. 711–725, Mar. 2018, published online.

[7] M. Cernak, A. Asaei, and A. Hyafil, "Cognitive speech coding: Examining the impact of cognitive speech processing on speech compression," *IEEE Signal Processing Magazine*, vol. 35, no. 3, pp. 97–109, Apr. 2018.

[8] F. Ficuciello, P. Falco, and S. Calinon, "A brief survey on the role of dimensionality reduction in manipulation learning and control," *IEEE Robotics and Automation Letters (RA-L)*, vol. 3, no. 3, pp. 2608–2615, 2018.

[9] A. Giusti, M. Zeestraten, E. Icer, A. Pereira, D. G. Caldwell, S. Calinon, and M. Althoff, "Flexible automation driven by demonstration: Leveraging strategies that simplify robotics," *IEEE Robotics and Automation Magazine (RAM)*, vol. 25, no. 2, pp. 18–27, Jun. 2018.

[10] I. Havoutis and S. Calinon, “Learning from demonstration for semi-autonomous teleoperation,” *Autonomous Robots*, 2018.

[11] P.-E. Honnet, B. Gerazov, A. Gjoreski, and P. N. Garner, “Intonation modelling using a muscle model and perceptually weighted matching pursuit,” *Speech Communication*, 2018.

[12] N. Jaquier and S. Calinon, “Improving the control of prosthetic hands with tactile sensing,” *Micro & Nano Magazine, Micronarc*, pp. 42–43, 2018.

[13] I. Kodrasi and S. Doclo, “Analysis of eigenvalue decomposition-based late reverberation power spectral density estimation,” *IEEE Transaction on Acoustics, Speech and Language Processing*, vol. 26, no. 6, pp. 1106–1118, Jun. 2018.

[14] F. Labhart, E. Kuntsche, and R. Engels, “What reminds young people that they drank more than intended on weekend nights: An event-level study,” *Journal of Studies on Alcohol and Drugs*, vol. 79, no. 4, pp. 644–648, Jul. 2018.

[15] F. Labhart, E. Kuntsche, M. Livingston, and R. Engels, “After how many drinks does someone experience acute consequences-determining thresholds for binge drinking based on two event-level studies: Optimal thresholds for binge drinking,” *Addiction*, vol. 113, no. 12, pp. 2235–2244, Dec. 2018.

[16] N. Le and J.-M. Odobez, “Improving speech embedding using crossmodal transfer learning with audio-visual data,” *Multimedia Tools and Applications*, pp. 1–24, Dec. 2018.

[17] E. Madonna, D. Ginsbourger, and O. Martius, “A poisson regression approach to model monthly hail occurrence in northern switzerland using large-scale environmental variables,” *Atmospheric Research*, vol. 203, pp. 261–274, May 2018.

[18] S. Marmin, D. Ginsbourger, J. Baccou, and J. Liandrat, “Warped gaussian processes and derivative-based sequential design for functions with heterogeneous variations,” *SIAM/ASA Journal on Uncertainty Quantification*, vol. 6, no. 3, pp. 991–1018, 2018.

[19] S. Muralidhar, M. Schmid Mast, and D. Gatica-Perez, “A tale of two interactions: Inferring performance in hospitality encounters from cross-situation social sensing,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 129, 2018.

[20] L. S. Nguyen, S. Ruiz-Correa, M. Schmid Mast, and D. Gatica-Perez, “Check out this place: Inferring ambiance from airbnb photos,” *IEEE transactions on Multimedia*, vol. 20, no. 6, pp. 1499–1511, Jun. 2018.

[21] S. Okada, L. S. Nguyen, O. Aran, and D. Gatica-Perez, “Modeling dyadic and group impressions with inter-modal and inter-person features,” *ACM Transactions on Multimedia Computing, Communications, and Applications*, 2018.

[22] D. Ram, A. Asaei, and H. Bourlard, “Phonetic subspace features for improved query by example spoken term detection,” *Speech Communication*, vol. 103, pp. 27–36, Oct. 2018.

[23] D. Ram, A. Asaei, and H. Bourlard, “Sparse subspace modeling for query by example spoken term detection,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2018.

[24] M. Razavi, R. Rasipuram, and M. Magimai.-Doss, “Towards weakly supervised acoustic subword unit discovery and lexicon development using hidden markov models,” *Speech Communication*, vol. 96, pp. 168–183, Feb. 2018.

[25] L. Rozo, H. Ben Amor, S. Calinon, A. Dragan, and D. Lee, “Special issue on robot learning for human-robot collaboration,” *Autonomous Robots*, vol. 42, no. 5, pp. 953–956, Jul. 2018.

[26] S. Ruiz-Correa, I. Ruiz-Correa, C. Olmos Carrillo, F. A. RENDÓN-HUERTA, B. Ramirez-Salazar, L. S. Nguyen, and D. Gatica-Perez, “Mi casa es su casa? examining airbnb hospitality exchange practices in a developing economy,” *ACM Transactions on Social Computing*, 2018.

[27] D. Santani, T.-M.-T. Do, F. Labhart, S. Landolt, E. Kuntsche, and D. Gatica-Perez, “Drinksense: Characterizing youth drinking behavior using smartphones,” *IEEE Transactions on Mobile Computing*, 2018.

[28] D. Santani, S. Ruiz-Correa, and D. Gatica-Perez, “Looking south: Learning urban perception in developing cities,” *ACM Transactions on Social Computing*, 2018.

[29] J. Silverio, S. Calinon, L. Rozo, and D. G. Caldwell, “Learning task priorities from demonstrations,” *IEEE Transactions on Robotics*, 2018.

[30] A. K. Tanwani and S. Calinon, “Small variance asymptotics for non-parametric online robot learning,” *International Journal of Robotics Research (IJRR)*, 2018.

[31] S. Tong, P. N. Garner, and H. Bourlard, “Cross-lingual adaptation of a ctc-based multilingual acoustic model,” *Speech Communication*, vol. 104, pp. 39–46, Nov. 2018.

[32] S. Tulyakov, A. Ivanov, N. Thomas, V. Roloff, A. Pommerol, G. Cremonese, T. Weigel, and F. Fleuret, “Geometric calibration of colour and stereo surface imaging system of ESA’s Trace Gas Orbiter,” *Advances in Space Research*, 2018.

[33] J. Vázquez-Canteli, S. Ulyanin, J. Kämpf, and Z. Nagy, “Fusing tensorflow with building energy simulation for intelligent energy management in smart cities,” *Sustainable Cities and Society*, Nov. 2018.

[34] Y. Yu, K. A. Funes Mora, and J.-M. Odobez, “Headfusion: 360 degree head pose tracking combining 3d morphable model and 3d reconstruction,” *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, vol. 40, no. 11, Nov. 2018.

[35] M. Zeestraten, I. Havoutis, and S. Calinon, “Programming by demonstration for shared control with an application in teleoperation,” *IEEE Robotics and Automation Letters (RA-L)*, vol. 3, no. 3, pp. 1848–1855, 2018.

5.3 PhD Theses

[1] S. Dey, “Phonetic aware techniques for speaker verification,” PhD thesis, EPFL, 2018.

[2] C. Jose, “Learning embeddings: Efficient algorithms and applications,” PhD thesis, École Polytechnique Fédérale de Lausanne, Feb. 2018.

[3] J. Newling, “Novel algorithms for clustering,” PhD thesis, École polytechnique fédérale de Lausanne, Jan. 2018.

[4] X. Pu, “Word sense consistency in statistical and neural machine translation,” PhD thesis, École Polytechnique Fédérale de Lausanne, 2018.

[5] A. K. Tanwani, “Generative models for learning robot manipulation skills from humans,” PhD thesis, Ecole Polytechnique Federale de Lausanne, Feb. 2018.

5.4 Articles in Conference Proceedings

[1] A. Asaei, D. Ram, and H. Bourlard, “Phonological posterior hashing for query by example spoken term detection,” in *Proceedings of Interspeech*, 2018.

[2] P. Baqué, E. Remelli, F. Fleuret, and P. Fua, “Geodesic convolutional shape optimization,” in *Proceedings of the International Conference on Machine Learning*, 2018.

[3] S. Bhattacharjee, A. Mohammadi, and S. Marcel, “Spoofing deep face recognition with custom silicone masks,” in *Proceedings of BTAS2018*, Oct. 2018.

[4] G. Can, Y. Benkhedda, and D. Gatica-Perez, “Ambiance in social media venues: Visual cue interpretation by machines and crowds,” in *IEEE CVPR Workshop on Visual Understanding of Subjective Attributes*, Jun. 2018.

[5] G. Canal, E. Pignat, G. Alenya, S. Calinon, and C. Torras, “Joining high-level symbolic planning with low-level motion primitives in adaptive hri: Application to dressing assistance,” in *Proc. IEEE Intl Conf. on Robotics and Automation (ICRA)*, 2018.

[6] Y. Cao, O. Canévet, and J.-M. Odobez, “Leveraging convolutional pose machines for fast and accurate head pose estimation,” in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Oct. 2018.

[7] M. Cernak and S. Tong, “Nasal speech sounds detection using connectionist temporal,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 2018.

[8] T. Chavdarova, P. Baqué, A. Maksai, S. Bouquet, C. Jose, L. Lettry, F. Fleuret, P. Fua, and L. V. Gool, “Wildtrack: A multi-camera hd dataset for dense unscripted pedestrian detection,” in *Proceedings of the IEEE international conference on Computer Vision and Pattern Recognition*, 2018.

[9] T. Chavdarova and F. Fleuret, “Sgan: An alternative training of generative adversarial networks,” in *Proceedings of the IEEE international conference on Computer Vision and Pattern Recognition*, 2018.

[10] T. Chavdarova, S. Stich, M. Jaggi, and F. Fleuret, “Stochastic variance reduced gradient optimization of generative adversarial networks,” in *International Conference on Machine Learning (ICML) workshop on Theoretical Foundations and Applications of Deep Generative Models*, Jul. 2018.

[11] G. Curtois, V. Grimaldi, H. Lissek, I. Kodrasi, and E. Georganti, “Experimental evaluation of speech enhancement methods in remote microphone systems for hearing aids,” in *Proc. EuroNoise 2018*, Crete, Greece, May 2018, pp. 351–358.

[12] S. Dey, T. Koshinaka, P. Motlicek, and S. Madikeri, “Dnn based speaker embedding using content information for text-dependent speaker verification,” in *Proceedings of 2018 IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2018.

[13] S. Dey, S. Madikeri, and P. Motlicek, “End-to-end text-dependent speaker verification using novel distance measures,” in *Proceedings of Interspeech 2018*, 2018.

[14] P. Dighe, H. Bourlard, and A. Asaei, “Far-field asr using low-rank and sparse soft targets from parallel data,” in *IEEE Workshop on Spoken Language Technology*, 2018.

[15] J. Dumoulin, O. Canévet, M. Villamizar, H. Nunes, O. A. Khaled, E. Mugellini, F. Moscheni, and J.-M. Odobez, “Unicity: A depth maps database for people detection in security airlocks,” in *IEEE International Conference on Advanced Video and Signal-based Surveillance Workshop*, 2018.

[16] S. Ebling, N. C. Camgoz, P. Boyes Braem, K. Tissi, S. Sidler-Miserez, S. Stoll, S. Hadfield, T. Haug, R. Bowden, S. Tornay, M. Razavi, and M. Magimai.-Doss, “Smile swiss german sign language dataset,” in *Language Resources and Evaluation Conference*, 2018.

[17] D. Gatica-Perez, D. Sanchez-Cortes, T.-M.-T. Do, D. B. Jayagopi, and K. Otsuka, “Vlogging over time: Longitudinal impressions and behavior in youtube,” in *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia*, Cairo, EGYPT, Nov. 2018, pp. 37–46.

[18] B. Gerazov, G. Bailly, O. Mohammed, Y. Xu, and P. N. Garner, “Embedding context-dependent variations of prosodic contours using variational encoding for decomposing the structure of speech prosody,” in *Workshop on Prosody and Meaning: Information Structure and Beyond*, Aix-en-Provence, France, Nov. 2018.

[19] W. He, P. Motlicek, and J.-M. Odobez, “Deep neural networks for multiple speaker detection and localization,” in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, Brisbane, AUSTRALIA, May 2018, pp. 74–79.

[20] W. He, P. Motlicek, and J.-M. Odobez, “Joint localization and classification of multiple sound sources using a multi-task neural network,” in *Proceedings of Interspeech*, 2018.

[21] E. Hermann and S. Goldwater, “Multilingual bottleneck features for subword modeling in zero-resource languages,” in *Proc. Interspeech 2018*, Sep. 2018, pp. 2668–2672.

[22] G. Heusch and S. Marcel, “Pulse-based features for face presentation attack detection,” in *Proceedings of BTAS 2018, special session on Image and Video Forensics in Biometrics*, 2018.

[23] M. Ivanova, S. Bhattacharjee, S. Marcel, A. Rozeva, and M. Durcheva, “Enhancing trust in eassessment - the tesla system solution,” in *Proceedings of 2018 Technology Enhanced Assessment Conference.*, Dec. 2018.

[24] N. Jaquier and S. Calinon, “Geometry-aware control and learning in robotics,” in *R:SS Pioneers Workshop*, Jun. 2018.

[25] N. Jaquier, L. Rozo, D. G. Caldwell, and S. Calinon, “Geometry-aware tracking of manipulability ellipsoids,” in *Robotics: Science and Systems*, Pittsburgh, USA, Jun. 2018.

[26] N. Jaquier, L. Rozo, and S. Calinon, “Geometry-aware robot manipulability transfer,” in *R:SS Workshop on Learning and Inference in Robotics: Integrating Structure, Priors and Models*, Jun. 2018.

[27] C. Jose, M. Cisse, and F. Fleuret, “Kronecker recurrent units,” in *Proceedings of the International Conference on Machine Learning*, 2018.

[28] S. H. Kabil, H. Muckenhirn, and M. Magimai.-Doss, “On learning to identify genders from raw speech signal using cnns,” in *Proceedings of Interspeech*, Sep. 2018.

[29] A. Katharopoulos and F. Fleuret, “Not all samples are created equal: Deep learning with importance sampling,” in *Proceedings of International Conference on Machine Learning*, 2018.

[30] K. Khelif, yann Mombrun, G. Hazzani, P. Motlicek, S. Madikeri, F. Sahito, D. Kelly, L. Scarpatto, E. Chatzigavriil, and G. Backfried, “Siip: An innovative speaker identification approach for law enforcement agencies,” in *Big Data and Artificial Intelligence for Military Decision Making*, Meeting Proceedings RDP, http://www.sto.nato.int/, STO, May 2018, PT–1 –1: PT–1 –14.

[31] M. Kleinert, H. Helmke, h. Ehr, K. Christian, D. Klakow, P. Motlicek, M. Singh, and G. Siol, “Building blocks of assistant based speech recognition for air traffic management applications,” in *Conference: SESAR Innovation Days 2018*, European Union, Eurocontrol, Salzburg, Austria: SESARJU, Dec. 2018.

[32] M. Kleinert, H. Helmke, G. Siol, h. Ehr, C. Aneta, K. Christian, D. Klakow, P. Motlicek, Y. Oualil, M. Singh, and A. Srinivasamurthy, “Semi-supervised adaptation of assistant based speech recognition models for different approach areas,” in *37th AIAA/IEEE Digital Avionics Systems Conference*, The best paper award in cathegory "ST-B: Human Factors & Performance for Aerospace Applications" (http://2018.dasconline.org/pages/award-winners), AIAA/IEEE, London, Sep. 2018.

[33] I. Kodrasi and H. Bourlard, “Single-channel late reverberation power spectral density estimation using denoising autoencoders,” in *Proc. Annual Conference of the International Speech Communication Association*, Hyderabad, India, Sep. 2018.

[34] I. Kodrasi and H. Bourlard, “Statistical modeling of speech spectral coefficients in patients with parkinson’s disease,” in *Proc. ITG conference on Speech Communication*, Oldenburg, Germany, Oct. 2018.

[35] I. Kodrasi and S. Doclo, “Joint late reverberation and noise power spectral density estimation in a spatially homogeneous noise field,” in *Proc. International Conference on Acoustics, Speech, and Signal Processing*, Calgary, Canada, Apr. 2018, pp. 441–445.

[36] P. Korshunov, A. R. Goncalves, R. P. V. Violato, F. O. Simões, and S. Marcel, “On the use of convolutional neural networks for speech presentation attack detection,” in *International Conference on Identity, Security and Behavior Analysis*, Jan. 2018.

[37] P. Korshunov and S. Marcel, “Speaker inconsistency detection in tampered video,” in *European Signal Processing Conference*, Sep. 2018.

[38] V. Krivokuca and S. Marcel, “Towards quantifying the entropy of fingervein patterns across different feature extractors,” in *2018 IEEE 4th International Conference on Identity, Security, and Behavior Analysis (ISBA)*, 2018.

[39] N. Le and J.-M. Odobez, “Robust and discriminative speaker embedding via intra-class distance variance regularization,” in *Proceedings of Interspeech*, 2018.

[40] T. S. Lembono, F. Suarez-Ruiz, and Q.-C. Pham, “Scalar - simultaneous calibration of 2d laser and robot’s kinematic parameters using three planar constraints,” in *International Conference on Intelligent Robots*, 2018.

[41] G. Liu, Y. Yu, K. A. Funes Mora, and J.-M. Odobez, “A differential approach for gaze estimation with calibration,” in *29TH BRITISH MACHINE VISION CONFERENCE*, 2018.

[42] S. Madikeri, S. Dey, and P. Motlicek, “Analysis of language dependent front-end for speaker recognition,” in *Proceedings of Interspeech 2018*, 2018.

[43] A. Martínez-González, M. Villamizar, O. Canévet, and J.-M. Odobez, “Investigating depth domain adaptation for efficient human pose estimation,” in *European Conference on Computer Vision - Workshops*, 2018.

[44] A. Martínez-González, M. Villamizar, O. Canévet, and J.-M. Odobez, “Real-time convolutional networks for depth-based human pose estimation,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2018.

[45] L. Miculicich, N. Pappas, D. Ram, and A. Popescu-Belis, “Self-attentive residual decoder for neural machine translation,” in *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, 2018.

[46] L. Miculicich, D. Ram, N. Pappas, and J. Henderson, “Document-level neural machine translation with hierarchical attention networks,” in *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2018.

[47] H. Muckenhirn, M. Magimai.-Doss, and S. Marcel, “On learning vocal tract system related speaker discriminative information from raw signal using cnns,” in *Proceedings of Interspeech*, Sep. 2018.

[48] H. Muckenhirn, M. Magimai.-Doss, and S. Marcel, “Towards directly modeling raw speech signal for speaker verification using cnns,” in *IEEE International Conference on Acoustics, Speech and Signal Processing*, Calgary, CANADA, Apr. 2018, pp. 4884–4888.

[49] S. Muralidhar, L. S. Nguyen, and D. Gatica-Perez, “Words worth: Verbal content and hirability impressions in youtube video resumes,” in *Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, 2018.

[50] S. Muralidhar, R. Siegfried, J.-M. Odobez, and D. Gatica-Perez, “Facing employers and customers: What do gaze and expressions tell about soft skills?” In *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia*, 2018.

[51] O. Nikisins, T. Eglitis, A. Anjos, and S. Marcel, “Fast cross-correlation based wrist vein recognition algorithm with rotation and translation compensation,” in *Sixth International Workshop on Biometrics and Forensics*, Jun. 2018.

[52] O. Nikisins, A. Mohammadi, A. Anjos, and S. Marcel, “On effectiveness of anomaly detection approaches against unseen presentation attacks in face anti-spoofing,” in *The 11th IAPR International Conference on Biometrics (ICB 2018)*, Feb. 2018.

[53] N. Pappas, L. Miculicich, and J. Henderson, “Beyond weight tying: Learning joint input-output embeddings for neural machine translation,” in *Proceedings of the Third Conference on Machine Translation (WMT)*, 2018.

[54] D. Ram, L. Miculicich, and H. Bourlard, “Cnn based query by example spoken term detection,” in *Proceedings of Interspeech*, 2018.

[55] G. Ramet, P. N. Garner, M. Baeriswyl, and A. Lazaridis, “Context-aware attention mechanism for speech emotion recognition,” in *IEEE Workshop on Spoken Language Technology*, Athens, Greece, Dec. 2018, pp. 126–131.

[56] B. Schnell and P. N. Garner, “A neural model to predict parameters for a generalized command response model of intonation,” in *Proc. Interspeech 2018*, Sep. 2018, pp. 3147–3151.

[57] J. Sebastian, M. Kumar, D. S. P. Kumar, M. Magimai.-Doss, H. A. Murthy, and S. Narayanan, “Denoising and raw-waveform networks for weakly-supervised gender identification on noisy speech,” in *Proceedings of Interspeech 2018*, 2018.

[58] A. Shajkofci and M. Liebling, “Semi-blind spatially-variant deconvolution in optical microscopy with local point spread function estimation by use of convolutional neural networks,” in *2018 25th IEEE International Conference on Image Processing (ICIP)*, IEEE, Sep. 2018, pp. 3818–3822.

[59] J. Silverio, S. Calinon, L. Rozo, and D. G. Caldwell, “Bimanual skill learning with pose and joint space constraints,” in *Proc. of the IEEE-RAS Intl Conf. on Humanoid Robots (Humanoids)*, 2018.

[60] J. Silverio, Y. Huang, L. Rozo, S. Calinon, and D. G. Caldwell, “Probabilistic learning of torque controllers from kinematic and force constraints,” in *Proc. of the IEEE/RSJ Intl Conf. on Intelligent Robots and Systems (IROS)*, 2018, pp. 6552–6559.

[61] S. Srinivas and F. Fleuret, “Knowledge transfer with jacobian matching,” in *Proceedings of the International Conference on Machine Learning*, 2018.

[62] A. Srinivasamurthy, P. Motlicek, M. Singh, Y. Oualil, M. Kleinert, h. Ehr, and H. Helmke, “Iterative learning of speech recognition models for air traffic control,” in *Proceedings of Interspeech 2018*, ISCA, Hyderabad, India, Sep. 2018, pp. 3519–3523.

[63] M. Tammen, I. Kodrasi, and S. Doclo, “Complexity reduction of eigenvalue decomposition-based diffuse power spectral density estimators using the power method,” in *Proc. International Conference on Acoustics, Speech, and Signal Processing*, Calgary, Canada, Apr. 2018, pp. 451–455.

[64] M. Tammen, I. Kodrasi, and S. Doclo, “Iterative alternating least-aquares approach to jointly estimate the retfs and the diffuse psd,” in *Proc. ITG conference on Speech Communication*, Oldenburg, Germany, Oct. 2018.

[65] A. K. Tanwani, J. Lee, B. Thananjeyan, M. Laskey, S. Krishnan, R. Fox, K. Goldberg, and S. Calinon, “Generalizing robot imitation learning with invariant hidden semi-markov models,” in *13th Intl Workshop on the Algorithmic Foundations of Robotics (WAFR)*, 2018.

[66] S. Tong, P. N. Garner, and H. Bourlard, “Fast language adaptation using phonological information,” in *Proceedings of Interspeech 2018*, 2018.

[67] S. Tulyakov, A. Ivanov, and F. Fleuret, “Practical deep stereo (pds): Toward applications-friendly deep stereo matching,” in *Proceedings of the international conference on Neural Information Processing Systems*, 2018.

[68] M. Villamizar, A. Martínez-González, O. Canévet, and J.-M. Odobez, “Watchnet: Efficient and depth-based network for people detection in video surveillance systems,” in *IEEE International Conference on Advanced Video and Signal-based Surveillance*, 2018.

[69] B. Vlasenko, J. Sebastian, D. S. P. Kumar, and M. Magimai.-Doss, “Implementing fusion techniques for the classification of paralinguistic information,” in *Proceedings of Interspeech 2018*, 2018.

[70] Y. Yu, G. Liu, and J.-M. Odobez, “Deep multitask gaze estimation with a constrained landmark-gaze model,” in *European Conference on Computer Vision Workshop*, Sep. 2018.

6 List of Patents

While evaluating the success of technology transfer is a difficult task, invention disclosures and patents are usually key metrics in the field. As a necessary first step to any patent filing, an invention disclosure aims to identify a technology with a certain level of maturity and some promising economic potential.

In 2018, 16 invention and software disclosures have been filled. A patent committee examines each invention disclosure and decides to move forward with a patent filing or not. The committee is composed of the Head of technology transfer, two representatives of the Direction and one representative of the researchers. Since 2012, 8 patents have been filed:

- Granted patents
  - US20130096922: A. Asaei, H. Bourlard, V. Cevher, “Method, apparatus and computer program product for determining the location of a plurality of speech sources”, Grant Number: 09689959
  - US20140089365: C. Dubout, F. Fleuret, “Object detection method, object detector and object detection computer program”, Grant Number: 09058541
  - US20170171210: S. Marcel, A. Anjos, P. Abbet, “Method and internet-connected server for reviewing a computer-executable experiment”, Grant Number: 09973503
- Patents having entered in national phase
  - WO2015192879 / EP3154407: K. A. Funes Mora, J-M. Odobez, “A gaze estimation method and apparatus”, International Application Number: PCT/EP2014/062604
  - WO2016023582 / EP3180736: S. Marcel, “A method of detecting a falsified presentation to a vascular recognition system”, International Application Number: PCT/EP2014/067290
  - WO2017221049: A. Anjos, S. Marcel, “A data-network connected server, a device, a platform and a method for conducting computer-executable experiments”, International Application Number: PCT/IB2016/053683
- Patents filed and published
  - US20140149104: N-H. Kim, P. Motlicek, P. N. Garner, D. Imseng, J-W. Lee, J-M. Cho, “Apparatus and method for constructing multilingual acoustic model and computer readable recording medium for storing program for performing the method”, Publication Number: 20140149104
  - US20170069306: A. Asaei, M. Cernak, H. Bourlard, “Signal processing method and apparatus based on structured sparsity of phonological features”, Publication Number: 20170069306