



TECHNOLOGY PORTFOLIO

Contents

1	VALUE PROPOSITION	1
1.1	Portrait	1
1.2	Research groups	1
1.3	Research and development engineers group	2
1.4	Technology Transfer Office	3
2	TECHNOLOGY TRANSFER	4
2.1	Project examples	5
2.2	IdeArk: Idiap startup incubator	7
3	PATENTS	8
3.1	Introduction	8
3.2	Patents list	8
4	TECHNOLOGY COMPONENTS	9
4.1	Audio and speech processing	9
4.1.1	Intonation Modeling	9
4.1.2	Multilingual and accented speech recognition	10
4.1.3	Keyword Spotting for Searching and Indexing Spoken Documents	11
4.1.4	Acoustic Simulator	12
4.1.5	Forensic Voice Comparison	13
4.1.6	Grapheme-based and Instance-based Methods for Speech Processing	14
4.1.7	Large Scale Speaker Identification	15
4.1.8	Speaker Verification	16
4.1.9	Speaker Segmentation and Linking	17
4.1.10	Speaker Verification – accessing secure systems	18
4.1.11	Automatic Speech Assessment	19
4.1.12	Speech-Based Just-In-Time Information Retrieval	20
4.1.13	Speech Synthesis	21
4.1.14	Speech-to-Text Transcription	22
4.1.15	Very Low Bit-Rate Speech Coding	23
4.1.16	Sparse Phonological Vocoding	24

4.1.17	Ad Hoc Microphone Arrays	25
4.1.18	Speaker Localization and Separation	26
4.2	Machine learning and patter recognition	27
4.2.1	Signal Processing and Machine Learning Toolbox (BOB)	27
4.2.2	Activity Monitoring from Ambient Sensors	28
4.2.3	Fast Object Detector	29
4.2.4	Large-scale Image Retrieval	30
4.2.5	Multi-camera Detection and Tracking	31
4.2.6	Unsupervised Activity Discovery, Monitoring, and Abnormality Detection	32
4.2.7	Scalable linear discriminant analysis for large scale pattern recognition	33
4.2.8	Soft-skill analytics from online video job interviews	34
4.3	Computer vision	35
4.3.1	Cardiac Imaging	35
4.3.2	Real-time Mutiple face and head pose tracking	36
4.3.3	Motion Detection, Segmentation, and Estimation. Motion Satbilization	37
4.3.4	Gaze Estimation and Non-Verbal Behavior Extraction from consumer cameras	39
4.3.5	Static Luggage Detection and Intrusion Detection	40
4.4	Biometrics	41
4.4.1	Face recognition library and awarded ISV technology	41
4.4.2	Fingervein recognition	42
4.4.3	A method of detecting a falsified presentation to a vascular recognition system	43
4.4.4	Palmvein recognition library	44
4.4.5	Speaker recognition library and award winning technology	45
4.5	Multimedia processing	46
4.5.1	Hybrid Recommender System for Multimedia	46
4.5.2	Machine Translation at the Document Level	47
4.5.3	Text detection and recognition in images and videos	48
4.5.4	Social context inference from proximity sensors	49
4.6	Robotics	50
4.6.1	Robot Skills Transfer Toolbox	50

1 VALUE PROPOSITION

1.1 Portrait

The Idiap Research Institute, based in Martigny, Valais - Switzerland, is a non-profit research foundation specialized in multimedia information management and multimodal man-machine interaction. The institution was founded in 1991 by the Town of Martigny, State of Valais, the Ecole Polytechnique Fédérale de Lausanne (EPFL), the University of Geneva and Swisscom. Idiap is connected to the EPFL through a Joint Development Plan.

1.2 Research groups

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

Machine Learning: The goal of the group is the development of new machine learning techniques, with a particular interest in their computational properties. The application domain is mainly computer vision and includes object detection, scene analysis, tracking of persons and biological structures, and image recognition in general.

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

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

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

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

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

Natural Language Understanding: The Idiap NLU group was created in September 2017 under the direction of James Henderson, in part as a continuation of the previous Natural Language Processing group which was lead by Andrei Popescu-Belis. The NLU group continues work on how semantic and discourse processing of text and dialog can improve statistical machine translation and information indexing, with a recent focus on neural machine translation and attention-based deep learning models. This fits well with the NLU group's new research direction of neural network structured prediction and representation learning for modeling the syntax and semantics of text and speech, including modeling abstraction (textual entailment) and summarization.

Biosignal Processing: The Biosignal Processing Group was created in 2018. Biosignals are signals from biomedical sensors 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 sensing for healthcare and other related 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.

Energy Informatics: 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.

Genomics and Health Informatics: The mass generation of high-content genomic, imaging and digital data has brought rapid advances in our understanding of cell biology and human diseases, and offers promising avenues for the field of precision medicine. A major obstacle revolves around delivering meaningful information from these data, requiring the deployment of a variety of methods such as statistics, machine learning and data visualisation in tight collaboration with clinicians. The Genomics and Health Informatics Group was created in 2019 to develop statistical and machine-learning methods to integrate genomic, clinical, and imaging data. The group aims to accelerate the diagnosis and improve the understanding and treatment of complex diseases such as neurodegenerative disorders in collaboration with clinical neuroscientists. The group has expertise in genomics, bioinformatics, RNA biology, neuroscience, data science, and data visualisation.

1.3 Research and development engineers group

Since the end of 2006, Idiap maintains an active group of a dozen of highly talented developers, led by Olivier Bornet, whoes goal is to be the essential links between research and industry. They are sharing multiple responsibilities, including:

- **Industrial projects:** Development of prototypes, and adaptation of research software to the needs of industries, with the goal to demonstrate feasibility of new products and systems, and in direct response to a contract with the industry or the needs of one of the Idiap' spin-offs (always in the framework of well defined projects and IPR contracts). This also includes the development part of CTI projects, where

responsibilities are then shared by the PI researchers and one of the developers.

- Academic projects: The Research and development engineers group is also responsible for the development of demonstration systems, either of general interest to Idiap (in which case efforts are funded by Idiap) or in the context of projects, e.g., EU project involving specific development efforts going beyond what researchers are able or willing to do.
- Internal support: The Research and development engineers group is also responsible to respond to in-house requests, e.g., helping out colleagues in resolving complex code issues, contributing in research or collaborative platform development, etc. In this context, the group is also responsible for the quality control and maintenance of the open source libraries made publicly available. As available for the Idiap web site, open-source libraries are available from <http://www.idiap.ch/technology-transfer/open-source-software>. Each library has to be thoroughly tested and validated before this release, and nobody at Idiap is allowed to release open source software without prior permission of the management and without doing this through the Research and development engineers group, who will make sure of the quality of what is being delivered, perform version control, etc.
- Stable software and platforms: This is not only part of the internal support discussed above but also for “close source” software which is being licensed to industries. There also, the Research and development engineers group is responsible for quality control, version control, and maintenance (if necessary). In collaboration with the TTO, they also have to make sure that we all follow the same IPR policy and keep track of the licensing and IPR status.

1.4 Technology Transfer Office

Idiap is very active in multiple, national and international, technology transfer activities, and is also involved in numerous projects with industries, ranging from large institutions to SMEs and startups.

The Technology Transfer Office represents the key link between researchers, research and development engineers, and industry. Besides maintaining Idiap’s technology portfolio (presented below), and responding to industrial needs and contacts, the TTO also pro-actively investigates new opportunities. In this context, it assists Idiap researchers and industries to develop joint projects. As also discussed below, and illustrated in the present document, the TTO is also responsible to maintain a clear IPR strategy and track IP status across licenses. The TTO is thus a key entry link between the scientists, the industries, and the Research and development engineers group.

2 TECHNOLOGY TRANSFER

Idiap is committed to the transfer of knowledge and technology to the business community through an active engagement and partnership. The Technology Transfer Unit works in close collaboration with businesses to realize joint research and sponsored development programs along the research areas of Idiap. From small startups to large corporations, Idiap Research Institute is an ideal partner in “avant garde” technology and research!

Technology transfer is one of Idiap Research Institute’s three core mission including fundamental research and education. With “Technology transfer” we mean the process that leads the technology (software, algorithm, knowledge, expertise) to be transferred from Idiap to an industrial partner or a spin-off for economic development purposes.

This transfer is usually done by giving the grant of rights on the commercial exploitation of this technology (through license). We can therefore consider the process of technology transfer as a creation of economic value out of scientific discoveries. If the concept sounds simple, the path that leads from a technology demonstrator to an industrial product is highly complex. One reason is that researchers and industrialists do not speak the “same language” and their aim could be totally different.

One of the fundamental challenges in “technology transfer” is to facilitate the interface between the knowledge and skills of the researcher and the needs of the industrial partner. Idiap has resolved this by providing a pluri-disciplinary team of developers, programers dedicated to technology transfer. Also, the incubator IdeArk (<http://www.ideark.ch/en/>, see Section 2.2), a spin-off from Idiap has the mission to enhance Idiap’s technologies and promote the emergence of new startups by providing workspace, financing and mentorship.

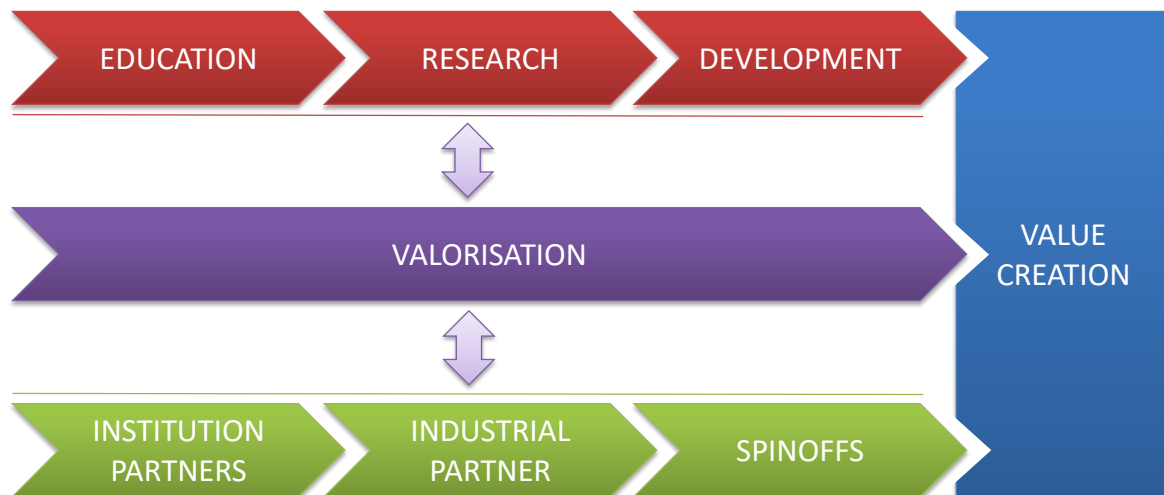


Figure 1: Value Creation

2.1 Project examples

Six-month project to design a novel alignment approach for a manufacturing process

The industrial partner was a Swiss-based supplier of solutions for manufacturing plastic components. The project was initiated through pre-existing contacts with Idiap employees.

The problem at hand was to add into a pre-existing manufacturing process a procedure to align automatically plastic elements as they were manipulated before being assembled.

The preliminary discussions between Idiap and the company lasted for three months, after which a proposal was drafted collaboratively and submitted to TheArk to fund one developer at 50% for six months at Idiap, with a commitment from the industrial partner for four months of work on their side. The funding for the project was accepted in less than a month.

The company then provided Idiap with a test machine to record measurements from a reference sensor. From there at Idiap a software developer funded through the project, the project manager and a scientific supervisor met weekly. They designed and benchmarked a novel algorithm able to estimate automatically deviations from a desired alignment in spite of heavy structured noise in the measurements. This development was done with a monthly face-to-face meeting with the industrial partner to discuss progresses.

The outcome of the project was a novel algorithm, a performance report, and a reference implementation in C that could be directly transferred to the industrial partner for integration into its manufacturing robots. There were a few additional technical exchanges with Idiap for minor software changes.

In the following months the company filed a patent to protect the method.

Resources: 4 PM Engineer

Funding: 50'000 CHF

One year project to develop activity analysis methods for a start-up company

The industrial partner is a Swiss startup in the domain of health. The objectives of the collaboration are confidential; the domain of work is activity analysis.

The project was initiated through contacts between TTO and the company. After few months of interaction and concept definition, a CTI proposal was prepared, submitted, and accepted. Idiap acted as lead research partner.

The CTI project supported one postdoctoral researcher at full-time for one year and a software developer at part-time during one year.

Interaction included weekly interactions between research partners and monthly-to-bimonthly team meetings.

The main goals of the project were achieved, but there was a need to re-orient some of the research due to limitations related to availability of labeled data for machine learning.

This project encouraged the partners to apply for a follow-up CTI project, who was ultimately granted.

Resources: 12 PM Post-Doc, 8 PM Engineer

Funding: 270'000 CHF

One year project to develop security solutions in the domain of mobile biometrics

The industrial partner is a large US-based Internet company. The contact was initiated following a short research stay within the company premises. After the research stay the company was enthusiastic by the skills of our researchers and decided to fund a follow-up research project.

The precise objective of the collaboration is confidential and we can only mention that it is related to biometrics security on mobile devices.

In only one month, Idiap and the company prepared a research and an IP-agreement, specifying milestones and deliverables. The industrial partner agreed to fund through direct payment two researchers (scientific collaborators).

Over the duration of the project, the researchers will develop and evaluate algorithms to address specific vulnerability aspects in mobile biometrics. Idiap remains in contact with the industrial partner through remote meetings with the project manager.

The deliverables are composed of python/C++ software, in a form of an independent archive based on our BOB library (<https://www.idiap.ch/software/bob>), which was provided under an exclusive license of use in a restrictive field.

Resources: 24 PM Engineer

Funding: 275'000 USD

Fourteen months project to develop a novel tool for prediction in the domain of biometrics

The industrial partner was a Russian Security company, and the contact was initiated by them by searching for Research Institutions with expertise in biometrics on the web.

The objective of the collaboration was to design a novel prediction algorithm for biometrics.

During two months, Idiap and the company conducted preliminary discussions and drafted a research and an IP-agreement, the former specifying milestones and deliverable. The industrial partner agreed to fund through direct payment two researchers at the post-doctoral level and a part-time software developer. It also provided Idiap with a test database for benchmarking the developed methods.

Over the duration of the project, the researchers funded on the project developed a new algorithm and a prototype implementation, and the software developer wrote a industrial-grade implementation packaged into a library, and developed an accompanying graphic interface for testing. Idiap remained in contact with the industrial partner through meetings with the CEO directly every two months.

The deliverables were composed of an evaluation report and an archive of the library, which was provided under an exclusive license of use in a restrictive field.

Resources: 24 PM Post-Doc

Funding: 250'000 EUR

Three year project on state of the art large vocabulary speech recognition

The industrial partner was Samsung in Korea. We were known to their research group mainly via some of our software open-sourced previously; we were invited to apply for their annual call for research partnerships.

Funding was granted under the Samsung [GRO program](#). The object of the collaboration was to develop and evaluate state of the art speech recognition for distributed platforms. The project was renewed once with the same funding scheme and a slightly different focus. It was then renewed for a third year, with funding coming directly from the research lab, rather than the GRO program.

The Samsung GRO has a standard IP sharing agreement.

The collaboration was eased by use of open-source solutions, and resulted in a patent being filed ([EP2736042A1](#)).

Resources: 20 PM Researchers, 8 PM Post-Doc

Funding: 310'000 USD

2.2 IdeArk: Idiap startup incubator

As part of the Valais initiative *TheArk*, IdeArk relays the concerns of companies to the scientists, identifies market trends and facilitates the conversion of research results into innovative solutions. IdeArk's task is also to support start-ups and SMEs active in Idiap-related fields.

The IdeArk Incubator provides multiple types of support and services to startups, including: free office spaces with furnitures, internet connection, coaching support, and financial support for collaborative projects with Idiap.

Many technologies from Idiap were developed into startups through the IdeArk support (KeyLemon, Klewel, AudioSearch, Koemei, recapp, ...) and naturally benefit from Idiap's proximity and expertise.

Once a year, an *International Create Challenge* (ICC, <http://www.createchallenge.org/>) is also being held within the Idiap premises. Initiated in 2012, ICC is a free of charge, 3-week immersive super accelerator program that gives entrepreneurs and potential entrepreneurs the opportunity to drive their projects to a "Minimum Viable Product" (MVP). It is a unique program combining state-of-art technologies, cutting edge research, mentors-led coaching, and micro-seed investment (in collaboration with VCs and KickVenture).

3 PATENTS

3.1 Introduction

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 2019, 32 invention and software disclosures have been filed. 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, 11 patents have been filed: 3 patents are granted, 7 applications are published and pending (1 of them has been sold) and 1 application was abandoned.

3.2 Patents list

Following is a list of our published patents (3 granted, 7 pending applications). If you are interested in enriching your IP portfolio by buying one of them or if you would like to know about the licensing possibilities and conditions, please contact us: tto@idiap.ch.

- Granted patents

IDIAP-1 [US 9,689,959 B2] A. Asaei, H. Bourlard, V. Cevher, “Method, apparatus and computer program product for determining the location of a plurality of speech sources”

IDIAP-2 [US 9,058,541 B2] C. Dubout, F. Fleuret, “Object detection method, object detector and object detection computer program”

IDIAP-8 [US 9,973,503 B2] S. Marcel, A. Anjos, P. Abbet, “Method and internet-connected server for reviewing a computer-executable experiment”

- Pending applications

SAMSUNG-1 [US 2014/0149104 A1, EP 2 736 042 A1] 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”

IDIAP-5 [WO 2015/192879 A1] K. A. Funes Mora, J-M. Odobez, “A gaze estimation method and apparatus”

IDIAP-9 [WO 2017/221049 A1, US 2019/0180040 A1] A. Anjos, S. Marcel, “A data-network connected server, a device, a platform and a method for conducting computer-executable experiments”

globID-1 [WO 2019/150254 A1] L. Sonna Momo, L. Cerqueira Torres, S. Marcel, A. Anjos, M. Liebling, A. Shajkofci, S. Amos, A. Woeffray, A. Sierro, “Method and device for biometric vascular recognition and/or identification”

IDIAP-10 [EP 3 691 258 A1] M. Liebling, C. Jacques, “System and method for acquiring images”

IDIAP-11 [EP 3 719 679 A1] V. Krivokuca Hahn, S. Marcel, “A method for protecting biometric templates, and a system and method for verifying a speaker’s identity”

- Sold

IDIAP-4 [WO 2016/023582 A1] S. Marcel, “A method of detecting a falsified presentation to a vascular recognition system”

4 TECHNOLOGY COMPONENTS

4.1 Audio and speech processing

4.1.1 Intonation Modeling

Keywords

Intonation modelling; text-to-speech systems; emotion recognition; speech signal processing

Key contact researcher(s)

Dr. Philip N. Garner

Phil.Garner@idiap.ch

Technology Transfer Office

Dr. Joël Dumoulin

tto@idiap.ch

File reference & version number:

Software disclosure 8700

Software disclosure 11457

Functional description

Intonation refers to the way pitch and energy of a speech signal vary in order to convey meaning, intent or emotion. It is normally ignored in speech recognition, but is essential for natural sounding speech synthesis. Idiap has novel intonation models with a physiological basis, closely matching the human production mechanism. Designed for cross-lingual emotion transfer in speech to speech translation, the models have potential application in emotion recognition and speech modification.

Innovative aspects

- Biologically plausible physiological model
- Cross-lingual adaptation

Commercial application examples

- Speech to speech translation
- Emotion recognition
- Emotional speech synthesis

More information

“Atom Decomposition-based Intonation Modelling”, Pierre-Edouard Honnet, Branislav Gerazov and Philip N. Garner, in: *IEEE 40th International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brisbane, Australia, IEEE, 2015.

Software & IPR status

Much of the underlying software is or will soon be open-source.

- SSP: <https://github.com/idiap/ssp>

4.1.2 Multilingual and accented speech recognition

Keywords

Speech recognition, mixed language speech recognition, non-native speech recognition, fast adaptation

Key contact researcher(s)

Dr. David Imseng

David.Imseng@idiap.ch

Technology Transfer Office

Dr. Joël Dumoulin

tto@idiap.ch

File reference & version number:

Spin-off Agreement 20150101

Functional description

Idiap is one of the world's leading providers of multilingual speech recognition technology. Multilingual speech recognition involves numerous research challenges, including adaptation on limited amount of training data and mixed language recognition where it is not even possible to assume that one knows in advance the language being spoken. Idiap's system can rapidly adapt to under-resourced languages, i.e. languages that have only a small amount of data available to train a speech recognition system, or to accented speech. Furthermore, Idiap's system allows to perform mixed language speech recognition for Swiss German, Swiss French and English.

Innovative aspects

- Parsimonious use of language adaptation data (adaptation with very little amount of data)
- Fast adaptation to unseen languages
- Generic probabilistic multilingual framework

Commercial application examples

- Robust non-native speech recognition
- Mixed language speech recognition
- Language identification

More information

“Automatic speech recognition and translation of a Swiss German dialect: Walliserdeutsch”, Philip N. Garner, David Imseng, and Thomas Meyer in: *Proceedings of Interspeech*, 2014.

Software & IPR status

Most of the underlying software is open source. Some databases associated with certain languages may require separate agreements.

The Idiap spin-off recapp IT (www.recapp.ch) commercialises parts of the technology and can provide multilingual speech recognition as a webservice.

4.1.3 Keyword Spotting for Searching and Indexing Spoken Documents

Keywords

Speech processing; keyword spotting; data mining; information retrieval

Key contact researcher(s)

Prof. Hervé Bourlard

bourlard@idiap.ch

and

Dr. Petr Motlicek

petr.motlicek@idiap.ch

and

Dr. Mathew Magimai Doss

Mathew.MagimaiDoss@idiap.ch

Technology Transfer Office

Dr. Joël Dumoulin

tto@idiap.ch

File reference & version number:

Software disclosure 11639

Functional description

Idiap offers solutions to search and index spoken documents (audio/multimedia archives) through keyword spotting (KWS) technology. The solutions support both text-based query input and query-by-example input i.e. spoken query. The KWS systems are in-parts or fully based on state-of-the-art automatic speech recognition technology. Consequently, Idiap can provide customized solutions of keyword spotting systems for different languages and domains.

Innovative aspects

- Neural network based acoustic modeling, technology allowing for speaker/channel/sentence-based adaptation
- Support of several speech transmission channels (landline telephone, mobile phone and voice over IP)
- Several times faster than real-time solution, enabling search in very large audio archives
- Robust solutions also with noisy recordings
- Unified framework for text query-based KWS and query-by-example based KWS

Commercial application examples

- Search for specific information in large audio archives
- Audio content data-mining and indexing

More information

Motlicek, Petr, Valente, Fabio and Szoke, Igor, “Improving Acoustic Based Keyword Spotting Using LVCSR Lattices”, *Proceedings on IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 4413-4416, Japan, 2012.

Software & IPR status

- Most of the underlying software is open source. Some databases associated with certain languages may require separate agreements.
- The Idiap spin-off AudioSearch SARL is commercializing a part of the keyword spotting technology.

4.1.4 Acoustic Simulator

Keywords

Acoustic simulator; Noise; Reverberation ;
Codec

Key contact researcher(s)

Dr. Srikanth Madikeri

srikanth.madikeri@idiap.ch

and

Dr. Petr Motlicek

petr.motlicek@idiap.ch

Technology Transfer Office

Dr. Joël Dumoulin

tto@idiap.ch

File reference & version number:

N/A

Functional description

Idiap maintains an open-source acoustic simulator that provides a large number of speech signal degradation processes. Based on a large database of 80h of real noise recordings, over a hundred impulse responses of speakers, cabinets, smartphones and rooms along with 15 different speech and audio codecs, this simulator stands out as a valuable pool of resources to develop detailed and robust models for speaker and speech recognition technology. It currently covers landline, cellular, satellite, VOIP, interview and playback scenarios.

Innovative aspects

- Large number of speech signal degradation processes
- Open-source data and code
- Expandable functionality

Commercial application examples

- Development of robust speech processing systems
- Low-resource speech applications

More information

Ferras, Marc, Madikeri, Srikanth, Dey Subhadeep, Motlicek, Petr Bourlard Herve, “A Large-Scale Open-Source Acoustic Simulator for Speaker Recognition”, *IEEE Signal Processing Letters*, 23(4), pp. 527-531, 2016.

Software & IPR status

IP lies in the copyright of the code. Noise data and impulse responses are online resources released under a Creative Commons Attribution license. Codec code has more restrictive licenses.

4.1.5 Forensic Voice Comparison

Keywords

Voice comparison; speaker recognition; security; forensics; tracking of criminal activities

Key contact researcher(s)

Dr. Petr Motlicek
pmotlic@idiap.ch

Technology Transfer Office

tto@idiap.ch

File reference & version number:

141

Functional description

Forensic voice comparison involves the expert assessment of voice recordings to determine the identity of a speaker (offender). The speakers involved are termed as known speakers and questioned speakers. The former is a set of one or more speakers whose identity is known while the latter is the speaker whose identity we wish to confirm with respect to the known speakers. A fully-automatic (or semi-automatic) approach available at Idiap exploits an automatic speaker recognition system to obtain, through a series of carefully designed steps, a score signifying the proximity of the voice characteristics of the offender to that of the known speaker.

Innovative aspects

- Language-independent technology based on a distributable service that takes into account relevant population data matching the questioned-speaker data
- Score evaluation and quality assurance provided in a final report
- Client-server interaction using the standard REST API

Commercial application examples

- Pay-Per-Use licensing scheme or fully commercial system can be offered
- Service specifically targeted for forensic voice comparison that exposes latest technologies in speaker recognition
- Latest technology with large interest for government security bureaus and law-enforcement agencies

Software & IPR status

The above mentioned approach to forensic voice comparison is being developed internally and for the partners in the relevant projects.

4.1.6 Grapheme-based and Instance-based Methods for Speech Processing

Keywords

Speech processing; Markov models; computational linguistics; graphemes; phonemes

Key contact researcher(s)

Dr. Mathew Magimai Doss

Mathew.MagimaiDoss@idiap.ch

Technology Transfer Office

Dr. Joël Dumoulin

tto@idiap.ch

File reference & version number:

Software disclosure 8700

Software disclosure 11457

Functional description

Phonetic lexicon and acoustic resources (transcribed speech data) are the primary resources needed for development of speech technologies, such as automatic speech recognition, text-to-speech synthesis, spoken term detection. Not all languages or domains may have such well developed resources. To overcome these limitations, Idiap is actively engaged in research on alternate ways to model the words in the language through acoustics. More specifically, (a) by modeling units of written form i.e. graphemes and (b) by modeling spoken instances.

Innovative aspects

- Speech recognition without the need of a phonetically transcribed lexicon
- Development of linguistic resources and speech technology for under-resourced languages and domains

Commercial application examples

- Speech and language technology applications where lexicon with phonetic transcription is not readily available
- Front-end of text-to-speech systems
- Search for specific information in large audio archives
- Audio content data-mining and indexing

More information

“Acoustic and Lexical Resource Constrained ASR using Language-Independent Acoustic Model and Language-Dependent Probabilistic Lexical Model”, Ramya Rasipuram and Mathew Magimai-Doss, *Speech Communication*, Vol. 68, April, 2015.

Software & IPR status

The complete software consists of open source softwares such as HTK, Quicknet, SSP (SSP: <https://github.com/idiap/ssp>) and Idiap proprietary code that interfaces the open source tools, which can be customized to the customers need. Speech databases and lexical resources used for development may require commercial license.

4.1.7 Large Scale Speaker Identification

Keywords

Speaker identification; search in large database; security; forensics; tracking of criminal activities

Key contact researcher(s)

Dr. Petr Motlicek

petr.motlicek@idiap.ch

Technology Transfer Office

Dr. Joël Dumoulin

tto@idiap.ch

File reference & version number:

N/A

Functional description

Idiap's speaker identification technology focuses on datasets with large speaker populations. It exploits several complementary technologies (iVectors, Subspace GMMs) to generate low-dimensional but representative speaker models, followed by channel compensation technique.

Our environment simulator provides massive resources for the development of speaker identification systems that are more robust to acoustic mismatch.

Innovative aspects

- Combining state-of-the-art speech and speaker recognition technologies
- Simulation of a large variety of acoustic environments for improved robustness
- The technology is language-, text- and channel-independent
- Possibility to detect phone exchange within the same speaker
- New speaker can be easily added by end-users
- The technology supports all speaker identification alternatives: 1:1, 1:N, N:M (group of speakers against group of speaker models)

Commercial application examples

- Search and retrieval of speaker identities (in large audio archives)
- Speaker identification of multi-styled and noisy speech
- The technology can be easily combined with gender and language identification

More information

Petr Motlicek et al., "Employment Of Subspace Gaussian Mixture Models in Speaker Recognition", in: Proc. of IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP), pp. 4755–4799, 2015.

Software & IPR status

The above mentioned approaches to speaker identification are developed within the open-source Kaldi toolkit. The acoustic simulator is available as an open-source package from publicly available resources.

4.1.8 Speaker Verification

Keywords

Speaker recognition; speaker verification; security; forensics; tracking of criminal activities

Key contact researcher(s)

Dr. Petr Motlicek

petr.motlicek@idiap.ch

Technology Transfer Office

Dr. Joël Dumoulin

tto@idiap.ch

File reference & version number:

N/A

Functional description

Idiap's speaker verification/authentication technology focuses on datasets with large speaker populations. Besides state-of-the-art approaches based on iVectors and SGMM modeling generating low-dimensional speaker models, followed by channel compensation techniques, a hybrid approach successfully integrates our state-of-the-art speech recognition (or key-word spotting) technology into a speaker recognition system that benefits from linguistic resources. Our environment simulator provides massive resources for the development of speaker verification systems that are more robust to acoustic mismatch.

Innovative aspects

- Combining state-of-the-art speech and speaker recognition technologies
- Exploiting the content of the spoken message to enhance verification capabilities
- Simulation of a large variety of acoustic environments for improved robustness
- Multiple enrollment solutions offering enhanced scoring capabilities

Commercial application examples

- Verification/authentication of caller against large database of speaker models
- Speaker verification of multi-styled and noisy speech
- Security: ID verification, forensics, tracking of criminal activities
- Gatekeeper for accessing secured systems (“voice-as-password” applications exploiting users' knowledge)

More information

S. Dey, P. Motlicek, S. Madikeri and M. Ferras, “Exploiting Sequence Information for Text-Dependent Speaker Verification”, in: Proc. of 2017 IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, New Orleans, USA, 2017.

Software & IPR status

The above mentioned approaches to speaker verification are developed for the open-source Kaldi toolkit. The acoustic simulator is available as an open-source package from publicly available resources.

4.1.9 Speaker Segmentation and Linking

Keywords

Speaker segmentation; speaker linking;
Speaker diarization; Information Bottleneck

Key contact researcher(s)

Prof. Hervé Bourlard

herve.bourlard@idiap.ch

Technology Transfer Office

Dr. Joël Dumoulin

tto@idiap.ch

File reference & version number:

Software disclosure 6536

Functional description

The speaker segmentation and linking systems automatically discover speaker boundaries in an audio and link similar speakers across audio recordings in a database. Idiap's speaker segmentation system, also referred to as the speaker diarization system, employs the Information Bottleneck approach for faster-than-real-time, unsupervised segmentation of speakers in an audio recording. The speakers discovered are then linked across audio recordings using the state-of-the-art speaker modelling and comparison approaches. Our systems have been demonstrated to work on a variety of recording conditions such as TV shows, talks, meetings, etc.

Innovative aspects

- Rapid speaker segmentation of audio with the Information Bottleneck approach
- Longitudinal linking of speakers across audio recordings with i-vectors
- Speaker overlap detection to reduce segmentation errors

Commercial application examples

- Pre-processing for speech processing systems (e.g., for speech recognition, speaker recognition, etc.)
- Mining speakers in a large audio collection

More information

“DiarTK: An Open Source Toolkit for Research in Multi-stream Speaker Diarization and its Application to Meetings Recordings”, Deepu Vijayasenan and Fabio Valente, *Proc. of Interspeech*, 2011

Software & IPR status

The Information Bottleneck based speaker diarization system is available as a free and open source software.

- <http://www.idiap.ch/scientific-research/resources/speaker-diarization-toolkit>

4.1.10 Speaker Verification – accessing secure systems

Keywords

Speaker verification; security; forensics; tracking of criminal activities

Key contact researcher(s)

Dr. Petr Motlicek

petr.motlicek@idiap.ch

Technology Transfer Office

tto@idiap.ch

File reference & version number:

141

Functional description

Idiap's speaker verification technology can be employed as a gatekeeper for accessing the secure system. It is primarily based on iVectors – state-of-the-art technology – to generate low-dimensional but representative speaker models, followed by channel compensation technique. It is combined with a hybrid approach which successfully integrates the state-of-the-art speech recognition technology into a system that benefits from linguistic resources.

Innovative aspects

- Text-dependent technology that can be used in “voice-as-a-password” applications
- The technology is language- and channel-independent
- New speaker can be easily added by end-users

Commercial application examples

- Verification/authentication of caller against large database of speaker models
- The robustness of the system can be increased by combining with speech recognition (users' knowledge included)
- Security: voice verification, forensics, tracking of criminal activities
- The technology can be easily combined with gender and language identification

More information

Petr Motlicek et al., “Employment Of Subspace Gaussian Mixture Models in Speaker Recognition”, in: Proc. of IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP), pp. 4755–4799, 2015.

Software & IPR status

The above mentioned approaches to speaker verification are developed for the open-source Kaldi toolkit.

4.1.11 Automatic Speech Assessment

Keywords

Speech assessment; speech signal processing; Markov models

Key contact researcher(s)

Dr. Mathew Magimai Doss

Mathew.MagimaiDoss@idiap.ch

Technology Transfer Office

Dr. Joël Dumoulin

tto@idiap.ch

File reference & version number:

N/A

Functional description

Idiap conducts research on the development of automatic speech assessment methods. The R&D activities range from objective assessment of speech intelligibility to automatic accentedness evaluation of non-native speech. Automatic speech assessment reduces the need for time-consuming listening tests with human subjects and aids in developing speech communication and assistive speech technologies, such as speech coding, speech synthesis, computer aided language learning (CALL).

Innovative aspects

- Unified framework to assess speech at various levels, such as phone/phoneme, word
- Quantitative results that can be interpreted like traditional scores from human subjects

Commercial application examples

- Benchmarking of speech telecommunication or synthesis systems
- Development of assistive technologies such as CALL, pathological speech processing systems

More information

“Objective Intelligibility Assessment of Text-to-Speech Systems Through Utterance Verification”, Raphael Ullmann, Ramya Rasipuram, Mathew Magimai-Doss and Hervé Bourlard, *Proceedings of Interspeech*, Dresden, Germany, 2015.

Software & IPR status

The complete software consists of open source softwares such as HTK, Quicknet, SSP (SSP: <https://github.com/idiap/ssp>) and Idiap proprietary code that interfaces the open source tools. Speech databases and lexical resources used for the development of the assessment system may require commercial license.

Application of automatic accent evaluation for CALL is presently licensed to an industrial partner.

4.1.12 Speech-Based Just-In-Time Information Retrieval

Keywords

Just-in-time information retrieval; real-time document recommendation; speech-based retrieval; meeting assistants

Key contact researcher

Dr. Andrei Popescu-Belis

Technology Transfer Office

Dr. Joël Dumoulin

tto@idiap.ch

File reference & version number:

N/A

Functional description

We have designed a system that uses words from a conversation between users to create implicit queries, submits them to a search and retrieval system, clusters the results and displays suggestions of documents that are potentially relevant to users at a given moment in the conversation.

At the core of the system lies an algorithm that extracts a relevant and diverse set of keywords from a conversation fragment, avoiding ASR noise. These are then clustered according to conversation topics. Retrieval results from several implicit queries are merged and ranked by rewarding diversity. Moreover, we can also allow users to ask explicit queries in speech, and expand them based on the conversational context.

Innovative aspects

- Extract relevant and diverse keywords, avoiding noise
- Use topical clustering to formulate implicit queries
- Rank results by rewarding diversity
- Expand explicit queries using context

Commercial application examples

- Real-time document recommender system in meetings
- Document recommender for lecture recordings

More information

Maryam Habibi and Andrei Popescu-Belis, “Keyword Extraction and Clustering for Document Recommendation in Conversations”, in *IEEE/ACM Transactions on Audio, Speech and Language Processing*, 23(4):746–759, 2015.

Software & IPR status

Research software, demonstrator (Idiap Showroom).

4.1.13 Speech Synthesis

Keywords

Speech synthesis; text-to-speech systems; Markov models; speech signal processing

Key contact researcher(s)

Dr. Philip N. Garner

Phil.Garner@idiap.ch

Technology Transfer Office

Dr. Joël Dumoulin

tto@idiap.ch

File reference & version number:

Software disclosure 5879

Software disclosure 8700

Software disclosure 11457

Functional description

Idiap's speech synthesis technology is based on years of experience in speech recognition, and in cross fertilisation of the two technologies. We are able to provide specific voices in several languages, including English and the main Swiss languages. We can also provide average voice based TTS, that is, generic voices that can be adapted quickly to more personal voices. We are active in further merging the technologies for recognition and synthesis allowing advances in either domain to benefit the other.

Innovative aspects

- Convergence of recognition and synthesis technology
- Cross-lingual adaptation

Commercial application examples

- Speech to speech translation
- Hands free feedback
- Prosthetic voices
- Assistance for blind and partially sighted

More information

“Combining Vocal Tract Length Normalization with Hierarchical Linear Transformations”, Lakshmi Saheer, Junichi Yamagishi, Philip N. Garner and John Dines, in: *IEEE Journal of Selected Topics in Signal Processing - Special Issue on Statistical Parametric Speech Synthesis*, 8(2):262–272, 2014.

Software & IPR status

Most of the underlying software is open source. Some databases associated with certain languages may require separate agreements.

- VTLN: <http://www.idiap.ch/software/hts-vtln/>
- SSP: <https://github.com/idiap/ssp>
- HTS: <http://hts.sp.nitech.ac.jp/>
- Kaldi: <http://kaldi.sourceforge.net/about.html>

4.1.14 Speech-to-Text Transcription

Keywords

Speech synthesis; text-to-speech systems; Markov models; speech signal processing

Key contact researcher(s)

Prof. Hervé Bourlard

bourlard@idiap.ch

and

Dr. Petr. Motlicek

petr.motlicek@idiap.ch

Technology Transfer Office

Dr. Joël Dumoulin

tto@idiap.ch

File reference & version number:

Software disclosure 2701

Software disclosure 8699

Functional description

Idiap has a growing collection of state of the art software to develop leading-edge speech-to-text systems and to perform speech recognition in various scenarios. Packages include Juicer, Tracter, ISS, and Kaldi. As one of the early pioneers in the development of hybrid systems using Hidden Markov Models (HMM) and Artificial Neural networks (ANN, now referred to DNN), we have leading-edge HMM/DNN technologies and speech recognizers available in several languages. Idiap can provide customized solutions of speech transcription systems for new languages (depending on availability of at least 100 hours of well transcribed audio data). Very large vocabulary (> 100 Kwords), continuous speech, speaker independent, HMM/GMM and HMM/DNN-based speech recognition available.

Innovative aspects

- Neural network based acoustic modeling, technology allowing for speaker/channel/sentence-based adaptation.
- Language model – domain customization using data crawling from Internet.
- Offline, near real-time, and real-time solutions, depending on algorithmic delay allowed by customers.

Commercial application examples

- Dictation system customized for specific domain
- Search for specific information in large audio archives
- Audio content data-mining and indexing

More information

Motlicek, Petr, Imseng, David, Potard, Blaise, Garner, Philip N. and Himawan, Ivan, “Exploiting foreign resources for DNN-based ASR”, *EURASIP Journal on Audio, Speech, and Music Processing*, 2015.

Software & IPR status

IP lies in the copyright of the code, plus in expertise allowing us to create recognisers from suitable third party databases, dictionaries and tools. Some databases associated with certain languages may require separate agreements.

- Kaldi: <http://kaldi.sourceforge.net/about.html>

4.1.15 Very Low Bit-Rate Speech Coding

Keywords

Speech processing; speech coding; very low bit rate real-time speech communication

Key contact researcher(s)

Dr. Philip N. Garner

Phil.Garner@idiap.ch

and

Dr. Milos Cernak

Technology Transfer Office

Dr. Joël Dumoulin

tto@idiap.ch

File reference & version number:

N/A

Functional description

Idiap's very low bit-rate (VLBR) speech coding system is a conventional speech coder with an acceptable communication delay, designed for real-time speech communication. Cascaded phone-based speech recognition and synthesis systems transmit phonetic and syllabic information, encoded independently, i.e., asynchronously. It operates at an uncompressed bit rate of 213 bits/sec and achieves an average communication delay of 243 ms. A more recent version also uses phonological speech representation instead of phonetic, and is purely based on artificial neural networks, and no hidden Markov models are then used.

Innovative aspects

- Speech coding inspired by the human speech signal processing, viz., has explicit simultaneous phonetic and syllabic components, which are asynchronously related.
- The recent version transmits binary phonological speech representation, and thus is more ready for a multi-lingual use.

Commercial application examples

End-to-end speech transmission suitable for military and tactical communication systems, and for environments with highly restricted bandwidth, such as under-see communication systems.

More information

M. Cernak, P. N. Garner, A. Lazaridis, P. Motlicek, and X. Na, "Incremental Syllable-Context Phonetic Vocoding," *IEEE/ACM Trans. on Audio, Speech, and Language Processing*, Vol. 23, No. 6, pp. 1019–1030, 2015.

Software & IPR status

Most of the underlying software is open source. Some databases associated with certain languages may require separate agreements.

- <https://gitlab.idiap.ch/milos.cernak/sct-ext>
- <https://gitlab.idiap.ch/milos.cernak/phonovoc>

4.1.16 Sparse Phonological Vocoding

Keywords

Deep phonological representation learning, Structured sparse coding, Structured compressive sensing, Linguistic parsing

Key contact researcher(s)

Prof. Hervé Bourlard

Dr. Afsaneh Asaei

Dr. Milos Cernak

herve.bourlard@idiap.ch

Technology Transfer Office

Dr. Joël Dumoulin

tto@idiap.ch

File reference & version number:

Software Disclosure 11851

Functional description

Phonological representation of speech signal and multimedia in general, are sparse and their coefficients are highly structured. The underlying structured sparsity patterns are related to segmental and supra-segmental speech parameters, associated to the production and perception of speech at different time scales correlated with physiology of articulation as well as linguistics. The structured sparsity can thus be utilized for innovative technologies in the context of speech production and ineligibility assessment. They also enable highly efficient speech representation. Our developed computational platform can achieve ultra low bit speech coding and it is applicable for automatic speech segmentation or linguistic parsing as well as assessment of speech production and ineligibility.

Innovative aspects

- Deep learning based phonological vocoding
- Structured compressive sensing of phonological representations
- Structured sparse coding of phonological representation
- Class-specific codebook of diverse linguistic structures

Commercial application examples

- Low bit rate speech coding
- Automatic speech segmentation and alignment
- Speech production and ineligibility assessment
- Robust recognition of articulatory distorted speech

More information

“A. Asaei, M. Cernak, H. Bourlard, “On Compressibility of Neural Network phonological Features for Low Bit Rate Speech Coding”, proceeding of Interspeech, Dresden, Germany, September 2015.

Software & IPR status

Some databases associated with certain languages may require separate agreements.

The rights are covered by a patent application: US2015846036: Signal processing method and apparatus based on structured sparsity of phonological features.

4.1.17 Ad Hoc Microphone Arrays

Keywords

Microphone position calibration, Synchronization, Distributed source localization

Key contact researcher(s)

Prof. Hervé Bourlard

herve.bourlard@idiap.ch

and

Dr. Afsaneh Asaei

Technology Transfer Office

Dr. Joël Dumoulin

tto@idiap.ch

File reference & version number:

Software disclosure 11610

Functional description

Ubiquitous sensing exploits the abundance of microphone-embedded devices, such as notebooks and smart phones, in developing innovative technologies capable of aggregating widely sensed spatial audio. The key challenges pertain to the unknown geometry of the microphones and asynchronous recordings to enable higher level speech applications. Idiap has the prototype of a technology for calibration and synchronization of ad hoc microphones relying on properties of the structured matrices as well as the acoustic reverberation.

Innovative aspects

- Calibration of microphones from partial distances
- Synchronization of microphone recordings
- Distributed speaker localization
- Ability to handle very large network of ad hoc microphones
- Ability to deal with malfunctioning microphones

Commercial application examples

- Hands-free high-quality speech recording
- 3D audio technologies
- Sound field reproduction
- Surveillance

More information

“Enhanced diffuse field model for ad hoc microphone array calibration”, Mohammad J. Taghizadeh, Philip N. Garner and Hervé Bourlard, *Signal Processing*, Vol. 101, pp. 242-255, 2014.

Software & IPR status

The research software is available upon agreement.

4.1.18 Speaker Localization and Separation

Keywords

Microphone array, Multiparty speech analysis, Source localization

Key contact researcher(s)

Prof. Hervé Boursard

Dr. Afsaneh Asaei

herve.boursard@idiap.ch

Technology Transfer Office

Dr. Joël Dumoulin

tto@idiap.ch

File reference & version number:

Software disclosure 11610

Invention disclosure 2013-4

Functional description

Multi-party speaker localization and separation play a key role in many applications involving hands-free meeting recordings and scene analysis. Idiap's multiparty speech processing technology provides a flexible framework for multi-speaker localization and speech separation based on the recently emerged techniques of the structured sparse coding. The principles rely on the models of the auditory system in structural grouping and spatial analysis.

Innovative aspects

- Ability to pinpoint the exact source location
- Ability to separate individuals from overlapping speech
- Requirement for a few number of microphones
- Flexible in ad-hoc and constrained microphone placement

Commercial application examples

- Distant speech recognition
- Hands-free video conferencing
- Sound field reproduction
- Surveillance
- Entertainment

More information

“Structured Sparsity Models for Reverberant Speech Separation”, A. Asaei, M. Golbabaee, H. Boursard, V. Cevher, *IEEE/ACM Transactions on Speech and Audio Processing*, Vol.22, No.3, pp. 620-633, 2014.

Software & IPR status

- The reference implementation is available upon agreement.
- The rights are partly covered by a patent application: US20130096922: Method, apparatus and computer program product for determining the location of a plurality of speech sources.

4.2 Machine learning and pattern recognition

4.2.1 Signal Processing and Machine Learning Toolbox (BOB)

Keywords

Signal processing; machine learning; toolbox

Key contact researcher(s)

Dr. Sébastien Marcel
marcel@idiap.ch

Technology Transfer Office

Dr. Joël Dumoulin
tto@idiap.ch

File reference & version number:

Software disclosure 6548

Functional description

Bob is a signal-processing and machine learning toolbox. The toolbox is written in a mix of Python and C++ and is designed to be both efficient and to reduce development time.

Innovative aspects

- Multi-dimensional arrays Blitz arrays (tensors up to 11 dimensions) integration
- LAPACK integration
- HDF5 scientific file format integration
- Signal processing (FFT, DCT)
- Image processing (SIFT bridge, LBP, DCT block, Gabor, Smoothing, Optical Flow, illumination normalization)
- Biometric database and protocol support
- PCA/LDA, Linear Machines, MLP, SVM bridge, k-Means, Gaussian Mixture Models, Joint Factor Analysis, Inter-Session Variability Modeling, Probabilistic Linear Discriminant Analysis
- Video support
- Performance evaluation toolkit

Commercial application examples

- Face recognition, speaker recognition, vein recognition, multi-modal processing
- Biometrics-enabled identity management systems (Automated Border Control, Access Control, ...)
- Multi-factor authentication security systems (Critical Infrastructures, e-Banking, ...)
- Forensic Science, Video surveillance, Entertainment, Robotics, Man-Machine interaction

More information

A. Anjos, L. El Shafey, R. Wallace, M. Günther, C. McCool, and S. Marcel. Bob: a free signal processing and machine learning toolbox for researchers, *ACM Multimedia 2012 International Conference*, 2012.

Software & IPR status

The library is available under the BSD license:

- BOB: <http://www.idiap.ch/software/bob/>

4.2.2 Activity Monitoring from Ambient Sensors

Keywords

activity analysis, ambient sensors

Key contact researcher

Prof. Daniel Gatica-Perez

gatica@idiap.ch

Technology Transfer Office

Dr. Joël Dumoulin

tto@idiap.ch

File reference & version number:

NA

Functional description

We have developed several algorithms related to the analysis of activities from ambient sensors. These algorithms allow the identification of recurrent activity patterns and the detection of changes in these activity patterns. Ambient sensing represents a challenge in terms of sensor types and configurations: ambient sensors can measure different types of information and work at different scales and precisions. This heterogeneity has been addressed by developing a sensor abstraction layer that allows for transversal comparisons between activity levels estimated in different environments.

Innovative aspects

- Detection of abnormal activity patterns
- Sensor abstraction layer
- Visualisation of activity patterns

Commercial application examples

- This solution is currently used by our industrial partner to remotely monitor elderly people

More information

Ambient sensing has been addressed during the CTI funded project: DomoCare - A new Home Care Preventive Protocol.

Software & IPR status

The Software is fully owned by our industrial partner. We have acquired a significant expertise in working and mining ambient sensor data that can be applied to any scenario involving ambient activity sensing.

4.2.3 Fast Object Detector

Keywords

Object detection

Key contact researcher(s)

Dr. François Fleuret

francois.fleuret@idiap.ch

Technology Transfer Office

Dr. Joël Dumoulin

tto@idiap.ch

File reference & version number:

Software disclosure 9496

Functional description

Object detection in images is a key component of many systems able to process images automatically. This can be for semantic analysis, or as a pre-processing for identification or robotic control. In many practical situation, this operation has to be done extremely rapidly, to allow for instance the processing of video streams in real time.

We have developed a novel strategy to speed up a large class of such methods, which is close to one order of magnitude faster than the best pre-existing algorithms.

Innovative aspects

- One order of magnitude speed-up with respect to pre-existing approaches
- Technique applicable to a large class of detection methods

Commercial application examples

- Pedestrian detection for surveillance in video stream
- Low-power object detection systems

More information

C. Dubout and F. Fleuret. Exact Acceleration of Linear Object Detectors. In Proceedings of the European Conference on Computer Vision (ECCV), pages 301-311, 2012.

Software & IPR status

The reference implementation is available under the GPL3 license.

- FFLD: <https://www.idiap.ch/scientific-research/resources/exact-acceleration-of-linear-object-detectors>
- The algorithm is covered by US patent (US20140089365) on “Object detection method, object detector and object detection computer program”.

4.2.4 Large-scale Image Retrieval

Keywords

Content-based image retrieval; large scale

Key contact researcher(s)

Dr. François Fleuret

francois.fleuret@idiap.ch

Technology Transfer Office

Dr. Joël Dumoulin

tto@idiap.ch

File reference & version number:

Software disclosure 6006

Functional description

Iterative Relevance Feedback is a software technology to facilitate search in very large image databases. The core idea is to provide the user with an interactive interface which displays candidate image among which he has to pick the most relevant. From that choice, the system update the scores attached to every image, and propose a new selection. The approach we develop relies on a sophisticated statistical model which maximize the information gathered from the user at every step.

Our technology can cope with very large databases. Concurrent methods with the same statistical models are limited to a few tens of thousands of images. Our technique allows to scale this up by two orders on a single PC, and has to potential to scale by several other orders of magnitude on a dedicated cluster of machines.

Innovative aspects

- Use an adaptive granularity for the image set representation
- Scales gracefully to very large data-sets

Commercial application examples

- On-line product search
- Interactive person identification
- Image bank browsing

More information

N. Suditu and F. Fleuret. Adaptive relevance feedback for large-scale image retrieval. Multimedia Tools and Applications (MTA), 2015. To appear.

Software & IPR status

The reference implementation is available under the GPL3 license:

- HEAT system: <http://www.idiap.ch/scientific-research/resources/heat-image-retrieval-system>

4.2.5 Multi-camera Detection and Tracking

Keywords

Multi-camera; pedestrian detection; surveillance; behavior analysis

Key contact researcher(s)

Dr. François Fleuret

francois.fleuret@idiap.ch

Technology Transfer Office

Dr. Joël Dumoulin

tto@idiap.ch

File reference & version number:

Software disclosure 7632

Functional description

This technology allows the detection of pedestrians or vehicles from multiple synchronized video streams taken from calibrated cameras looking at a common area of interest. It can either performs the detection in real-time in separate frames, or process a batch of frames to exploit temporal consistency to increase the accuracy.

Innovative aspects

- Robust system with a limited number of parameters
- Can handle very a degraded signal
- Consistent estimates in case of occlusion
- Meaningful probabilistic prediction

Commercial application examples

- Video-surveillance
- Behavioral analysis
- Crowd counting

More information

J. Berclaz, F. Fleuret, E. Turetken, and P. Fua, “Multiple Object Tracking using K-Shortest Paths Optimization,” *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 33(9):1806-1819, 2011.

Software & IPR status

- An implementation of the tracker (“Multi-Tracked Paths”) is available under the GPL3 license: <http://www.idiap.ch/scientific-research/resources/mtp>
- The algorithms were developed in collaboration with the CVLab at EPFL and are covered by two (EPFL) patents:
 - International patent WO2013072401 on “Tracklet-based Multi-Commodity Network Flow for Tracking Multiple People”.
 - US patent US20130177200 on “A method and apparatus for multiple object tracking with k-shortest paths”.

4.2.6 Unsupervised Activity Discovery, Monitoring, and Abnormality Detection

Keywords

Activity Discovery, Domotic, Surveillance, Abnormality detection.

Key contact researcher(s)

Dr. Jean-Marc Odobez
odobez@idiap.ch

Technology Transfer Office

Dr. Joël Dumoulin
tto@idiap.ch

File reference & version number:

Software disclosure 11637

Functional description

Idiap had developed methods and software for the unsupervised discovery of recurrent temporal pattern in multivariate time series, where observed values are caused by the superposition of multiple phenomena.

This typical situation occurs when multiple sensors records the activities of multiple objects/people, like multimodal sensors (proximity, water, light sensors, etc) in domotics, multiple cameras -or even microphone-arrays- in indoor/outdoor complex scenes.

Discovered patterns correspond to recurrent activities like trajectories of different object classes in cameras, or sequences of activated sensors in domotics. In addition, the method allows the identification of abnormal situations, and can be exploited to select streams to be displayed in control rooms of large public spaces (to direct the attention of operators towards relevant information).

Innovative aspects

- Capture sub-activities temporal ordering
- No Prior on scene needed
- Small number of parameter to set
- Estimate automatically the number of activities

Commercial application examples

- Activity discovery from large multi-sensor datasets
- Surveillance (anomaly detection, stream selection)
- Video content filtering (summary generation)
- Adaptation to user's behaviors (heating,...)

More information

- R. Emonet, J. Varadarajan and J.-M. Odobez. Temporal Analysis of Motif Mixtures using Dirichlet Processes. *IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI)*, 36(1):140-156, Jan. 2014.
- R. Emonet and J.-M. Odobez. Unsupervised Methods for Activity Analysis and Detection of Abnormal Events. in *Intelligent Video Surveillance Systems (ISTE)*. Wiley 2013.

Software status

- Recurrent pattern learning <http://www.idiap.ch/resource/research/probamod-v1/>
- More complete prototype for surveillance videos

4.2.7 Scalable linear discriminant analysis for large scale pattern recognition

Keywords

An efficient and scalable version of probabilistic linear discriminant analysis (PLDA).

Key contact researcher(s)

Dr. Sébastien Marcel
marcel@idiap.ch

Technology Transfer Office

Dr. Joël Dumoulin
tto@idiap.ch

File reference & version number:

Software disclosure 7088

Functional description

This implements probabilistic linear discriminant analysis (PLDA). The implementation is a scalable and efficient implementation.

Innovative aspects

The commercial application is potentially widespread as it provides an exact and scalable way of training and computing the likelihood for PLDA. Thus, it is applicable to many pattern recognition tasks involving discriminant analysis and it has already been shown to be applicable to face and speaker recognition tasks to obtain state-of-the-art performance. Other general pattern recognition and object recognition applications are also foreseen.

Commercial application examples

- Object recognition
- Biometrics-enabled identity management systems (Automated Border Control, Access Control, ...)
- Multi-factor authentication security systems (Critical Infrastructures, e-Banking, ...)
- Forensic Science, Video surveillance, Entertainment, Robotics, Man-Machine interaction

More information

L. Shafey, C. McCool, R. Wallace and S. Marcel: “A Scalable Formulation of Probabilistic Linear Discriminant Analysis: Applied to Face Recognition”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35, (7), pp. 1788–1794. http://publications.idiap.ch/downloads/papers/2013/ElShafey_TPAMI_2013.pdf

Software & IPR status

Open sourced at <https://pypi.python.org/pypi/xbob.paper.tpami2013> but re-licensing possible for commercial purposes.

4.2.8 Soft-skill analytics from online video job interviews

Keywords

online video, nonverbal behavior, job interviews, crowdsourcing

Key contact researcher

Prof. Daniel Gatica-Perez
gatica@idiap.ch

Technology Transfer Office

Dr. Joël Dumoulin
tto@idiap.ch

File reference & version number:

Invention disclosure 2015-01

Functional description

A system to profile, filter, screen, compare, search, and rank job candidates on the basis of soft-skills analytics. Soft-skills is an umbrella term to refer to a cluster of personality traits, communication, language, interpersonal skills, teamwork, leadership, etc. that characterize people in the working place, that are as relevant as technical or hard skills, and are typically not shown in a resume. We integrate online-video recordings, a crowdsourcing system that uses a crowd to collect a ground-truth of soft-skills, a set of machine learning algorithms to estimate soft-skills based on annotations and automatic methods that extract nonverbal and verbal behavioral cues, and a suit of machine machine learning algorithms to compare, search, and rank job candidates on the basis of soft-skills.

Innovative aspects

- Crowdsourcing of high-agreement soft-skill annotations
- Consistent measures of soft skills on the basis of behavioral video data
- Scales the understanding of soft-skills from personal video collections

Commercial application examples

- Online video job interviewing processes

More information

J.-I. Biel and D. Gatica-Perez, “The YouTube Lens: Crowdsourced Personality Impressions and Audiovisual Analysis of Vlogs”, in *IEEE Transactions on Multimedia*, Vol. 15, No. 1, pp. 41-55, Jan. 2013.

Software & IPR status

- Internal Invention Disclosure

4.3 Computer vision

4.3.1 Cardiac Imaging

Keywords

Biological microscopy, heart, cardiac, image processing, image reconstruction, dynamic imaging, movie synchronization, 3D imaging

Key contact researcher(s)

Dr. Michael Liebling

michael.liebling@idiap.ch

Technology Transfer Office

Dr. Joël Dumoulin

tto@idiap.ch

File reference & version number:

N/A

Functional description

Development of imaging protocols and reconstruction algorithms for 3D cardiac imaging.

Innovative aspects

- Volumetric and dynamic imaging of the heart
- Extraction of cardiac function parameters
- Minimal user interaction

Commercial application examples

- Combined microscopy-software tools for live biological imaging
- 3D image reconstruction from multiple views

Software & IPR status

The underlying software consists of a set of *Matlab*, *C*, and *Java* tools to perform problem-specific microscopy acquisition protocols, digital image reconstruction of dynamic 3D volumes, analysis and display routines.

Note on IPR: Ongoing efforts, part of software suite was developed while at Caltech and UCSB. Some of the methods are covered by patent:

M. Dickinson, A. Farouhar, S. E. Fraser, M. Gharib, and M. Liebling, “Four-dimensional imaging of periodically moving objects via post-acquisition synchronization of nongated slice-sequences,” U.S. Patent No. 7,372,984, May 13, 2008.

4.3.2 Real-time Multiple face and head pose tracking

Keywords

Computer Vision, Face tracking, Head orientation

Key contact researcher(s)

Dr. Jean-Marc Odobez
odobez@idiap.ch

Technology Transfer Office

Dr. Joël Dumoulin
tto@idiap.ch

File reference & version number:

Software disclosure 7164
Software disclosure 7165

Functional description

Idiap has developed a real-time system that can track continuously from 1 to 5 faces, which in addition can extract for each detected face the head orientation, providing some hint on where the person is looking at.

The technology does not only rely on face detection but also tracking, since due to the large variety of head poses that people can take (e.g. when looking down at paper, or in some profile views) face detection is rapidly insufficient when a person does not look straight at the camera.

Innovative aspects

- Robust to background noise and illumination variability can tolerate camera motion
- Track faces even when they are not frontal
- Provide head pose cue, and visual focus of attention cues (e.g. are people looking at the camera)
- Robust track creation and deletion mechanisms for lower false alarm rates.

Commercial application examples

Tracking people faces is a prerequisite in many applications before further behavior analysis, e.g. when interacting implicitly (e.g. in the remote conferencing systems) or explicitly (e.g. with a robot) with a device.

- HCI application with one-to-many interactions
- Communication systems (e.g. webcam) that can exploit behavior cues for communication improvement (e.g. zooming on individuals)
- Behavior analysis study systems.

More information

- S. Duffner and J.-M. Odobez. A Track Creation and Deletion Framework for Long-Term Online Multi-Face Tracking. *IEEE Transaction on Image Processing*, March 2013.
- S. Ba and J.-M. Odobez. Recognizing Visual Focus of Attention from Head Pose in Natural Meetings. *IEEE Trans. on System, Man and Cybernetics: part B, Cybernetics*, Vol. 39. No. 1. Feb 2009.

Software & IPR status

A demonstration software is available.

4.3.3 Motion Detection, Segmentation, and Estimation. Motion Satbilization

Keywords

Computer Vision, Video Stabilization, Motion analysis in videos.

Key contact researcher(s)

Dr. Jean-Marc Odobez
odobez@idiap.ch

Technology Transfer Office

Dr. Joël Dumoulin
tto@idiap.ch

File reference & version number:

Licence 20090914-A02

Functional description

Idiap has expertise and software on video motion analysis, which is important in both the surveillance and video content analysis domains. Robust motion estimation is often a way to stabilize shaky videos, and can be exploited as such or prior to further video content analysis tasks.

Innovative aspects

- Robust automatic estimation of dominant motion (works even in presence of secondary motion sources)
- Accurate image motion estimation
- Proven efficiency in both consumer video application (video stabilization) and industrial background (Alcatel, Ifremer,...)
- Real-time

Commercial application examples

- Video Stabilization (e.g. from mobile or table video acquisition)
- Surveillance

More information

J-M. Odobez and P. Bouthemy. Robust multiresolution estimation of parametric motion model. *Journal of Visual Communication and Image Representation*, Vol. 6, Number 4 pp 348-365, December 1995.

Software status

- Commercial motion estimation and stabilization software available.
- Non-exclusive license granted to Cinetis SA.

4.3.4 Gaze Estimation and Non-Verbal Behavior Extraction from consumer cameras

Keywords

Computer Vision, Gaze, Behavior analysis, Interactions, HCI, HRI.

Key contact researcher(s)

Dr. Jean-Marc Odobez
odobez@idiap.ch

Technology Transfer Office

Dr. Joël Dumoulin
tto@idiap.ch

File reference & version number:

Software disclosure 10352
Software disclosure 10353
Invention disclosure 2013-06

Functional description

We have developed softwares for and a long-standing experience with the automatic analysis of a variety of real-life human (non-verbal) behaviors in interaction modeling (attention, head gestures, addressee, engagement,...).

Using RGBD sensors, we have developed precise eye gaze estimation systems comprising the fitting of personalized 3D mesh face models, a real-time estimation head pose tracker, and gaze estimation modules.

When relying on RGB cameras only (typically at further distance from the sensor), multiperson real-time head pose trackers have been developed.

In both domains, software have been developed to infer the attention of people towards visual targets (people, objects or scene regions), and recognize head gestures.

Innovative aspects

- No user restriction (non-intrusive system)
- Generic or personalized head and eye gaze models
- Handles gaze tracking under head pose changes
- 3D reasoning (not only for screen-based applications)
- Works under low-resolution eye imaging

Commercial application examples

- Attention modeling (e.g. consumer preference analysis)
- User studies, eg in sociology (NVB and gaze coding; dyadic and group interactions analysis)

More information

- K. Funes and J.-M. Odobez. Geometric Generative Gaze Estimation (G3E) for Remote RGB-D Cameras. Computer Vision and Pattern Recognition conference (CVPR), 2014.
- K. Funes, L. Nguyen, D. Gatica-Perez and J.-M. Odobez. A Semi-Automated System for Accurate Gaze Coding in Natural Dyadic Interactions Proc. Int. Conf. on Multimodal Interfaces (ICMI), 2013.
- L. Nguyen, J.-M. Odobez, and D. Gatica-Perez. Using Self-Context for Multimodal Detection of Head Nods in Face-to-Face Interactions ICMI, 2012.

Software and IPR status

- RGBD simplified version - <http://www.idiap.ch/scientific-research/resources/hg3d>)
- Gaze tracking under patenting

4.3.5 Static Luggage Detection and Intrusion Detection

Keywords

Computer Vision, Security, Surveillance, Activity Monitoring, Luggage Detection

Key contact researcher(s)

Dr. Jean-Marc Odobez
odobez@idiap.ch

Technology Transfer Office

Dr. Joël Dumoulin
tto@idiap.ch

File reference & version number:

Software disclosure 11636

Functional description

Detecting changes in videos is an essential component of surveillance applications. Idiap has expertise in this domain, with efficient background subtraction schemes and further exploitation of this technology for automatic activity analysis from outdoor or indoor scenes.

Several modules have been developed in this direction. The first one detects and analyses changes in perimeters and space surveillance application, allowing for detecting intruders. The second detects static abandoned luggage left by passengers, including the detection of potential people remaining close to these objects.

Innovative aspects

- Robust multi-cue background subtraction scheme
- Light robust multi-tracker scheme
- Combination of long-term cues (background) and dynamic cues (tracklets)

Commercial application examples

- Public space monitoring (luggage detection)
- Perimeter or indoor monitoring (intrusion detection)

More information

- J. Yao and J.-M. Odobez. Multi-Layer Background Subtraction Based on Color and Texture. in Proc. CVPR Visual Surveillance workshop, June 2007.
- A. Lopez-Mendez, F. Monay and J.-M. Odobez. Exploiting Scene Cues for Dropped Object Detection. Int. Conf. on Computer Vision Theory and Applications, Jan. 2014.

Software status

Demonstration software validated in real conditions.

4.4 Biometrics

4.4.1 Face recognition library and awarded ISV technology

Keywords

Face recognition library

Key contact researcher(s)

Dr. Sébastien Marcel

marcel@idiap.ch

Technology Transfer Office

Dr. Joël Dumoulin

tto@idiap.ch

File reference & version number:

Software disclosure 9141

Functional description

The Face Recognition Library (FaceRecLib) is designed to perform a fair comparison of face recognition algorithms. It contains scripts to execute various kinds of face recognition experiments on a variety of facial image databases, and running baseline algorithms.

Innovative aspects

The library contains interfaces to many publicly available facial image databases, and default evaluation protocols. Default implementations of face recognition algorithms, which rely on BOB, are provided, e.g., for: Eigenfaces, Linear Discriminant Analysis, Probabilistic Linear Discriminant Analysis (PLDA), Local Gabor Binary Pattern Histogram Sequences, Gabor Graph Matching, Gaussian Mixture Modeling, Inter-Session Variability Modeling (ISV), and the Bayesian Intrapersonal/Extrapersonal Classifier.

Commercial application examples

- Biometrics-enabled identity management systems (Automated Border Control, Access Control, ...)
- Multi-factor authentication security systems (Critical Infrastructures, e-Banking, ...)
- Forensic Science, Video surveillance, Entertainment, Robotics, Man-Machine interaction

More information

- M. Günther, R. Wallace and S. Marcel: “An Open Source Framework for Standardized Comparisons of Face Recognition Algorithms”, ECCV – Workshops and Demonstrations, 2012. http://publications.idiap.ch/downloads/papers/2012/Gunther_BEFIT2012_2012.pdf
- C. McCool, R. Wallace, M. McLaren, L. El Shafey and S. Marcel: “Session variability modelling for face authentication”, IET Biometrics, 2013, 2, (3), pp. 117–129. <http://publications.idiap.ch/index.php/publications/show/2606> **IET Biometrics Premium Award 2015**

Software & IPR status

Open sourced at <https://pypi.python.org/pypi/bob.bio.face> but re-licensing possible for commercial purposes.

4.4.2 Fingervein recognition

Keywords

Fingervein recognition library

Key contact researcher(s)

Dr. Sébastien Marcel

marcel@idiap.ch

Technology Transfer Office

Dr. Joël Dumoulin

tto@idiap.ch

File reference & version number:

Software disclosure 10264

Functional description

This library is designed to perform a fair comparison of finger vein recognition algorithms.

Innovative aspects

The library contains interfaces to publicly available fingervein image databases, and default evaluation protocols. Default implementations of vein recognition algorithms, which rely on BOB, are provided, e.g., for: Maximum Curvature, Repeated Line Tracking and Wide Line Detector.

Commercial application examples

- Biometrics-enabled identity management systems (Automated Border Control, Access Control, ...)
- Multi-factor authentication security systems (Critical Infrastructures, e-Banking, ...)

More information

P. Tome, M. Vanoni and S. Marcel: “On the Vulnerability of Finger Vein Recognition to Spoofing”, IEEE International Conference of the Biometrics Special Interest Group (BIOSIG), 2012. <http://publications.idiap.ch/index.php/publications/show/2910>

Software & IPR status

Open sourced at <https://pypi.python.org/pypi/xbob.fingervein> but re-licensing possible for commercial purposes.

4.4.3 A method of detecting a falsified presentation to a vascular recognition system

Keywords

vascular biometrics, anti-spoofing

Key contact researcher(s)

Dr. Sébastien Marcel

marcel@idiap.ch

Technology Transfer Office

Dr. Joël Dumoulin

tto@idiap.ch

File reference & version number:

Invention disclosure 2013-07

Functional description

Biometric vascular recognition is the recognition of an individual through the vascular (veins and arteries) pattern of their hand/finger. If the biometric is stolen it has been shown that it is possible to create a fake. Presenting a fake biometric on the sensor of a biometric recognition system is denoted as spoofing or presentation attack. Anti-spoofing or presentation attack detection, also termed as liveness detection, is the process of detecting a fake biometric.

This invention disclosure describes processes to detect a presentation attack with a fake vascular biometric.

Innovative aspects

The invention is based on a multi-algorithm and multi-spectrum framework processing both Visual spectra and Near-InfraRed spectra static and dynamic images.

Commercial application examples

- Biometrics-enabled identity management systems (Automated Border Control, Access Control, ...)
- Multi-factor authentication security systems (Critical Infrastructures, e-Banking, ...)

More information

On request (patent pending)

4.4.4 Palmvein recognition library

Keywords

Palmvein recognition library

Key contact researcher(s)

Dr. Sébastien Marcel

marcel@idiap.ch

Technology Transfer Office

Dr. Joël Dumoulin

tto@idiap.ch

File reference & version number:

Software disclosure 11285

Functional description

This library is designed to perform a fair comparison of palm vein recognition algorithms.

Innovative aspects

The library contains interfaces to publicly available palm vein image databases, and default evaluation protocols. Default implementations of vein recognition algorithms, which rely on BOB, are provided, e.g., for: Maximum Curvature, Repeated Line Tracking, Wide Line Detector and Local Binary Pattern Histogram Sequences.

Commercial application examples

- Biometrics-enabled identity management systems (Automated Border Control, Access Control, ...)
- Multi-factor authentication security systems (Critical Infrastructures, e-Banking, ...)

More information

P. Tome and S. Marcel: “On the Vulnerability of Palm Vein Recognition to Spoofing Attacks”, The 8th IAPR International Conference on Biometrics (ICB), 2015. http://publications.idiap.ch/downloads/papers/2015/Tome_ICB2015-SpoofingPalmvein.pdf

Software & IPR status

Open sourced at <https://pypi.python.org/pypi/bob.palmvein> but re-licensing possible for commercial purposes.

4.4.5 Speaker recognition library and award winning technology

Keywords

Speaker recognition library

Key contact researcher(s)

Dr. Sébastien Marcel

marcel@idiap.ch

Technology Transfer Office

Dr. Joël Dumoulin

tto@idiap.ch

File reference & version number:

Software disclosure 9142

Functional description

This library is designed to run speaker verification experiments.

Innovative aspects

Different state-of-the-art algorithms for voice activity detection, feature extraction (MFCC, LFCC), modelling techniques (Gaussian Mixture Models, Session Variability Modelling, Total Variability Modelling – iVectors) and database interfaces are implemented.

Commercial application examples

- Biometrics-enabled identity management systems (Automated Border Control, Access Control, ...)
- Multi-factor authentication security systems (Critical Infrastructures, e-Banking, ...)
- Forensic Science, Video surveillance, Entertainment, Robotics, Man-Machine interaction

More information

- E. Khoury, L. El-Shafey and S. Marcel: “Spear: An open source toolbox for speaker recognition based on BOB”, IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP), 2014. http://publications.idiap.ch/downloads/papers/2014/Khoury_ICASSP_2014.pdf
- E. Khoury, M. Ferras, L. El-Shafey and S. Marcel: “Hierarchical speaker clustering methods for the nist i-vector challenge”, In Odyssey: The Speaker and Language Recognition Workshop, 2014. http://publications.idiap.ch/downloads/papers/2014/Khoury_ODYSSEY_2014.pdf **NIST Speaker i-vector Machine Learning Challenge 2014**

Software & IPR status

Open sourced at <https://pypi.python.org/pypi/bob.bio.spear> but re-licensing possible for commercial purposes.

4.5 Multimedia processing

4.5.1 Hybrid Recommender System for Multimedia

Keywords

Recommender system; content-based recommendation; collaborative filtering; hybrid recommender system; sentiment analysis; multimedia recommendation

Key contact researcher

Dr. Andrei Popescu-Belis

Technology Transfer Office

Dr. Joël Dumoulin
tto@idiap.ch

File reference & version number:

Software disclosure 10782

Functional description

This system performs generic and personalized recommendation of multimedia content: typically A/V lectures, but also images or videos. We have identified the metadata most relevant for content-based (CB) recommendation. We have defined a method to combine CB and collaborative filtering (CF), which is applicable to both cold-start and non-cold start settings. In one-class CF problems (when users only mark items as favorites, or ignore them), we can complement user action information with user sentiment extracted from user-generated texts such as comments. Furthermore, we can perform fine-grained sentiment analysis in terms of aspects, learning to detect which sentences of reviews or comments are likely to refer to given aspects, for review segmentation and summarization.

Innovative aspects

- Hybrid content-based / collaborative filtering recommender
- Complement one-class ratings with sentiment of comments
- Aspect-based analysis of reviews, extracting most representative sentences

Commercial application examples

- Recommender systems for large or small collections
- Explanation of recommendations based on user reviews

More information

Nikolaos Pappas and Andrei Popescu-Belis, “Combining Content with User Preferences for Non-Fiction Multimedia Recommendation: A Study on TED Lectures”, *Multimedia Tools and Applications*, Vo.74, No.4, pp. 1175–1197, 2015.

Software & IPR status

Open sources:

- CBRec v1.0: <https://github.com/idiap/cbrec>
- EMORec v1.0: <https://github.com/idiap/cbrec>

4.5.2 Machine Translation at the Document Level

Keywords

Statistical machine translation; discourse structure; connectives; pronouns; verbs

Key contact researcher

Dr. Andrei Popescu-Belis

Technology Transfer Office

Dr. Joël Dumoulin

tto@idiap.ch

File reference & version number:

Software disclosure 8452

Software disclosure 9011

Functional description

This technology improves current systems for automatic translation, by using text-level information. Several pre-processors detect inter-sentence dependencies and use them to label different types of words, prior to translation. The labels are used by a third-party statistical phrase-based machine translation system (open source) which otherwise translates sentence-by-sentence. We currently deal with discourse connectives (e.g. ‘since’ or ‘while’), pronouns (‘it’ vs. ‘il’ or ‘elle’) and verb tenses in English, and with Chinese and German compounds. Our method improves the translation of entire texts.

Innovative aspects

- Coupling of discourse-level classifiers with SMT
- Learning to detect semantic features for connectives, pronouns, compounds, and verbs from parallel corpora

Commercial application examples

- Improved MT systems for long coherent texts
- Post-editing tools for existing MT engines

More information

Thomas Meyer, Najeh Hajlaoui, and Andrei Popescu-Belis, “Disambiguating Discourse Connectives for Statistical Machine Translation”, *IEEE/ACM Transactions on Audio, Speech and Language Processing*, 23(7):1184–1197, 2015.

Software & IPR status

- Open source software for translation of connectives <https://github.com/idiap/DiscoConn-Classifier>
- and their evaluation <https://github.com/idiap/act>

Research software and know-how for other types of words (verbs, pronouns, compounds).

4.5.3 Text detection and recognition in images and videos

Keywords

Computer Vision, text detection, text recognition, video content analysis.

Key contact researcher(s)

Dr. Jean-Marc Odobez
odobez@idiap.ch

Technology Transfer Office

Dr. Joël Dumoulin
tto@idiap.ch

File reference & version number:

Licence 20121015-A01

Functional description

Consumer pictures and broadcast videos often contain text that provide semantic information about the media content (e.g. name of a person appearing in the video).

Idiap has developed a set of robust tools for detecting and transcribing the text present in such media, which have been validated in several European research projects and Swiss technology transfer projects.

Innovative aspects

- Works with non black/white content (colored text, text superimposed on textured background)
- Multi-hypothesis text segmentation and transcription
- High text recall

Commercial application examples

- Broadcast data analysis
- Scanned document analysis

More information

D. Chen, J-M. Odobez and H. Bourlard. Text Detection and Recognition in Images and Videos. Pattern Recognition, Vol. 37, pp 595-608, 2004.

Software status

A demonstration software is available.

IPR status

A non-exclusive license has been granted to Klewel SA.

4.5.4 Social context inference from proximity sensors

Keywords

proximity sensors, mobile phones, wearables

Key contact researcher

Prof. Daniel Gatica-Perez

gatica@idiap.ch

Technology Transfer Office

Dr. Joël Dumoulin

tto@idiap.ch

File reference & version number:

N/A

Functional description

A method to infer social context from proximity sensors like bluetooth, for example embedded on mobile phones but also in wearable devices. The method uses longitudinal proximity data generated by a population of mobile device users and generates, for each member of the population, the groups of people with whom each user is in proximity, and the typical times when this occurs. The method can be used for group discovery or for event mining applications.

Innovative aspects

- Statistical machine learning approach that discovers groups and their temporal context from raw observations of physical proximity.
- The method is applicable to various types of sensor data related to physical proximity.

Commercial application examples

- Group discovery and event mining in large gatherings.
- Organizational mapping.

More information

T. Do and D. Gatica-Perez, “Human Interaction Discovery in Smartphone Proximity Networks,” *Personal and Ubiquitous Computing*, Vol. 17, Issue 3, pp. 413-431, Mar. 2013.

Software & IPR status

- Matlab code

4.6 Robotics

4.6.1 Robot Skills Transfer Toolbox

Keywords

Human-robot skills transfer; robot programming by demonstration; interfaces for robot programming; movement primitives; optimal control; adaptive motion planning; gesture synthesis

Key contact researcher(s)

Dr. Sylvain Calinon

sylvain.calinon@idiap.ch

Technology Transfer Office

Dr. Joël Dumoulin

tto@idiap.ch

File reference & version number:

Software disclosure 11619

Functional description

PbDlib is a collection of source codes for robot programming by demonstration (learning from demonstration). It is used to develop intuitive interface that can be used by laypeople to transfer skills to robots. This toolset includes various functionalities at the crossroad of statistical learning, dynamical systems, optimal control and Riemannian geometry.

PbDlib can be used in applications requiring task adaptation, human-robot skill transfer, safe controllers based on minimal intervention principle, adaptive motion planning. It relies on probabilistic representations that can be used for motion analysis and synthesis in multiple coordinate systems, as well as for the automatic analysis and generation of movements, gestures and manipulation skills.

Innovative aspects

- Layman interface for robot programming
- Automatic adaptation of movements to new situations
- Handling of task variations and options
- Safe robot controlled by minimal intervention principle

Commercial application examples

- Interface for human-robot skill transfer
- Adaptive planning
- Optimal control learned from human demonstrations
- Motion analysis and synthesis

More information

- <http://www.idiap.ch/software/pbdlib/>
- Calinon, S. and Lee, D. (2019). Learning Control. Vadakkepat, P. and Goswami, A. (eds.). Humanoid Robotics: a Reference. Springer.
- Calinon, S. (2016). A Tutorial on Task-Parameterized Movement Learning and Retrieval. Intelligent Service Robotics (Springer), 9:1, 1-29.

Software & IPR status

The underlying software comes in three distinct and independent versions in *Matlab*, *C++* and *Python*, with minimal dependencies to facilitate its integration to other softwares. The *Matlab* version is fully compatible with the *GNU Octave* open source software.



Contact

Dr. Joël Dumoulin

tto@idiap.ch



Centre du Parc, Rue Marconi 19, CH-1920 Martigny
T: +41 27 721 77 11 | info@idiap.ch | www.idiap.ch