Activity Report 2018

Project-Team STARS

Spatio-Temporal Activity Recognition Systems

RESEARCH CENTER
Sophia Antipolis - Méditerranée

THEME
Vision, perception and multimedia interpretation

Table of contents

Project-Team STARS

Creation of the Team: 2012 January 01, updated into Project-Team: 2013 January 01

Keywords:

Computer Science and Digital Science:

A2.1.9. - Synchronous languages
A2.1.11. - Proof languages
A2.3.3. - Real-time systems
A2.4.2. - Model-checking
A2.4.3. - Proofs
A2.5. - Software engineering
A3.2.1. - Knowledge bases
A3.3.2. - Data mining
A3.4.1. - Supervised learning
A3.4.2. - Unsupervised learning
A3.4.5. - Bayesian methods
A3.4.6. - Neural networks
A4.7. - Access control
A5.1. - Human-Computer Interaction
A5.3.2. - Sparse modeling and image representation
A5.3.3. - Pattern recognition
A5.4.1. - Object recognition
A5.4.2. - Activity recognition
A5.4.3. - Content retrieval
A5.4.5. - Object tracking and motion analysis
A9.1. - Knowledge
A9.2. - Machine learning
A9.3. - Signal analysis

Other Research Topics and Application Domains:

B1.2.2. - Cognitive science
B2.1. - Well being
B7.1.1. - Pedestrian traffic and crowds
B8.1. - Smart building/home
B8.4. - Security and personal assistance

1. Team, Visitors, External Collaborators

Research Scientists

Francois Brémond [Team leader, Inria, Senior Researcher, HDR]
Sabine Moisan [Inria, Researcher, HDR]
Annie Ressouche [Inria, Researcher, until Jan 2018]
Jean-Paul Rigault [Univ de Nice - Sophia Antipolis, Emeritus]
Monique Thonnat [Inria, Senior Researcher, HDR]

Antitza Dantcheva [Inria, Researcher, until December 2019]

Faculty Member

Frederic Precioso [Univ de Nice - Sophia Antipolis, Associate Professor, from Mar 2018 until Sep 2018]

Post-Doctoral Fellows

Abhijit Das [Inria]
S L Happy [Inria, from Sep 2018]
Alexandra Konig [Inria]
Furqan Muhammad Khan [Inria, until Apr 2018]
Michal Koperski [Inria, until May 2018]

PhD Students

Srijan Das [Univ de Nice - Sophia Antipolis]
Juan Diego Gonzales Zuniga [KONTRON, from Apr 2018]
S L Happy [Inria, until Jun 2018]
Jen Cheng Hou [Inria, from Nov 2018]
Thibaud Lyvonnet [Inria, from Dec 2018]
Farhood Negin [Inria, until Sep 2018]
Thi Lan Anh Nguyen [Inria, until May 2018]
Ines Sarray [Inria, until Dec 2018]
Ujjwal Ujjwal [VEDECOM]
Yaohui Wang [Inria]

Technical staff

Abdelrahman Gaber Abubakr [Inria, until Sep 2018]
Rui Dai [Inria, from Oct 2018]
Sebastien Gilabert [Inria, from Dec 2018]
Soumik Mallick [Inria]
Hung Nguyen [Inria, until May 2018]
Minh Khue Phan Tran [Inria, granted by BPIFRANCE FINANCEMENT SA]

Interns

Lea Baudon [Speech Therapy School, Nice, from Oct 2018 until Nov 2018]
Valentin Charlet [Speech Therapy School, Nice, from Oct 2018 until Nov 2018]
Arpit Chaudhary [Inria, from May 2018 until Aug 2018]
Beatrice Eula Fantozzi [Speech Therapy School, Nice, from Oct 2018 until Nov 2018]
Abhishek Goel [Inria, until Feb 2018]
Stefan Kostic [Campus ID School, Sophia Antipolis, from Jul 2018 until Aug 2018]
Kuan Ru Lee [Inria, until Feb 2018]
Aimen Neffati [Campus ID School, Sophia Antipolis, from Jul 2018 until Sep 2018]
Kaustubh Sanjay Sakhalkar [Inria, until May 2018]
Kaustav Tamuly [Inria, from May 2018 until Jul 2018]
Vikas Thamizharasan [Inria, from Aug 2018]
Dimitri Wyzlic [Inria, from Apr 2018 until Sep 2018]
Abdelrahman Gaber Abubakr [Univ Grenoble Alpes, from Oct 2018]
Hung Nguyen [Institut Telecom ex GET Groupe des Ecoles des Télécommunications, from Aug 2018, Jun 2018]

Administrative Assistant

Laurence Briffa [Inria]

Visiting Scientists

Nagi Aly [Université de Nouakchott, Mauritanie, from Apr 2018 until Sep 2018]
Carlos Antonio Caetano Junior [Universidade Federal de Minas Gerais, Brasil, from Feb 2018 until May 2018]
Adlen Kerboua [Université 20 Aout 55 Skikda, Algérie, from May 2018 until Jun 2018]
Xue Le [Laboratoire I3S, Sophia Antipolis, until Jan 2018]

Ion Mosnoi [Univ de Nice - Sophia Antipolis, until Jun 2018]

External Collaborators

Hao Chen [ESI, from Sep 2018]
Daniel Gaffe [Univ de Nice - Sophia Antipolis, from Feb 2018]
Sebastien Gilabert [Campus ID School, Sophia Antipolis, from Mar 2018 until Nov 2018]
Juan Diego Gonzales Zuniga [KONTRON, from Feb 2018 until Apr 2018]
Annie Ressouche [Retired, from Feb 2018 until Dec 2018]
Philippe Robert [Nice Hospital, until May 2018]
Jean-Yves Tigli [Univ de Nice - Sophia Antipolis, from Apr 2018 until Mar 2018]
Piotr Tadeusz Bilinski [University of Oxford, from Jul 2018, until Jun 2018]
Carlos-Fernando Crispim Junior [Univ Lumière, from Jul 2018 until Jun 2018]
Elisabetta de Maria [Univ de Nice - Sophia Antipolis, from Sep 2018]
Baptiste Fosty [EKINNOX, from Mar 2018 until Feb 2018]
Rachid Guerchouche [Nice Hospital, from Feb 2018]

2. Overall Objectives

2.1. Presentation

The **STARS (Spatio-Temporal Activity Recognition Systems)** team focuses on the design of cognitive vision systems for Activity Recognition. More precisely, we are interested in the real-time semantic interpretation of dynamic scenes observed by video cameras and other sensors. We study long-term spatio-temporal activities performed by agents such as human beings, animals or vehicles in the physical world. The major issue in semantic interpretation of dynamic scenes is to bridge the gap between the subjective interpretation of data and the objective measures provided by sensors. To address this problem Stars develops new techniques in the field of cognitive vision and cognitive systems for physical object detection, activity understanding, activity learning, vision system design and evaluation. We focus on two principal application domains: visual surveillance and healthcare monitoring.

2.1.1. Research Themes

Stars is focused on the design of cognitive systems for Activity Recognition. We aim at endowing cognitive systems with perceptual capabilities to reason about an observed environment, to provide a variety of services to people living in this environment while preserving their privacy. In today world, a huge amount of new sensors and new hardware devices are currently available, addressing potentially new needs of the modern society. However the lack of automated processes (with no human interaction) able to extract a meaningful and accurate information (i.e. a correct understanding of the situation) has often generated frustrations among the society and especially among older people. Therefore, Stars objective is to propose novel autonomous systems for the **real-time semantic interpretation of dynamic scenes** observed by sensors. We study long-term spatio-temporal activities performed by several interacting agents such as human beings, animals and vehicles in the physical world. Such systems also raise fundamental software engineering problems to specify them as well as to adapt them at run time.

We propose new techniques at the frontier between computer vision, knowledge engineering, machine learning and software engineering. The major challenge in semantic interpretation of dynamic scenes is to bridge the gap between the task dependent interpretation of data and the flood of measures provided by sensors. The problems we address range from physical object detection, activity understanding, activity learning to vision system design and evaluation. The two principal classes of human activities we focus on, are assistance to older adults and video analytic.

A typical example of a complex activity is shown in Figure 1 and Figure 2 for a homecare application. In this example, the duration of the monitoring of an older person apartment could last several months. The activities involve interactions between the observed person and several pieces of equipment. The application goal is to recognize the everyday activities at home through formal activity models (as shown in Figure 3) and data captured by a network of sensors embedded in the apartment. Here typical services include an objective assessment of the frailty level of the observed person to be able to provide a more personalized care and to monitor the effectiveness of a prescribed therapy. The assessment of the frailty level is performed by an Activity Recognition System which transmits a textual report (containing only meta-data) to the general practitioner who follows the older person. Thanks to the recognized activities, the quality of life of the observed people can thus be improved and their personal information can be preserved.

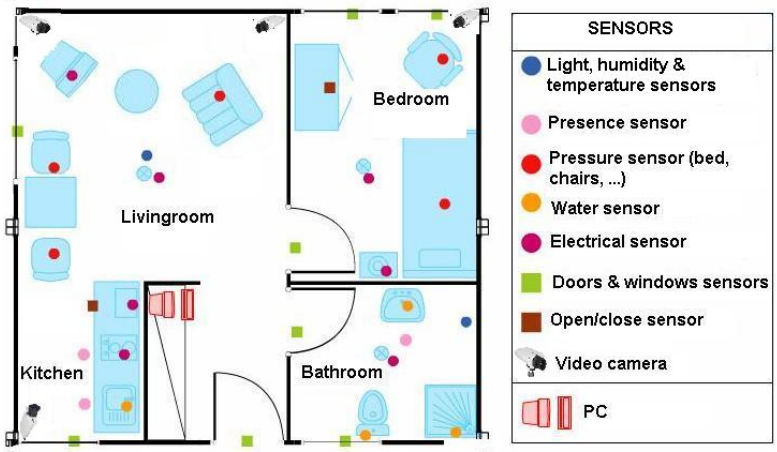


Figure 1. Homecare monitoring: the set of sensors embedded in an apartment

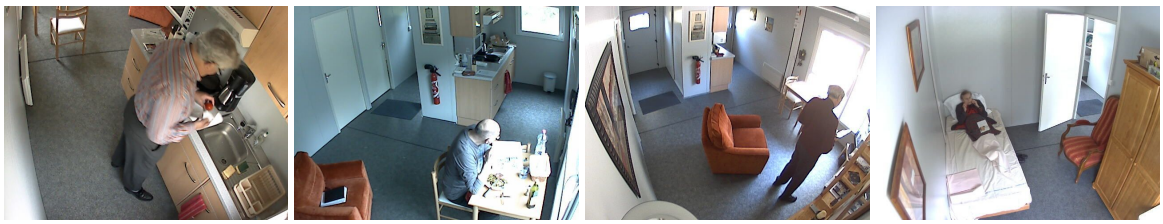


Figure 2. Homecare monitoring: the different views of the apartment captured by 4 video cameras

The ultimate goal is for cognitive systems to perceive and understand their environment to be able to provide appropriate services to a potential user. An important step is to propose a computational representation of people activities to adapt these services to them. Up to now, the most effective sensors have been video cameras due to the rich information they can provide on the observed environment. These sensors are currently perceived as intrusive ones. A key issue is to capture the pertinent raw data for adapting the services to the people while preserving their privacy. We plan to study different solutions including of course the local processing of the data without transmission of images and the utilization of new compact sensors developed

Activity (*PrepareMeal*,
PhysicalObjects((p : Person), (z : Zone), (eq : Equipment))
Components((s_inside : InsideKitchen(p, z))
(s_close : CloseToCountertop(p, eq))
(s_stand : PersonStandingInKitchen(p, z)))
Constraints((z->Name = Kitchen)
(eq->Name = Countertop)
(s_close->Duration >= 100)
(s_stand->Duration >= 100))
Annotation(AText("prepare meal")
AType("not urgent")))

Figure 3. Homecare monitoring: example of an activity model describing a scenario related to the preparation of a meal with a high-level language

for interaction (also called RGB-Depth sensors, an example being the Kinect) or networks of small non visual sensors.

2.1.2. International and Industrial Cooperation

Our work has been applied in the context of more than 10 European projects such as COFRIEND, ADVISOR, SERKET, CARETAKER, VANAHEIM, SUPPORT, DEM@CARE, VICOMO. We had or have industrial collaborations in several domains: *transportation* (CCI Airport Toulouse Blagnac, SNCF, Inrets, Alstom, Ratp, GTT (Italy), Turin GTT (Italy)), *banking* (Crédit Agricole Bank Corporation, Eurotelis and Ciel), *security* (Thales R&T FR, Thales Security Syst, EADS, Sagem, Bertin, Alcatel, Keeneo), *multimedia* (Multitel (Belgium), Thales Communications, Idiap (Switzerland)), *civil engineering* (Centre Scientifique et Technique du Bâtiment (CSTB)), *computer industry* (BULL), *software industry* (AKKA), *hardware industry* (ST-Microelectronics) and *health industry* (Philips, Link Care Services, Vistek).

We have international cooperations with research centers such as Reading University (UK), ENSI Tunis (Tunisia), National Cheng Kung University, National Taiwan University (Taiwan), MICA (Vietnam), IPAL, I2R (Singapore), University of Southern California, University of South Florida, University of Maryland (USA).

3. Research Program

3.1. Introduction

Stars follows three main research directions: perception for activity recognition, semantic activity recognition, and software engineering for activity recognition. **These three research directions are interleaved**: *the software engineering* research direction provides new methodologies for building safe activity recognition systems and *the perception* and *the semantic activity recognition* directions provide new activity recognition techniques which are designed and validated for concrete video analytic and healthcare applications. Conversely, these concrete systems raise new software issues that enrich the software engineering research direction.

Transversely, we consider a *new research axis in machine learning*, combining a priori knowledge and learning techniques, to set up the various models of an activity recognition system. A major objective is to automate model building or model enrichment at the perception level and at the understanding level.

3.2. Perception for Activity Recognition

Participants: François Brémond, Sabine Moisan, Monique Thonnat.

: Activity Recognition, Scene Understanding, Machine Learning, Computer Vision, Cognitive Vision Systems, Software Engineering

3.2.1. Introduction

Our main goal in perception is to develop vision algorithms able to address the large variety of conditions characterizing real world scenes in terms of sensor conditions, hardware requirements, lighting conditions, physical objects, and application objectives. We have also several issues related to perception which combine machine learning and perception techniques: learning people appearance, parameters for system control and shape statistics.

3.2.2. Appearance Models and People Tracking

An important issue is to detect in real-time physical objects from perceptual features and predefined 3D models. It requires finding a good balance between efficient methods and precise spatio-temporal models. Many improvements and analysis need to be performed in order to tackle the large range of people detection scenarios.

Appearance models. In particular, we study the temporal variation of the features characterizing the appearance of a human. This task could be achieved by clustering potential candidates depending on their position and their reliability. This task can provide any people tracking algorithms with reliable features allowing for instance to (1) better track people or their body parts during occlusion, or to (2) model people appearance for re-identification purposes in mono and multi-camera networks, which is still an open issue. The underlying challenge of the person re-identification problem arises from significant differences in illumination, pose and camera parameters. The re-identification approaches have two aspects: (1) establishing correspondences between body parts and (2) generating signatures that are invariant to different color responses. As we have already several descriptors which are color invariant, we now focus more on aligning two people detection and on finding their corresponding body parts. Having detected body parts, the approach can handle pose variations. Further, different body parts might have different influence on finding the correct match among a whole gallery dataset. Thus, the re-identification approaches have to search for matching strategies. As the results of the re-identification are always given as the ranking list, re-identification focuses on learning to rank. "Learning to rank" is a type of machine learning problem, in which the goal is to automatically construct a ranking model from a training data.

Therefore, we work on information fusion to handle perceptual features coming from various sensors (several cameras covering a large scale area or heterogeneous sensors capturing more or less precise and rich information). New 3D RGB-D sensors are also investigated, to help in getting an accurate segmentation for specific scene conditions.

Long term tracking. For activity recognition we need robust and coherent object tracking over long periods of time (often several hours in videosurveillance and several days in healthcare). To guarantee the long term coherence of tracked objects, spatio-temporal reasoning is required. Modeling and managing the uncertainty of these processes is also an open issue. In Stars we propose to add a reasoning layer to a classical Bayesian framework modeling the uncertainty of the tracked objects. This reasoning layer can take into account the a priori knowledge of the scene for outlier elimination and long-term coherency checking.

Controlling system parameters. Another research direction is to manage a library of video processing programs. We are building a perception library by selecting robust algorithms for feature extraction, by insuring they work efficiently with real time constraints and by formalizing their conditions of use within a program supervision model. In the case of video cameras, at least two problems are still open: robust image segmentation and meaningful feature extraction. For these issues, we are developing new learning techniques.

3.3. Semantic Activity Recognition

Participants: François Brémond, Sabine Moisan, Monique Thonnat.

: Activity Recognition, Scene Understanding, Computer Vision

3.3.1. Introduction

Semantic activity recognition is a complex process where information is abstracted through four levels: signal (e.g. pixel, sound), perceptual features, physical objects and activities. The signal and the feature levels are characterized by strong noise, ambiguous, corrupted and missing data. The whole process of scene understanding consists in analyzing this information to bring forth pertinent insight of the scene and its dynamics while handling the low level noise. Moreover, to obtain a semantic abstraction, building activity models is a crucial point. A still open issue consists in determining whether these models should be given a priori or learned. Another challenge consists in organizing this knowledge in order to capitalize experience, share it with others and update it along with experimentation. To face this challenge, tools in knowledge engineering such as machine learning or ontology are needed.

Thus we work along the following research axes: high level understanding (to recognize the activities of physical objects based on high level activity models), learning (how to learn the models needed for activity recognition) and activity recognition and discrete event systems.

3.3.2. High Level Understanding

A challenging research axis is to recognize subjective activities of physical objects (i.e. human beings, animals, vehicles) based on a priori models and objective perceptual measures (e.g. robust and coherent object tracks).

To reach this goal, we have defined original activity recognition algorithms and activity models. Activity recognition algorithms include the computation of spatio-temporal relationships between physical objects. All the possible relationships may correspond to activities of interest and all have to be explored in an efficient way. The variety of these activities, generally called video events, is huge and depends on their spatial and temporal granularity, on the number of physical objects involved in the events, and on the event complexity (number of components constituting the event).

Concerning the modeling of activities, we are working towards two directions: the uncertainty management for representing probability distributions and knowledge acquisition facilities based on ontological engineering techniques. For the first direction, we are investigating classical statistical techniques and logical approaches. For the second direction, we built a language for video event modeling and a visual concept ontology (including color, texture and spatial concepts) to be extended with temporal concepts (motion, trajectories, events ...) and other perceptual concepts (physiological sensor concepts ...).

3.3.3. Learning for Activity Recognition

Given the difficulty of building an activity recognition system with a priori knowledge for a new application, we study how machine learning techniques can automate building or completing models at the perception level and at the understanding level.

At the understanding level, we are learning primitive event detectors. This can be done for example by learning visual concept detectors using SVMs (Support Vector Machines) with perceptual feature samples. An open question is how far can we go in weakly supervised learning for each type of perceptual concept (i.e. leveraging the human annotation task). A second direction is to learn typical composite event models for frequent activities using trajectory clustering or data mining techniques. We name composite event a particular combination of several primitive events.

3.3.4. Activity Recognition and Discrete Event Systems

The previous research axes are unavoidable to cope with the semantic interpretations. However they tend to let aside the pure event driven aspects of scenario recognition. These aspects have been studied for a long time at a theoretical level and led to methods and tools that may bring extra value to activity recognition, the most important being the possibility of formal analysis, verification and validation.

We have thus started to specify a formal model to define, analyze, simulate, and prove scenarios. This model deals with both absolute time (to be realistic and efficient in the analysis phase) and logical time (to benefit from well-known mathematical models providing re-usability, easy extension, and verification). Our purpose is to offer a generic tool to express and recognize activities associated with a concrete language to specify activities in the form of a set of scenarios with temporal constraints. The theoretical foundations and the tools being shared with Software Engineering aspects, they will be detailed in section 3.4.

The results of the research performed in perception and semantic activity recognition (first and second research directions) produce new techniques for scene understanding and contribute to specify the needs for new software architectures (third research direction).

3.4. Software Engineering for Activity Recognition

Participants: Sabine Moisan, Annie Ressouche, Jean-Paul Rigault, François Brémond.

: Software Engineering, Generic Components, Knowledge-based Systems, Software Component Platform, Object-oriented Frameworks, Software Reuse, Model-driven Engineering

The aim of this research axis is to build general solutions and tools to develop systems dedicated to activity recognition. For this, we rely on state-of-the art Software Engineering practices to ensure both sound design and easy use, providing genericity, modularity, adaptability, reusability, extensibility, dependability, and maintainability.

This research requires theoretical studies combined with validation based on concrete experiments conducted in Stars. We work on the following three research axes: *models* (adapted to the activity recognition domain), *platform architecture* (to cope with deployment constraints and run time adaptation), and *system verification* (to generate dependable systems). For all these tasks we follow state of the art Software Engineering practices and, if needed, we attempt to set up new ones.

3.4.1. Platform Architecture for Activity Recognition

In the former project teams Orion and Pulsar, we have developed two platforms, one (VSIP), a library of real-time video understanding modules and another one, LAMA [14], a software platform enabling to design not only knowledge bases, but also inference engines, and additional tools. LAMA offers toolkits to build and to adapt all the software elements that compose a knowledge-based system.

Figure 4 presents our conceptual vision for the architecture of an activity recognition platform. It consists of three levels:

- The **Component Level**, the lowest one, offers software components providing elementary operations and data for perception, understanding, and learning.
  - *Perception components* contain algorithms for sensor management, image and signal analysis, image and video processing (segmentation, tracking...), etc.
  - *Understanding components* provide the building blocks for Knowledge-based Systems: knowledge representation and management, elements for controlling inference engine strategies, etc.
  - *Learning components* implement different learning strategies, such as Support Vector Machines (SVM), Case-based Learning (CBL), clustering, etc.

  An Activity Recognition system is likely to pick components from these three packages. Hence, tools must be provided to configure (select, assemble), simulate, verify the resulting component combination. Other support tools may help to generate task or application dedicated languages or graphic interfaces.
- The **Task Level**, the middle one, contains executable realizations of individual tasks that will collaborate in a particular final application. Of course, the code of these tasks is built on top of the components from the previous level. We have already identified several of these important tasks: Object Recognition, Tracking, Scenario Recognition... In the future, other tasks will probably enrich this level.

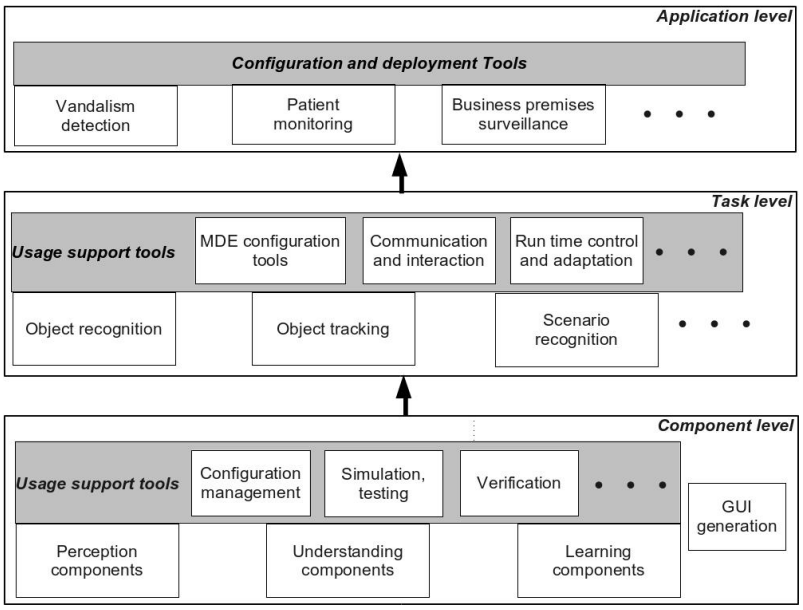


Figure 4. Global Architecture of an Activity Recognition The gray areas contain software engineering support modules whereas the other modules correspond to software components (at Task and Component levels) or to generated systems (at Application level).

For these tasks to nicely collaborate, communication and interaction facilities are needed. We shall also add MDE-enhanced tools for configuration and run-time adaptation.

- The **Application Level** integrates several of these tasks to build a system for a particular type of application, e.g., vandalism detection, patient monitoring, aircraft loading/unloading surveillance, etc.. Each system is parameterized to adapt to its local environment (number, type, location of sensors, scene geometry, visual parameters, number of objects of interest...). Thus configuration and deployment facilities are required.

The philosophy of this architecture is to offer at each level a balance between the widest possible genericity and the maximum effective reusability, in particular at the code level.

To cope with real application requirements, we shall also investigate distributed architecture, real time implementation, and user interfaces.

Concerning implementation issues, we shall use when possible existing open standard tools such as NuSMV for model-checking, Eclipse for graphic interfaces or model engineering support, Alloy for constraint representation and SAT solving for verification, etc. Note that, in Figure 4, some of the boxes can be naturally adapted from SUP existing elements (many perception and understanding components, program supervision, scenario recognition...) whereas others are to be developed, completely or partially (learning components, most support and configuration tools).

3.4.2. Discrete Event Models of Activities

As mentioned in the previous section (3.3) we have started to specify a formal model of scenario dealing with both absolute time and logical time. Our scenario and time models as well as the platform verification tools rely on a formal basis, namely the synchronous paradigm. To recognize scenarios, we consider activity

descriptions as synchronous reactive systems and we apply general modeling methods to express scenario behavior.

Activity recognition systems usually exhibit many safeness issues. From the software engineering point of view we only consider software security. Our previous work on verification and validation has to be pursued; in particular, we need to test its scalability and to develop associated tools. Model-checking is an appealing technique since it can be automatized and helps to produce a code that has been formally proved. Our verification method follows a compositional approach, a well-known way to cope with scalability problems in model-checking.

Moreover, recognizing real scenarios is not a purely deterministic process. Sensor performance, precision of image analysis, scenario descriptions may induce various kinds of uncertainty. While taking into account this uncertainty, we should still keep our model of time deterministic, modular, and formally verifiable. To formally describe probabilistic timed systems, the most popular approach involves probabilistic extension of timed automata. New model checking techniques can be used as verification means, but relying on model checking techniques is not sufficient. Model checking is a powerful tool to prove decidable properties but introducing uncertainty may lead to infinite state or even undecidable properties. Thus model checking validation has to be completed with non exhaustive methods such as abstract interpretation.

3.4.3. Model-Driven Engineering for Configuration and Control and Control of Video Surveillance systems

Model-driven engineering techniques can support the configuration and dynamic adaptation of video surveillance systems designed with our SUP activity recognition platform. The challenge is to cope with the many—functional as well as nonfunctional—causes of variability both in the video application specification and in the concrete SUP implementation. We have used *feature models* to define two models: a generic model of video surveillance applications and a model of configuration for SUP components and chains. Both of them express variability factors. Ultimately, we wish to automatically generate a SUP component assembly from an application specification, using models to represent transformations [58]. Our models are enriched with intra- and inter-models constraints. Inter-models constraints specify models to represent transformations. Feature models are appropriate to describe variants; they are simple enough for video surveillance experts to express their requirements. Yet, they are powerful enough to be liable to static analysis [69]. In particular, the constraints can be analyzed as a SAT problem.

An additional challenge is to manage the possible run-time changes of implementation due to context variations (e.g., lighting conditions, changes in the reference scene, etc.). Video surveillance systems have to dynamically adapt to a changing environment. The use of models at run-time is a solution. We are defining adaptation rules corresponding to the dependency constraints between specification elements in one model and software variants in the other [57], [74], [72].

4. Application Domains

4.1. Introduction

While in our research the focus is to develop techniques, models and platforms that are generic and reusable, we also make effort in the development of real applications. The motivation is twofold. The first is to validate the new ideas and approaches we introduce. The second is to demonstrate how to build working systems for real applications of various domains based on the techniques and tools developed. Indeed, Stars focuses on two main domains: **video analytic** and **healthcare monitoring**.

4.2. Video Analytics

Our experience in video analytic [6], [1], [8] (also referred to as visual surveillance) is a strong basis which ensures both a precise view of the research topics to develop and a network of industrial partners ranging from end-users, integrators and software editors to provide data, objectives, evaluation and funding.

For instance, the Keeneo start-up was created in July 2005 for the industrialization and exploitation of Orion and Pulsar results in video analytic (VSIP library, which was a previous version of SUP). Keeneo has been bought by Digital Barriers in August 2011 and is now independent from Inria. However, Stars continues to maintain a close cooperation with Keeneo for impact analysis of SUP and for exploitation of new results.

Moreover new challenges are arising from the visual surveillance community. For instance, people detection and tracking in a crowded environment are still open issues despite the high competition on these topics. Also detecting abnormal activities may require to discover rare events from very large video data bases often characterized by noise or incomplete data.

4.3. Healthcare Monitoring

Since 2011, we have initiated a strategic partnership (called CobTek) with Nice hospital [62], [75] (CHU Nice, Prof P. Robert) to start ambitious research activities dedicated to healthcare monitoring and to assistive technologies. These new studies address the analysis of more complex spatio-temporal activities (e.g. complex interactions, long term activities).

4.3.1. Research

To achieve this objective, several topics need to be tackled. These topics can be summarized within two points: finer activity description and longitudinal experimentation. Finer activity description is needed for instance, to discriminate the activities (e.g. sitting, walking, eating) of Alzheimer patients from the ones of healthy older people. It is essential to be able to pre-diagnose dementia and to provide a better and more specialized care. Longer analysis is required when people monitoring aims at measuring the evolution of patient behavioral disorders. Setting up such long experimentation with dementia people has never been tried before but is necessary to have real-world validation. This is one of the challenge of the European FP7 project Dem@Care where several patient homes should be monitored over several months.

For this domain, a goal for Stars is to allow people with dementia to continue living in a self-sufficient manner in their own homes or residential centers, away from a hospital, as well as to allow clinicians and caregivers remotely provide effective care and management. For all this to become possible, comprehensive monitoring of the daily life of the person with dementia is deemed necessary, since caregivers and clinicians will need a comprehensive view of the person's daily activities, behavioral patterns, lifestyle, as well as changes in them, indicating the progression of their condition.

4.3.2. Ethical and Acceptability Issues

The development and ultimate use of novel assistive technologies by a vulnerable user group such as individuals with dementia, and the assessment methodologies planned by Stars are not free of ethical, or even legal concerns, even if many studies have shown how these Information and Communication Technologies (ICT) can be useful and well accepted by older people with or without impairments. Thus one goal of Stars team is to design the right technologies that can provide the appropriate information to the medical carers while preserving people privacy. Moreover, Stars will pay particular attention to ethical, acceptability, legal and privacy concerns that may arise, addressing them in a professional way following the corresponding established EU and national laws and regulations, especially when outside France. Now, Stars can benefit from the support of the COERLE (Comité Opérationnel d'Evaluation des Risques Légaux et Ethiques) to help it to respect ethical policies in its applications.

As presented in 3.1, Stars aims at designing cognitive vision systems with perceptual capabilities to monitor efficiently people activities. As a matter of fact, vision sensors can be seen as intrusive ones, even if no images are acquired or transmitted (only meta-data describing activities need to be collected). Therefore new communication paradigms and other sensors (e.g. accelerometers, RFID, and new sensors to come in the future) are also envisaged to provide the most appropriate services to the observed people, while preserving their privacy. To better understand ethical issues, Stars members are already involved in several ethical organizations. For instance, F. Brémond has been a member of the ODEGAM - "Commission Ethique et Droit" (a local association in Nice area for ethical issues related to older people) from 2010 to 2011 and a

member of the French scientific council for the national seminar on “La maladie d’Alzheimer et les nouvelles technologies - Enjeux Ã©thiques et questions de société” in 2011. This council has in particular proposed a chart and guidelines for conducting researches with dementia patients.

For addressing the acceptability issues, focus groups and HMI (Human Machine Interaction) experts, will be consulted on the most adequate range of mechanisms to interact and display information to older people.

5. Highlights of the Year

5.1. Highlights of the Year

5.1.1. Awards

Abhijit Das, Antitza Dantcheva and Francois Brémond were winners of the Bias Estimation in Face Analytics (BEFA) Challenge at the European Conference on Computer Vision (ECCV 2018).

6. New Software and Platforms

6.1. SUP

Scene Understanding Platform
KEYWORDS: Activity recognition - 3D - Dynamic scene
FUNCTIONAL DESCRIPTION: SUP is a software platform for perceiving, analyzing and interpreting a 3D dynamic scene observed through a network of sensors. It encompasses algorithms allowing for the modeling of interesting activities for users to enable their recognition in real-world applications requiring high-throughput.

- Participants: Etienne Corvée, François Brémond, Thanh Hung Nguyen and Vasanth Bathrinarayanan
- Partners: CEA - CHU Nice - USC Californie - Université de Hamburg - I2R
- Contact: François Brémond
- URL: https://team.inria.fr/stars/software

6.2. VISEVAL

FUNCTIONAL DESCRIPTION: ViSEval is a software dedicated to the evaluation and visualization of video processing algorithm outputs. The evaluation of video processing algorithm results is an important step in video analysis research. In video processing, we identify 4 different tasks to evaluate: detection, classification and tracking of physical objects of interest and event recognition.

- Participants: Bernard Boulay and François Brémond
- Contact: François Brémond
- URL: http://www-sop.inria.fr/teams/pulsar/EvaluationTool/ViSEvAl_Description.html

7. New Results

7.1. Introduction

This year Stars has proposed new results related to its three main research axes : perception for activity recognition, semantic activity recognition and software engineering for activity recognition.

7.1.1. Perception for Activity Recognition

Participants: François Brémond, Juan Diego Gonzales Zuniga, Abhijit Das, Antitza Dancheva, Furqan Muhammad Khan, Michal Koperski, Thi Lan Anh Nguyen, Remi Trichet, Ujjwal Ujjval, Srijan Das, Vikas Thamizharasan, Monique Thonnat.

The new results for perception for activity recognition are:

- Late Fusion of multiple convolutional layers for pedestrian detection (see 7.2)
- Deep Learning applied on Embedded Systems for People Tracking (see 7.3)
- Cross Domain Residual Transfer Learning for Person Re-identification (see 7.4)
- Face-based Attribute Classification (see 7.5)
- Face Attribute manipulation
- From attribute-labels to faces: face generation using a conditional generative adversarial network (see 7.6)
- Face analysis in structured light images (see 7.7)

7.1.2. Semantic Activity Recognition

Participants: François Brémond, Antitza Dantcheva, Farhood Negin, Thanh Hung Nguyen, Michal Koperski, Srijan Das, Kaustubh Sakhalkar, Arpit Chaudhary, Abhishek Goel, Abdelrahman Abubakr, Abhijit Das, Yaohui Wang, S L Happy, Alexandra König, Guillaume Sacco, Philippe Robert, Soumik Mallick, Julien Badie, Monique Thonnat.

For this research axis, the contributions are :

- Deep-Temporal LSTM for Daily Living Action Recognition (see 7.8)
- A New Hybrid Architecture for Human Activity Recognition from RGB-D videos (see 7.10)
- Where to focus on for Human Action Recognition? (see 7.11)
- Online temporal detection of daily-living human activities in long untrimmed video streams (see 7.12)
- Activity Detection in Long-term Untrimmed Videos (see 7.13)
- Video based face analysis for health monitoring (see 7.14)
- Mobile biometrics (see 7.15)
- Comparing methods for assessment of facial dynamics in patients with major neurocognitive disorders (see 7.16)
- Combating the issue of low sample size in facial expression recognition (see 7.17)
- Serious exergames for Cognitive Stimulation (see 7.18)
- Fully Automatic Speech-Based Analysis of the Semantic Verbal Fluency Task (see 7.19)
- Language Modelling in the Clinical Semantic Verbal Fluency Task (see 7.19.2)
- Telephone-based Dementia Screening I: Automated Semantic Verbal Fluency Assessment (see 7.19.3)
- Automatic Detection of Apathy using Acoustic Markers extracted from Free Emotional Speech and using Automatic Speech Analysis (see 7.19.4)
- Monitoring the Behaviors of Retail Customers (see 7.20)

7.1.3. Software Engineering for Activity Recognition

Participants: Sabine Moisan, Annie Ressouche, Jean-Paul Rigault, Ines Sarray, Daniel Gaffé, Julien Badie, François Brémond, Minh Khue Phan Tran.

The contributions for this research axis are:

- A Synchronous Approach to Activity Recognition (see 7.21)
- A Probabilistic Activity Description Language (see 7.22)

7.2. Late Fusion of Multiple Convolutional Layers for Pedestrian Detection

Participants: Ujjwal Ujjwal, François Brémond, Aziz Dziri [VEDECOM], Bertrand Leroy [VEDECOM].

One of the prominent problems in pedestrian detection is handling scale and occlusion. These problems are quite well aligned with the recent interests in autonomous vehicles. Successful detection of far-scale pedestrians can assist the vehicle in making safety maneuvers well ahead in time, thereby promoting a safer traffic environment. The same is true for surveillance systems in high security environment like airports and ports.

We propose a system design for pedestrian detection by leveraging the power of multiple convolutional layers explicitly (see Figure 5). We quantify the effect of different convolutional layers on the detection of pedestrians of varying scales and occlusion level. We show that earlier convolutional layers are better at handling small-scale and partially occluded pedestrians. We take cue from these conclusions and propose a pedestrian detection system design based on Faster-RCNN which leverages multiple convolutional layers by late fusion. In our design, we introduce height-awareness in the loss function to make the network emphasize on pedestrian heights which are misclassified during the training process. The proposed system design achieves a log-average miss-rate of 9.25% on the caltech-reasonable dataset. This is within 1.5% of the current state-of-art approach, while being a more compact system. The work was published in the 15^{th} IEEE International Conference on Advanced Video and Signal-based Surveillance (AVSS)-2018 [51].

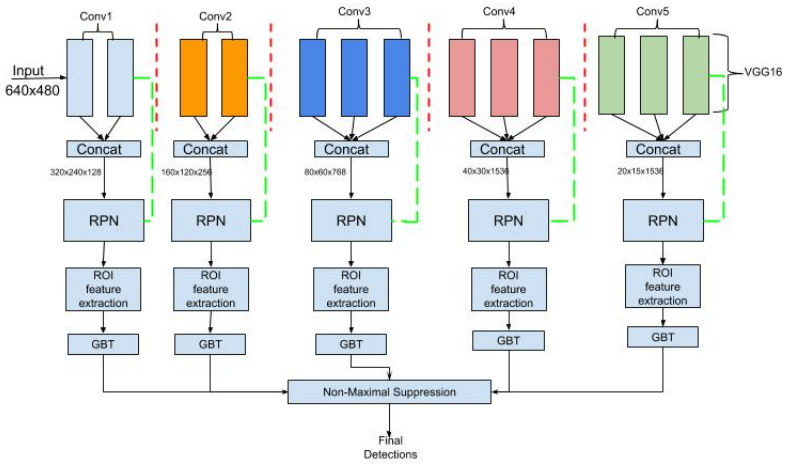


Figure 5. Block diagram of our proposed Multiple-RPN pedestrian detection system

7.3. Deep Learning applied on Embedded Systems for People Tracking

Participants: Juan Diego Gonzales Zuniga, Thi Lan Anh Nguyen, Francois Brémond, Serge Tissot [KONTRON].

Keywords: Deep Learning, Embedded Systems, Multiple Object Tracking

One of the main issues with people detection and tracking is the amount of resources it consumes for real time applications. Most architectures either require great amounts of memory or large computing time to achieve a state-of-the-art performance, these results are mostly achieved with dedicated hardware at data centers. The applications for an embedded hardware with these capabilities are limitless: automotive, security and surveillance, augmented reality and health-care just to name a few. But the state-of-the-art architectures are mostly focused on accuracy rather than resource consumption.

In our work, we have to consider improving the systems' accuracy and reducing resources for real-time applications. We are creating a shared effort of hardware adaptation and agnostic software optimization for all deep learning based solutions.

We here focus our work on two separated but linked problems.

First, we improve the feature representation of tracklets for the Multiple Object Tracking challenge. This is based on the concept of Residual Transfer Learning [44]. Second, we are creating a viable platform to run our algorithms on different target hardware, mainly, Intel Xeon Processors, FPGAs and AMD GPUs.

7.3.1. Residual Transfer Learning :

We present a smart training alternative for transfer learning based on the concept of ResNet [65]. In ResNet, a layer learns the estimate residual between the input and output signals. We cast transfer learning as a residual learning problem, since the objective is to close the gap between the initial network and the desired one. Achieving this goal is done by adding residual units for a number of layers to an existing model that needs to be transferred from one task to another. The existing model can thus be able to perform a new task by adding and optimizing residual units as shown in Figure 6. The main advantage of using residual units for transfer learning is the flexibility in terms of modelling the difference between two tasks.

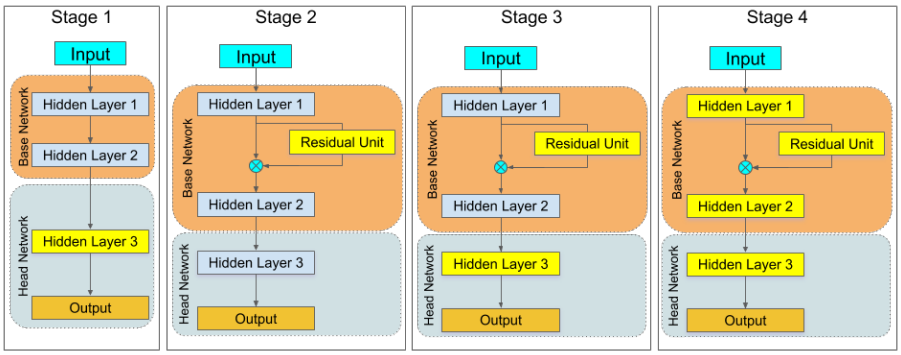


Figure 6. Training stages of Residual Transfer Learning method. Each stage only trains the layers shown in yellow, and fixes the layers in grey. The residual units are added at the second stage

7.3.2. Deep Learning Platform on Multiple Target Hardware :

Deep learning algorithms need an extensive allocation of resources to be executed, most of the research is accomplished under NVIDIA GPU's. This is limiting because it reduces the possibilities on how to optimize certain blocks that directly depend on the hardware configuration. The main cause is the lack of a flexible platform that would support different targets: AMD GPUs, Intel Xeon processors and specialized FPGAs.

We work with two hardware based platforms; ROCm and Openvino. The ROCm stack, shown in Figure 7, allows us to perform a variety of layer computations on AMD GPUs. We have managed to import different deep learning networks such as VGG16, ResNet and Inception to AMD's Radeon graphics card. On the other

hand, Openvino's main goal is to reduce the inference time of a network. For this solution, we count on the Openvino Optimizer, shown in Figure 8, which main goal is to transform the network model from Caffe or Tensorflow into an Inference Model for Intel's processors and FPGAs.

We also built docker images on top of the above mention platforms, this is done to speed the deployment stage by being operating system independent.

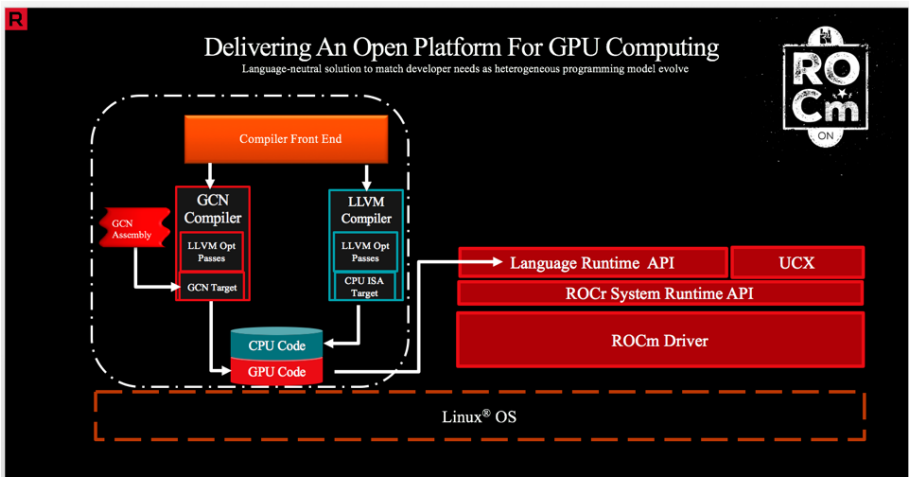


Figure 7. The ROCm System Runtime is language independent and makes heavy use of the Heterogeneous System Architecture.

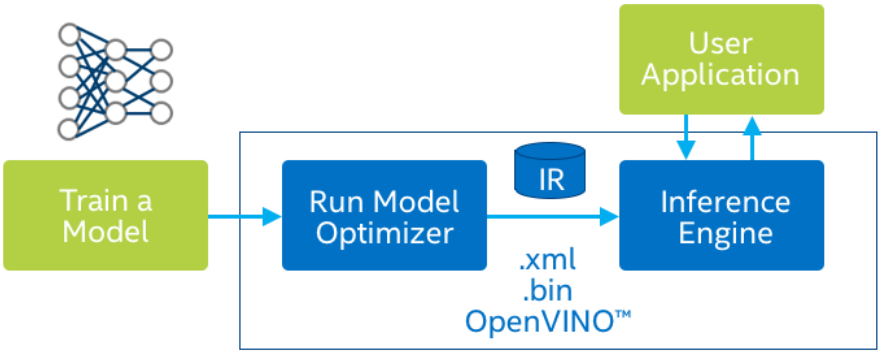


Figure 8. Openvino: When you run a pre-trained model through the Model Optimizer, your output is an Intermediate Representation of the network.

7.4. Cross Domain Residual Transfer Learning for Person Re-identification

Participants: Furqan Khan, Francois Brémond.

Keywords: multi-shot person re-identification, transfer learning, residual unit

Person re-identification (re-ID) refers to the retrieval task where the goal is to search for a given person (query) in disjoint camera views (gallery). Performance of appearance based person re-ID methods depends on the similarity metric and the feature descriptor used to build a person's appearance model from given image(s).

A novel way is proposed to transfer model weights from one domain to another using residual learning framework instead of direct fine-tuning. It also argues for hybrid models that use learned (deep) features and statistical metric learning for multi-shot person re-identification when training sets are small. This is in contrast to popular end-to-end neural network based models or models that use hand-crafted features with adaptive matching models (neural nets or statistical metrics). Our experiments demonstrate that a hybrid model with residual transfer learning can yield significantly better re-identification performance than an end-to-end model when training set is small. On iLIDS-VID [78] and PRID [67] datasets, we achieve rank1 recognition rates of 89.8% and 95%, respectively, which is a significant improvement over state-of-the-art.

7.4.1. Residual Transfer Learning

We use RTL to transfer a model trained on Imagenet [63] for object classification to perform person re-ID. We chose to use 16-layer VGG model due to its superior performance in comparison to AlexNet and overlooked ResNet for its extreme depth because our target datasets are small and do not warrant such a deep model for higher performance.

One advantage of using residual learning [66] for model transfer is that it allows more flexibility in terms of modeling the difference between two tasks through a number of residual units and their composition. We noted that when residual units are added to the network with a different network head, training loss is significantly higher in the beginning which pushes the network far away from pre-trained solution by trying to over compensate through residual units. To avoid this, we propose to train the network in 4 stages, with fourth stage being optional (Fig. 9). The proposed work has been published in [45].

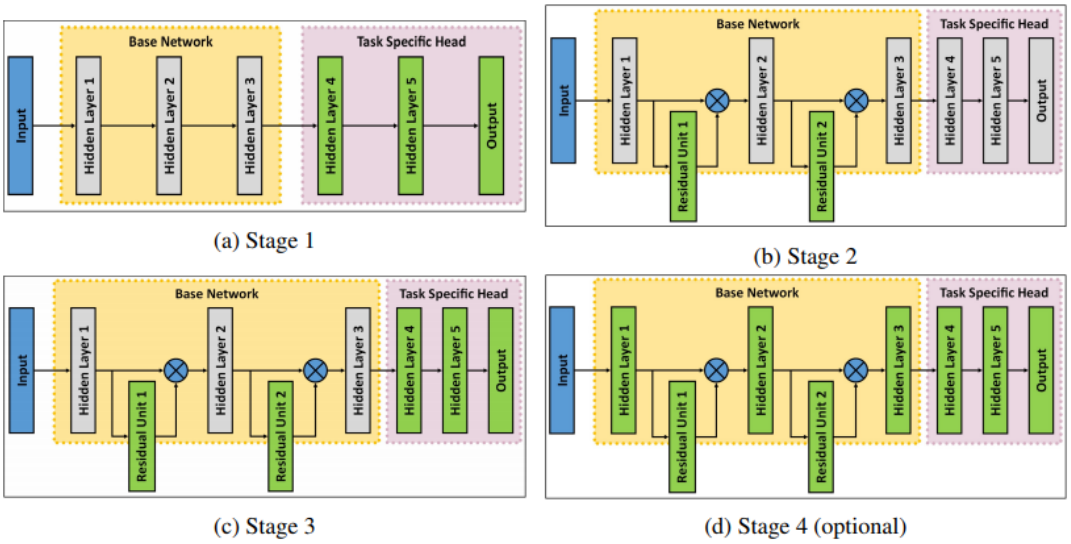


Figure 9. Residual Transfer Learning in 4 stages. During each stage only the selected layers (shown in green) are trained. Residual Units are added to the network after first stage of RTL.

- **Stage 1**: In the first stage, we replace original head of the network with a task specific head and initialize it randomly. At this stage, we do not add any residual units to the network and train only the parameters of the replaced head of the network. Thus only the head layers are considered to contribute to the loss. This allows the network to learn noisy high level representation for the desired task and decrease the network loss without affecting lower order layers.
- **Stage 2**: In the second stage, we add residual units to the network and initialize them randomly. Then we freeze all other layers, including the network head, and optimize the parameters of added residual units. As the head and other layers are fixed, residual units are considered as the source of loss. As we start with a reasonably low loss value, residual units are not forced to over compensate for the loss.
- **Stage 3**: In the third stage, we train the network by learning parameters of both added residual units and network head, thus allowing both the lower and higher order representations to adjust to the specific task.
- **Stage 4 (Optional)**: We noticed in our experiments on different datasets that the loss function generally gets low enough by the end of third stage. However, if needed, the whole network can be trained to further improve performance.

7.4.2. Conclusion

When using identity loss and large amount of training data, RTL gives comparable performance to direct fine-tuning of network parameters. However, the performance difference between two transfer learning approaches is considerably in favor of RTL when training sets are small. The reason is that when using RTL only a few parameters are modified to compensate for the residual error of the network. Still, the higher order layers of the network are prone to over-fitting. Therefore, we propose using hybrid models where higher order domain specific layers are replaced with statistical metric learning. We demonstrate that the hybrid model performs significantly better on small datasets and gives comparable performance on large datasets. The ability of the model to generalize well from small amount of data is crucial for practical applications because frequent data collection in large amount for training is nit possible.

7.5. Face-based Attribute Classification and Manipulation

Participants: Abhijit Das, Antitza Dantcheva, Francois Brémond.

Keywords: Face, Attribute, GAN, Biometrics

Due to the biasness of face analytic datasets, with respect to factors such as age, gender, ethnicity, pose and resolution, systems based on a skewed training dataset are bound to produce skewed results. Further, it has been exhibited in the literature [59] that such biases may have serious impacts on performance in challenging situations where the outcome is critical. In order to progress toward balanced face recognition and attribute estimation, the 1st International Workshop on Bias Estimation in Face Analytics was organized in conjunction with ECCV 2018. The workshop also organized a challenge to introduce a well-balanced dataset across multiple factors: age, gender, ethnicity, pose and resolution and requested for algorithms to estimate biases.

We proposed a Multi-Task Convolutional Neural Network (MTCNN) algorithm that jointly learned [37] gender, age and ethnicity by a loss function involving joint dynamic loss weight adjustment and was successful, as well as relatively unbiased in estimating age, gender and ethnicity. Our algorithm was found to be the best algorithm focusing the aim of the competition and the above mentioned research problem.

7.5.1. Generative Adversal Network (GAN)

models are autoregressive models depending on the global information, which can be potentially affected by its employment on local feature/ attribute-based erasuring. In addition, these models are typically trained depending on the maximum likelihood to find the intense difference between the regression domains, as a result after a certain limit of learning it can produce very naive development in the interpolation of the regression carried out for the purpose of local attribute removal. Hence, to mitigate an aforementioned couple of pitfalls we propose a method for localizing the Cycle GAN (C-GAN) for local feature-based regression. We

trained the C-GAN with domain-specific local feature and end model was recurrently imposed on the testing images. We experimented the Local C-GAN (L-C-GAN) on facial attribute (eyeglass and moustache/ bearded) auto-regression. Our qualitative performance on partial CelebA dataset and a couple of datasets we collected is promising. Moreover, ensuring the facial attributes have also been found to achieve better performance accuracy with respect to the presence of these attributes.

7.6. From Attribute-labels to Faces: Face Generation using a Conditional Generative Adversarial Network ,

Participants: Yaohui Wang, Antitza Dantcheva, Francois Brémond.

Keywords: Generative Adversarial Networks, Face generation

Facial attributes are instrumental in semantically characterizing faces. Automated classification of such attributes (i.e., age, gender, ethnicity) has been a well studied topic. We here seek to explore the inverse problem, namely given attribute-labels the *generation of attribute-associated faces*. The interest in this topic is fueled by related applications in law enforcement and entertainment. In this work, we propose two models for attribute-label based facial image and video generation incorporating 2D (see Figure 10) and 3D (see Figure 11) deep conditional generative adversarial networks (DCGAN). The attribute-labels serve as a tool to determine the specific representations of generated images and videos. While these are early results (see Figure 12 and 13), our findings indicate the methods' ability to generate realistic faces from attribute labels.

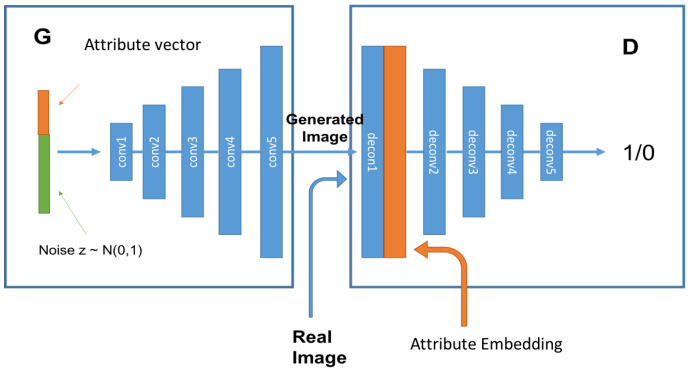


Figure 10. Architecture of proposed 2D method consisting of two modules, a discriminator D and a generator G. While D learns to distinguish between real and fake images, classifying based on attribute-labels, G accepts as input both, noise and attribute-labels in order to generate realistic face images.

7.7. Face Analysis in Structured Light Images

Participants: Vikas Thamizharasan, Antitza Dantcheva, Francois Brémond.

Keywords: Structured light, Face analysis

The main objective has been to perform face analysis tasks like authentication, gender, age and ethnicity classification by generating low-dimensional face embedding from the raw data acquired from structured light (see Figure 14) sensors using deep learning techniques. In this context we studied depth/disparity map extraction (see Figure 15), as well as other models.

7.8. Deep-Temporal LSTM for Daily Living Action Recognition

Participants: Srijan Das, Michal Koperski, Francois Brémond, Gianpiero Francesca.

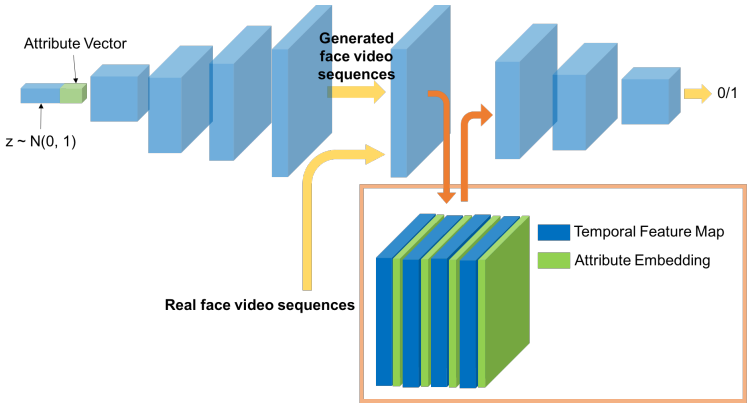


Figure 11. Architecture of proposed 3D model for face video generation

Keywords: Temporal sequences, Appearance, LSTM

We have proposed to improve the traditional use of RNNs by employing a many to many model for video classification. We analyzed the importance of modeling spatial layout and temporal encoding for daily living action recognition. Many RGB methods focus only on short term temporal information obtained from optical flow. Skeleton based methods on the other hand show that modeling long term skeleton evolution improves action recognition accuracy. In this work, we proposed a deep-temporal LSTM architecture (see fig. 16) which extends standard LSTM and allows better encoding of temporal information. In addition, we have proposed to fuse 3D skeleton geometry with deep static appearance. We validated our approach on publicly available datasets (CAD60, MSRDailyActivity3D and NTU-RGB+D), achieving competitive performance as compared to the state-of-the art. The proposed framework has been published in AVSS 2018 [39].

7.9. Spatio-Temporal Grids for Daily Living Action Recognition

Participants: Srijan Das, Kaustubh Sakhalkar, Michal Koperski, Francois Brémond.

Keywords: Spatio-temporal, Grids, Multi-modal

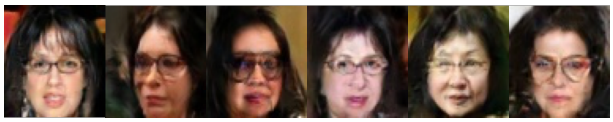
This work addresses the recognition of short-term daily living actions from RGB-D videos. Most of the existing approaches ignore spatio-temporal contextual relationships in the action videos. So, we have proposed to explore the spatial layout to better model the appearance. In order to encode temporal information, we divided the action sequence into temporal grids. We address the challenge of subject invariance by applying clustering on the appearance features and velocity features to partition the temporal grids. We validated our approach on four public datasets. The results show that our method is competitive with the state-of-the-art. The proposed architecture has been published in ICVGIP 2018 [40].

7.10. A New Hybrid Architecture for Human Activity Recognition from RGB-D videos

Participants: Srijan Das, Monique Thonnat, Kaustubh Sakhalkar, Michal Koperski, Francois Brémond, Gianpiero Francesca.

Keywords: Visual cues, Data fusion, RGB-D videos

(a) no glasses, female, black hair, smiling, young



(b) glasses, female, black hair, not smiling, old

(c) no glasses, male, no black hair, smiling, young

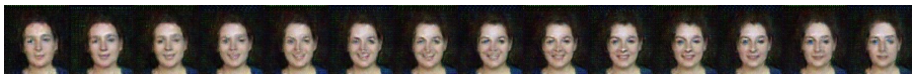
(d) glasses, male, no black hair, not smiling, old

Figure 12. Example images generated by the proposed 2D model.

(a) male, adolescent

(b) male, adult

(c) female, adolescent



(d) male, adult

Figure 13. Chosen output samples from 3DGAN

Figure 14. Structured light. A calibrated camera and projector (typically both near infrared) are placed at a fixed, known baseline. The structured light pattern helps establish correspondence between observed and projected pixels. Depth is derived for each corresponding pixel through triangulation. The process is akin to two stereo cameras, but with the projector system replacing the second camera, and aiding the correspondence problem.

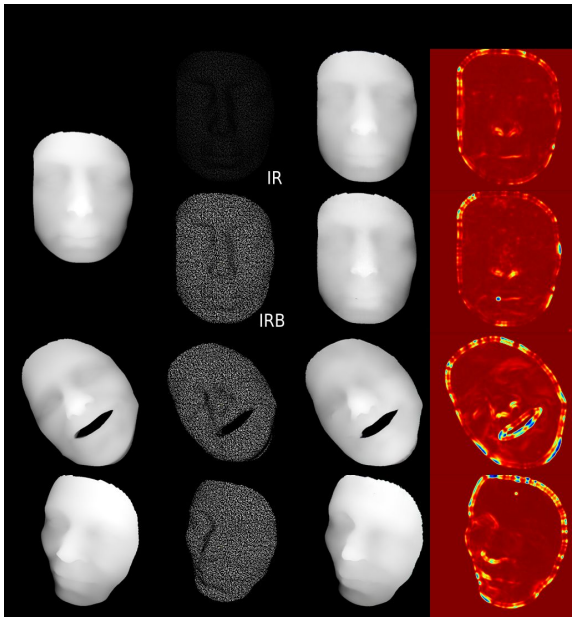


*Figure 15. *IR - Infrared image, IRB - Binarized Infrared image*

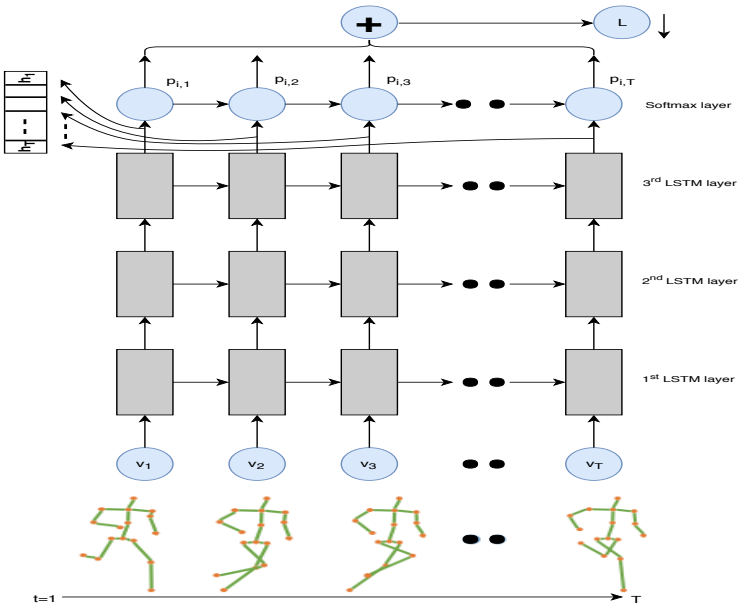


Figure 16. Framework of the deep-temporal LSTM proposed approach in [39]

Activity Recognition from RGB-D videos is still an open problem due to the presence of large varieties of actions. We have proposed a new architecture by mixing a high level handcrafted strategy and machine learning techniques. In order to address the problem of large variety of actions, we proposed a novel two level fusion strategy to combine motion, appearance and 3D pose information. For 3D pose information, we use the work published in AVSS 18 (described above). As similar actions are common in daily living activities, we also proposed a mechanism for similar action discrimination using dedicated SVMs. We validated our approach on four public datasets, CAD-60, CAD-120, MSRDailyActivity3D, and NTU-RGB+D improving the state-of-the-art results on them. The proposed architecture has been published in the industrial session of MMM 2019 [41].

7.11. Where to Focus on for Human Action Recognition?

Participants: Srijan Das, Arpit Chaudhary, Francois Brémond, Monique Thonnat.
Keywords: Spatial attention, Body parts, End-to-end

We proposed a spatial attention mechanism based on 3D articulated pose to focus on the most relevant body parts involved in the action. For action classification, we proposed a classification network compounded of spatio-temporal subnetworks modeling the appearance of human body parts and RNN attention subnetwork implementing our attention mechanism. Furthermore, we trained our proposed network end-to-end using a regularized cross-entropy loss, leading to a joint training of the RNN delivering attention globally to the whole set of spatio-temporal features, extracted from 3D ConvNets. Our method outperforms the State-of-the-art methods on the largest human activity recognition dataset available to-date (NTU RGB+D Dataset) which is also multi-views and on a human action recognition dataset with object interaction (Northwestern-UCLA Multiview Action 3D Dataset). The proposed framework will be published in WACV 2019. Sample visual results displaying the attention scores attained for each body parts can be seen in fig. 17.

7.12. Online Temporal Detection of Daily-Living Human Activities in Long Untrimmed Video Streams

Participants: Abhishek Goel, Abdelrahman G. Abubakr, Michal Koperski, Francois Brémond.
keywords: Daily-living activity recognition, Human activity detection, Video surveillance, Smarthome

Many approaches were proposed to solve the problem of activity recognition in short clipped videos, which achieved impressive results with hand-crafted and deep features. However, it is not practical to have clipped videos in real life, where cameras provide continuous video streams in applications such as robotics, video surveillance, and smart-homes. Here comes the importance of activity detection to help recognizing and localizing each activity happening in long videos. Activity detection can be defined as the ability to localize starting and ending of each human activity happening in the video, in addition to recognizing each activity label. A more challenging category of human activities is the daily-living activities, such as eating, reading, cooking, etc, which have low inter-class variation and environment where actions are performed are similar. In this work we focus on solving the problem of detection of daily-living activities in untrimmed video streams. We introduce new online activity detection pipeline that utilizes single sliding window approach in a novel way; the classifier is trained with sub-parts of training activities, and an online frame-level early detection is done for sub-parts of long activities during detection. Finally, a greedy Markov model based post processing algorithm is applied to remove false detection and achieve better results. We test our approaches on two daily-living datasets, DAHLIA and GAADRD, outperforming state of the art results by more than 10%. The proposed work has been published in [43].

7.12.1. The Work Flow of processing untrimmed videos is composed of three tasks:

- **Feature extraction** consists in extracting the Person-Centered CNN (PC-CNN) features as shown in fig. 18.
- **Classifier Training:** All training videos are first divided into relatively small windows of size W frames, which represent activity sub-videos (subparts). Then the features are generated for all these windows and the training is done with linear SVM classifier using all activities sub-videos.
- **Majority voting filtering**, as depicted in fig. 19, looks up for neighbors within a certain range that have the same label apply majority-voting between the labels

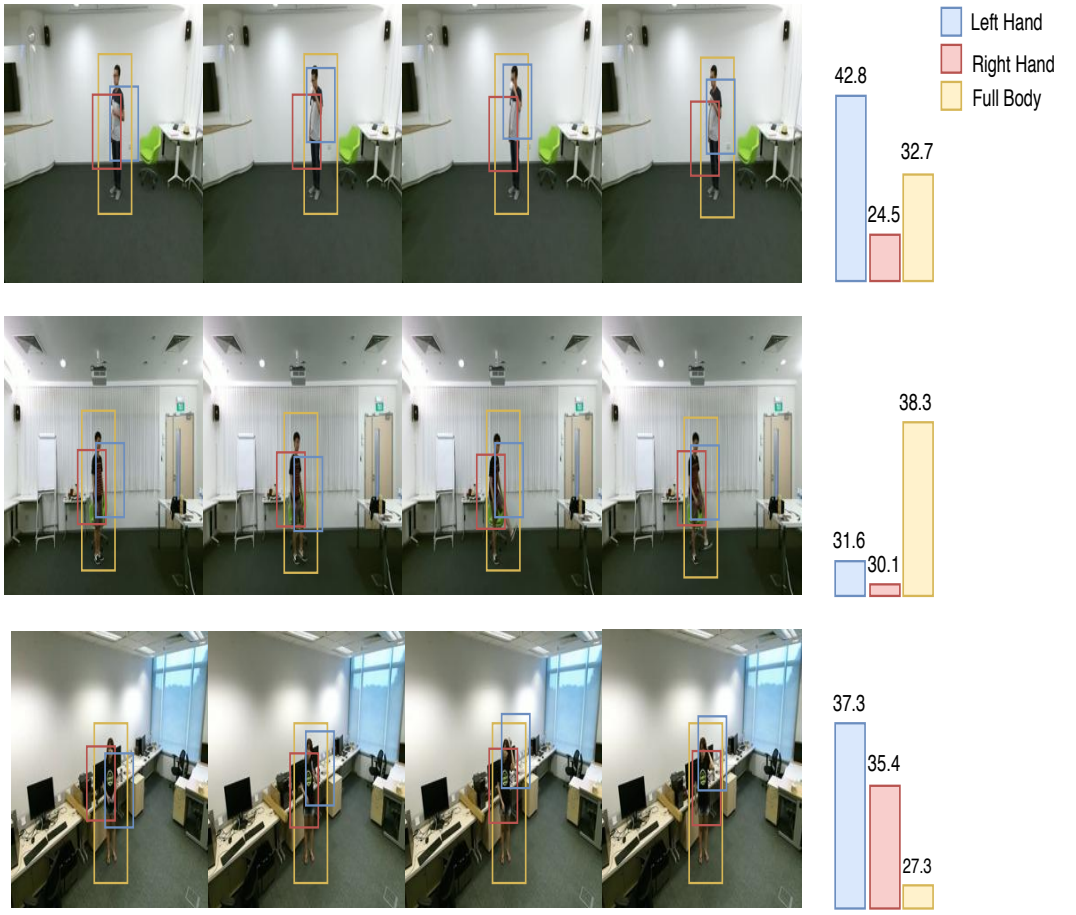


Figure 17. Example of video sequences with their respective attention scores. The action categories presented are drinking water with left hand (1st row), kicking (2nd row) and brushing hair with left hand (last row).

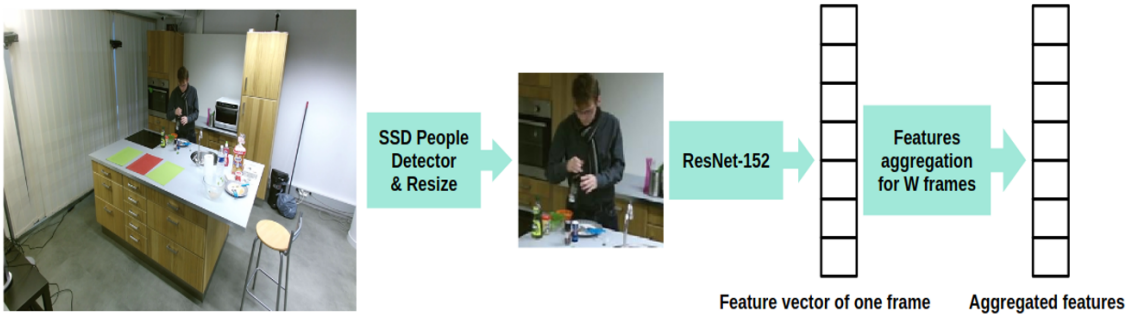


Figure 18. Extracting the Person-Centered CNN (PC-CNN) features

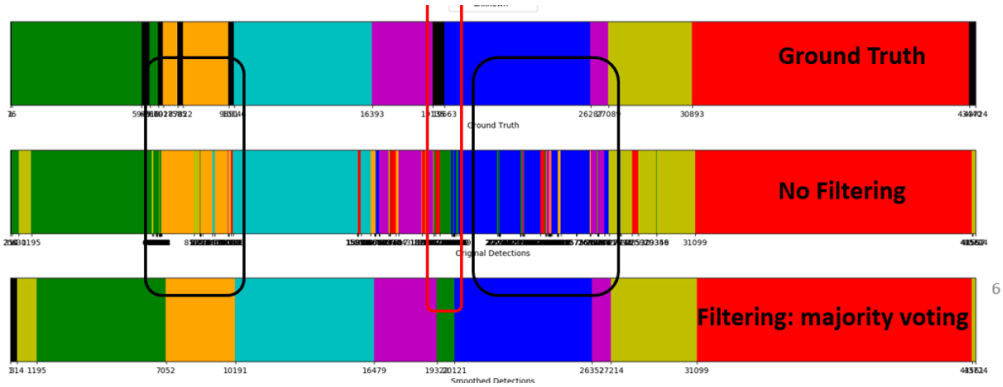


Figure 19. Post-filtering

7.13. Activity Detection in Long-term Untrimmed Videos by discovering sub-activities

Participants: Farhood Negin, Abhishek Goel, Abdelrahman G. Abubakr, Gianpiero Francesca, Francois Brémond.

Keywords: Activity detection, Semi-supervised learning, Sub-activity detection.

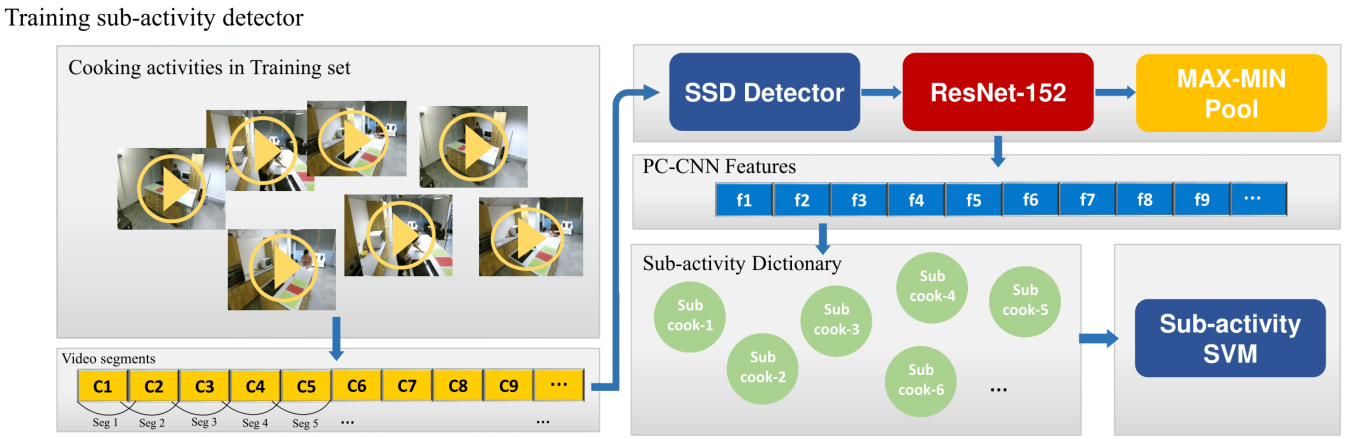


Figure 20. The process of extracting PC-CNN features and training of a weakly supervised sub-activity detector for the "Cooking" activity.

Detecting temporal delineation of activities is important to analyze large-scale videos. However, there are still challenges yet to be overcome in order to have an accurate temporal segmentation of activities. Detection of daily-living activities is even more challenging due to their high intra-class and low inter-class variations, complex temporal relationships of sub-activities performed in realistic settings. To tackle these problems, we

propose an online activity detection framework based on the discovery of sub-activities. We consider a long-term activity as a sequence of short-term sub-activities. Our contributions can be summarized as follows:

- We introduce a new online frame-level activity detection pipeline which uses single-sized window approach. A weakly supervised classifier is trained directly on sub-activities discovered by clustering and operates on test videos to capture sub-activities of long videos within a fixed temporal window.
- To alleviate the noisy detections especially in activity boundaries, we propose a novel greedy post-processing method based on Markov models.
- We have extensively evaluated our proposed method on untrimmed videos from DAHLIA [68] and GAADRD [77] datasets and achieved state-of-the-art performances.

7.13.1. Proposed Method:

Our framework produces frame-level activity labels in an online manner by two major steps followed by a novel greedy post-processing technique. In order to handle long activities, activities are decomposed into a sequence of fixed-length overlapping temporal clips. We then extract deep features from the clips. We suggested a person-centric feature (PC-CNN) based on SSD detector that satisfies required processing efficiency of online systems. We then proposed a weakly-supervised method for the discovery of sub-activities of long-term activities which benefits from clustering and model selection methods to find the optimal sub-activities of the given activities. In order to characterize each activity with constituent sub-activities, we use K-means to cluster that activity's clips and construct a specific sub-activity dictionary. Therefore, we have one sub-activity dictionary for each main activity. We represent an activity sequence with sub-activity assignments using the trained dictionary. Then, for each activity class, we train a binary SVM classifier (one versus all) based on its sub-activities (Figure 20). The trained classifiers are then simultaneously used to produce frame-level activity labels with the help of a sliding window architecture. It should be noticed that unlike multi-scale sliding window methods, we only use a single fixed-size temporal window thanks to recognition of fixed length sub-activities. Finally, assuming temporal progression of sub-activities, we developed a greedy algorithm based on Markov models to refine noisy sub-activity proposals in middle and boundary regions of long activities. We evaluated the proposed method on two daily-living activity datasets and achieved state-of-the-art performances.

Table 1. The activity detection results obtained on the DAHLIA. Values in bold represent the best performance.

	ELS			**Max Subgraph Search**			**DOHT (HOG)**			**Sub Activity**		
	FA_1	F_score	IoU	FA_1	F_score	IoU	FA_1	F_score	IoU	FA_1	F_score	IoU
View 1	0.18	0.18	0.11	-	0.25	0.15	0.80	0.77	0.64	**0.85**	**0.81**	**0.73**
View 2	0.27	0.26	0.16	-	0.18	0.10	0.81	0.79	0.66	**0.87**	**0.82**	**0.75**
View 3	0.52	0.55	0.39	-	0.44	0.31	0.80	**0.77**	0.65	**0.82**	0.76	**0.69**

Table 2. Detection results obtained on the GAADRD dataset.

Method	FA_1	F_score	IoU
simple sliding window(HOG)	0.68	0.52	0.40
simple sliding window(PC-CNN)	0.61	0.55	0.44

Tables 1 and 2 show the results of applying the developed frameworks on DAHLIA and GAADRD respectively. It can be noticed that in DAHLIA dataset (compared to [71], [61], [60]), we significantly outperformed state-of-the-art results in all of the categories except in camera view 3 when the F-Score metric is used. We reported the results of GAADRD dataset with the two types of features HOG and PC-CNN. As it can be seen, even with hand-crafted features our framework produces comparable results. In future work, we are going

to improve the sub-activity discovery algorithm by making it able to distinguish similar sub-activities in two different activities.

7.14. Video based Face Analysis for Health Monitoring

Participants: Abhijit Das, Antitza Dantcheva, Francois Brémond.

Keywords: Face, Attribute, GAN, Biometrics

Video based analysis in severely demented Alzheimer's Disease (AD) patients can be helpful for the analysis of their neuropsychiatric symptom such as apathy, depression. Even for the doctors it can be hard to know whether a person has depression or apathy. The main difference is that a person with depression will have feelings of sadness, be tearful, feel hopeless or have low self-esteem. Whereas, symptoms of person suffering from apathy can make the person's life less enjoyable. Therefore, a psychological protocol scenario can be used for video-based emotion analysis and facial movement can be used for discriminating apathetic person and non-apathetic person.

We proposed to use a) the facial expressions (neutral + 6 basic emotions: anger, disgust, happiness, surprise, sadness, fear) extracted using 50 layer Resnet, b) facial movements employing 68 facial landmark points, c) action unit intensity and frequency for AU 1, 2, 4, 5, 6, 7, 9, 10, 12, 14, 15, 17, 20, 23, 25, 26, and 45 using OpenFace and d) lip movements employing the 3D mouth open vector using the mean of upper lip and mean of bottom lip extracted from the facial landmarks detected around the lip as feature for each frame of the video. We post-process the features and calculated the amplitude, SD (Standard Deviations) and mean of each clip (10 seconds per clip) and these features were passed inputs to GRU. The GRU is connected to the Fully Connected layers, these fully connected features are mean pooled to get the apathy/non-apathy classification.

7.15. Mobile Biometrics

Participants: Abhijit Das, Antitza Dantcheva, Francois Brémond.

Keywords: Mobile biometrics

The prevalent commercial deployment of mobile biometrics as a robust authentication method on mobile devices has fueled increasingly scientific attention. Motivated by this, in this work [38] we seek to provide insight on recent development in mobile biometrics. We present parallels and dissimilarities of mobile biometrics and classical biometrics, enumerate related strengths and challenges. Further, we provide an overview of recent techniques in mobile bio-metrics, as well as application systems adopted by industry. Finally, we discuss open research problems in this field.

7.16. Comparing Methods for Assessment of Facial Dynamics in Patients with Major Neurocognitive Disorders

Participants: Yaohui Wang, Antitza Dantcheva, Francois Brémond.

Keywords: Face Analysis

Assessing facial dynamics in patients with major neurocognitive disorders and specifically with Alzheimer's disease (AD) has shown to be highly challenging. Classically such assessment is performed by clinical staff, evaluating verbal and non-verbal language of AD-patients, since they have lost a substantial amount of their cognitive capacity, and hence communication ability. In addition, patients need to communicate important messages, such as discomfort or pain. Automated methods would support the current healthcare system by allowing for telemedicine, *i.e.,* lesser costly and logistically inconvenient examination. In this work [52], we compare methods for assessing facial dynamics such as talking, singing, neutral and smiling in AD-patients, captured during music mnemotherapy sessions. Specifically, we compare 3D ConvNets (see Figure 21), Very Deep Neural Network based Two-Stream ConvNets (see Figure 22), as well as Improved Dense Trajectories. We have adapted these methods from prominent action recognition methods and our promising results suggest that the methods generalize well to the context of facial dynamics. The Two-Stream ConvNets in combination with ResNet-152 obtains the best performance on our dataset (Table 3), capturing well even minor facial dynamics and has thus sparked high interest in the medical community.

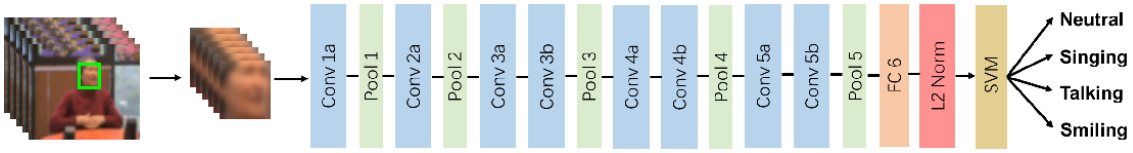
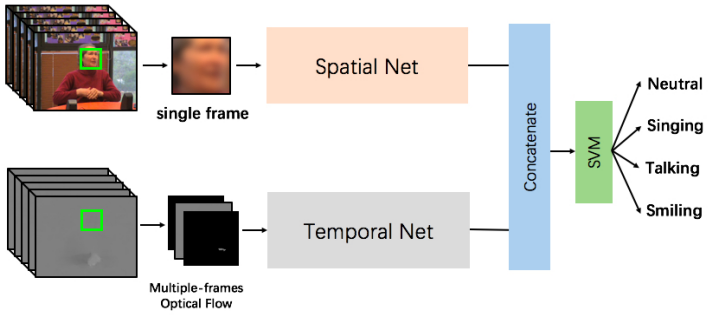
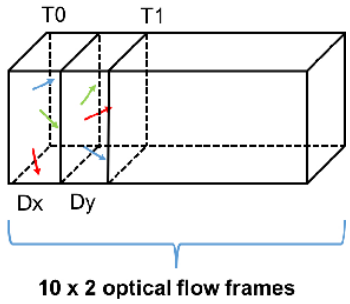


Figure 21. ***C3D based facial dynamics detection:*** *For each video sequence, faces are detected and the face sequences are passed into a pre-trained C3D network to extract a 4096-dim feature vector for each video. Finally a SVM classifier is trained to predict the final classification result. We have blurred the faces of the subject in this figure, in order to preserve the patient's privacy.*



(a) Two-Stream Architecture



(b) Stacked Optical Flow Field volume

Figure 22. (a) While the spatial ConvNet accepts a single RGB frame as input, the temporal ConvNet's input is the D_x and D_y of 10 consecutive frames, namely 20 input channels. Both described inputs are fed into the Two-stream ConvNets, respectively. We use in this work two variations of Very Deep Two Stream ConvNets, incorporating VGG-16 [76] ResNet-152 [65] for both streams respectively. (b) The optical flow of each frame has two components, namely D_x and D_y. We stack 10 times D_y after D_x for each frame to form a 20 frames length input volume.

Table 3. Classification accuracies of C3D, Very Deep Two-Stream ConvNets, iDT, as well as fusion thereof on the presented ADP-dataset. We report the Mean Accuracy (MA) associated to the compared methods. Abbreviations used: SN...Spatial Net, TN...Temporal Net.

Method	**MA (%)**
C3D	67.4
SN of Two-Stream ConvNets (VGG-16)	65.2
TN of Two-Stream ConvNets (VGG-16)	69.9
Two-Stream ConvNets (VGG-16)	76.1
SN of Two-Stream ConvNets (ResNet-152)	69.6
TN of Two-Stream ConvNets (ResNet-152)	75.8
Two-Stream ConvNets (ResNet-152)	76.4
iDT	61.2
C3D + iDT	71.1
Two-Stream ConvNets (VGG-16) + iDT	78.9
Two-Stream ConvNets (ResNet-152) + iDT	**79.5**

7.17. Combating the Issue of Low Sample Size in Facial Expression Recognition

Participants: S L Happy, Antitza Dantcheva, Francois Brémond.

Keywords: Face analysis, Expression recognition

The universal hypothesis suggests that the six basic emotions - anger, disgust, fear, happiness, sadness, and surprise - are being expressed by similar facial expressions by all humans. While existing datasets support the universal hypothesis and contain images and videos with discrete disjoint labels of profound emotions, real-life data contain jointly occurring emotions and expressions of different intensities. Reliable data annotation is a major problem in this field, which results in publicly available datasets with low sample size. Transfer learning [73], [64] is usually used to combat the low sample size problem by capturing high level facial semantics learned on different tasks. However, models which are trained using categorical one-hot vectors often over-fit and fail to recognize low or moderate expression intensities. Motivated by the above, as well as by the lack of sufficient annotated data, we here propose a weakly supervised learning technique for expression classification, which leverages the information of unannotated data. In weak supervision scenarios, a portion of training data might not be annotated or wrongly annotated [79]. Crucial in our approach is that we first train a convolutional neural network (CNN) with label smoothing in a supervised manner and proceed to tune the CNN-weights with both labelled and unlabelled data simultaneously. The learning method learns the expression intensities in addition to classifying them into discrete categories. This bootstrapping of a fraction of unlabelled samples, replacing labelled data for model-update, while maintaining the confidence level of the model on supervised data improves the model performance.

Table 4. Cross database classification performance when using CK+ database for training.

Test databases	Percentage of training data		
	25%	50%	80%
CK+ (test-set)	88.79%	91.29%	95.16%
RaFD	64.25%	65.25%	78.46%
lifespan	35.13%	40.51%	60.83%

7.17.1. Experimental Results

Experiments were conducted on three publicly available expression datasets, namely CK+, RaFD, and lifespan. Substantial experiments on these datasets demonstrate large performance gain in cross-database performance,

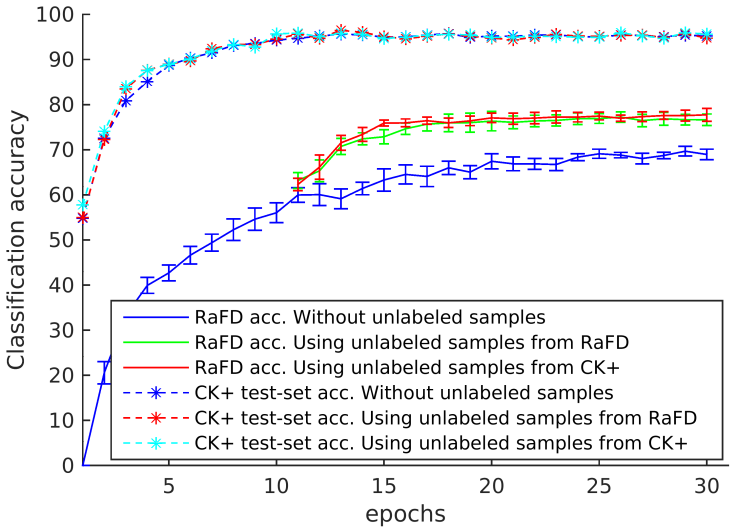


Figure 23. Cross-database experiments show significant performance improvement.

as well as show that the proposed method achieves to learn different expression intensities, even when trained with categorical samples. As can be seen in Fig. 23, when the model is trained on CK+ with unlabelled data, the model-performance improved by 11% in RaFD cross database evaluation. We observe that the use of unlabelled data from either CK+ or RaFD results in similar performances. Utilizing unlabelled images from CK+, the network sees varying expression-intensities and adapts to it. Table 4 reports the self and cross-database classification results with respect to varying number of training samples. Significant classification accuracy has been obtained with merely 25% of the training data. Use of a larger labelled training set strikingly boosts the cross-database performance. In future, we are planning to further improve the performance with unsupervised learning of expression patterns.

7.18. Serious Exergames for Cognitive Stimulation

Participants: Guillaume Sacco, Monique Thonnat.

Keywords: Neurocognitive disorders, Serious games, Geriatrics, Executive functions, Physical exercise, Cognitive training

A PhD thesis has been defended on the 8th of June at Nice University on this topic by Guillaume Sacco. This thesis presents a clinical and therapeutic approach aiming to create new care for patients with neurocognitive disorder. Serious exergames are serious video games integrating physical activity. Serious exergames could be tools to product enriched environment associating physical exercise and cognitive training. The aim of this thesis is to investigate whether serious exergames can contribute to the non-pharmacological management of neurocognitive disorders. In this thesis we have made two types of contributions. The first type are general contributions. One presents our integrative clinical approach associating physical exercise and cognitive training using serious exergames. The other one presents recommendations concerning the use of serious exergames. The second type of contributions are experimental. The first one aims to confirm a theoretical base of our clinical approach. The two other experiments assess the implementation of our approached in a population of patients with neurocognitive disorder. This year the integrative clinical approach associating physical exercise and cognitive training using serious exergames has been published [32] and presented at the International Conference on Gerontechnology ISG in Saint Petersburg, Florida, USA in May 2018.

7.19. Speech-Based Analysis for older people with dementia

Participants: Alexandra König, Philippe Robert, Nicklas Linz, Johannes Tröger, Jan Alexandersson.

Keywords: Alzheimer's disease, Dementia, Mild cognitive impairment, Neuropsychology , Assessment, Semantic verbal fluency, Speech recognition, Speech processing, Machine learning

7.19.1. Fully Automatic Speech-Based Analysis of the Semantic Verbal Fluency Task:

Semantic verbal fluency (SVF) tests are routinely used in screening for mild cognitive impairment (MCI). In this task, participants name as many items as possible of a semantic category under a time constraint. Clinicians measure task performance manually by summing the number of correct words and errors. More fine-grained variables add valuable information to clinical assessment, but are time-consuming. Therefore, the aim of this study is to investigate whether automatic analysis of the SVF could provide measures as accurate as the manual ones and thus, support qualitative screening of neurocognitive impairment.

Methods: SVF data were collected from 95 older people with MCI (n = 47), Alzheimer's or related dementias (ADRD; n = 24), and healthy controls (HC; n = 24). All data were annotated manually and automatically with clusters and switches. The obtained metrics were validated using a classifier to distinguish HC, MCI, and ADRD.

Results: Automatically extracted clusters and switches were highly correlated (r = 0.9) with manually established values, and performed as well on the classification task, separating HC from persons with ADRD (area under curve [AUC] = 0.939) and MCI (AUC = 0.758).

Conclusion: The results show that it is possible to automate fine-grained analyses of SVF data for the assessment of cognitive decline [70].

7.19.2. Language Modelling in the Clinical Semantic Verbal Fluency Task:

We employed language modelling (LM) as a natural technique to model production in this task. Comparing different LMs, we show that perplexity of a person's SVF production predicts dementia well (F1 = 0.83). Demented patients show significantly lower perplexity, thus are more predictable. Persons in advanced stages of dementia differ in predictability of word choice and production strategy - people in early stages differ only in predictability of production strategy (Linz et al., 2018a).

7.19.3. Telephone-based Dementia Screening I: Automated Semantic Verbal Fluency Assessment:

Despite encouraging results, there are still two main issues in leveraging pervasive sensing technologies for automatic dementia screening: significant hardware costs or installation efforts and the challenge of an effective pattern recognition. Conversely, automatic speech recognition (ASR) and speech analysis have reached sufficient maturity and allow for low-tech remote telephone-based screening scenarios. Therefore, we examine the technological feasibility of automatically assessing a neuropsychological test—Semantic Verbal Fluency (SVF)–via a telephone-based solution. We investigate its suitability for inclusion into an automated dementia frontline screening and global risk assessment, based on concise telephone-sampled speech, ASR and machine learning classification. Results are encouraging showing an area under the curve (AUC) of 0.85. We observe a relatively low word error rate of 33% despite phone-quality speech samples and a mean age of 77 years of the participants. The automated classification pipeline performs equally well compared to the classifier trained on manual transcriptions of the same speech data. Our results indicate SVF as a prime candidate for inclusion into an automated telephone-screening system [50].

7.19.4. Using Acoustic Markers extracted from Free Emotional Speech:

Apathy is a frequent neuropsychiatric syndrome in people with dementia. It leads to diminished motivation for physical, cognitive and emotional activity. Apathy is highly underdiagnosed since its criteria have been only recently established and rely heavily on the subjective evaluation of human observers. We analyzed speech samples from demented people with and without apathy. Speech was provoked by asking patients two emotional questions. Acoustic features were extracted and used in a classification task. The resulting models

show performances of AUC = 0.71 and AUC = 0.63. This is a decent first step into the direction of automatic detection of apathy from speech. Usefulness of stimuli to elicit free speech is found to depend on patients' gender [46].

7.19.5. Using Automatic Speech Analysis:

Apathy is present in several psychiatric and neurological conditions and found to have a severe negative effect on patients' life. In older people, it can be a predictor of increased dementia risk. Current assessment methods seem insufficiently objective and sensitive, thus new diagnostic tools and broad-scale screening technologies are needed. This study is the first of its kind aiming to investigate whether automatic speech analysis could be used for characterization and detection of apathy.

Methods: A group of apathetic and non-apathetic patients (n = 60) was recorded while performing two short narrative speech tasks. Paralinguistic markers relating to prosodic, formant, source and temporal qualities of speech were automatically extracted, examined between the groups and compared to baseline assessments. Machine learning experiments were carried out to validate the diagnosis power of extracted markers.

Results: Correlations between apathy sub-scales and features revealed a relation between temporal aspects of speech and the subdomains of reduction in interest and initiative, as well as between prosody features and the affective domain. Group differences were found to vary for males and females, depending on the task. Differences in temporal aspects of speech were found to be the most consistent difference between apathetic and non-apathetic patients. Machine learning models trained on speech features achieved top performances of AUC = 0.88 for males and AUC = 0.77 for females (article under review).

An additional study in this context analyses transcripts of responses to emotional questions (positive and negative) for sentiment using a French emotion dictionary (FEEL) and for psycholinguistic properties (LIWC). Significant reductions in the number of words, the magnitude of sentiment, the overall sentiment and the range between sentiment in the positive and negative questions are found for the apathetic population. This effect is consistent between the positive and the negative stories. When training machine learning classifiers to detect apathy based on these features, the best model showed an AUC of 0.874 using only sentiment features. LIWC features mostly showed no predictive power. When ASR technology was introduced to automatically create transcripts, the performance of predictive models dropped slightly to AUC = 0.864. ASR errors were consistent over all categories of sentiment words. These results highlight the potential of computational linguistic analysis in screening for apathy (article under review).

7.20. Monitoring the Behaviors of Retail Customers

Participants: Soumik Mallick, Julien Badie, Francois Brémond.

Keywords: Ontology, Event detection, Multi-sensor data fusion, Real-time person tracking

The future shops will be connected and distributors as well as shopkeepers need to fulfill their promise to provide a personalized shopping experience to the customers, for example: advising and guiding customers in real time. It could not only enrich the productivity of the staffs but also increase the product sale. Implementing digital service and information in the store (like using beacons) is of primary importance. Sellers can keep their promise by providing the customer's contextual support tool in order to sell more product. To improve the performance of the store, this digital service can help to analyze customer displacement and the reaction to the product which can help to reduce the operational costs of the store by optimizing store process. It can also help to adjust store prices, merchandising and commercial operation. Thus connected digital store is a major level for new consumer services and an efficient way to manage the store.

We use multiple video cameras to detect customer in real-time inside the store. Furthermore, data are collected from different sensors like mobile phone, video camera, GPS location or Beacon. It helps to provide us with the trajectory information of the customer. A trajectory is composed of a set of points. The trajectory points are collected with the help of sensor API. Then, the calculation of distance of points in subsequent frames is performed. Every point has a minimum distance to a certain threshold of time. If there is a difference between a distance on a certain period of time that will be considered as a moving subject. For example, if we have

2 tracklets from different sensors (and generally with a different frequency of points), we cut both tracklets just to keep the intersection (in terms of time) and then apply Dynamic Time Warping (DTW) on this section. When we have the results for all tracklet pairs, we order them by distance and we decide to authorize to merge the data from the different sensors or not, with help of fusion algorithms to pass the information from the sensors to the ontology. After that, only one trajectory is sent to the ontology. Then we create a SPARQL request to extract trajectory-based events and execute it.

In this storeConnect project, we are investigating to improve the event recognition model. It will help to identify customer activity in the different zone inside the store as well as moving and stopping positions of the customer. Furthermore, inside the ontology, we want to add different attributes such as emotion, gender etc.

7.21. Synchronous Approach to Activity Recognition

Participants: Daniel Gaffé, Sabine Moisan, Annie Ressouche, Jean-Paul Rigault, Ines Sarray.

Activity Recognition aims at recognizing and understanding sequences of actions and movements of mobile objects (human beings, animals or artefacts), that follow the predefined model of an activity. We propose to describe activities as a series of actions, triggered and driven by environmental events.

Due to the large range of application domains (surveillance, safety, health care ...), we propose a generic approach to design activity recognition systems that interact continuously with their environment and react to its stimuli at run-time. Such recognition systems must satisfy stringent requirements: dependability, real time, cost effectiveness, security and safety, correctness, completeness ... To enforce most of these properties, our approach is to base the configuration of the system as well as its execution on formal techniques. We chose the *Synchronous Approach* which provides formal bases to perform static analysis, verification and validation, but also direct implementation.

Based on the synchronous approach, we designed a new user-oriented activity description language (named ADeL) to express activities and to automatically generate recognition automata. This language relies on two formal semantics, a behavioral and an equational one [48]. We also developed a component, called Synchronizer, to transform asynchronous sensor events into synchronous "instants", necessary for the synchronous approach. This year, we mainly worked on the ADeL compiler to generate synchronous automata, on the graphical tool of this language and on the Synchronizer component.

7.21.1. ADeL Compilation:

To compile an ADeL program, we first transform it into an equation system which represents its synchronous automaton. Then we directly implement this equation system, transforming it into a Boolean equation system. This equation system provides an effective implementation of the initial ADeL program for our runtime recognition engine. The internal representation as Boolean equation systems also makes it possible to verify and validate ADeL programs, by generating a format suitable for a dedicated model checker such as the off-the-shelf NuSMV model-checker.

7.21.2. Synchronizer:

The role of the Synchronizer is to filter physical asynchronous events, to decide which ones may be considered as "simultaneous" and to aggregate the latter into logical instants. The sequence of these instants constitutes the logical time of our recognition systems. The runtime recognition engine interacts with the synchronizer and uses these instants to run the automata corresponding to the activities currently recognized. In general, no exact decision algorithm exists but several empirical strategies and heuristics may be used e.g., for determining instant boundaries. This year we completed the specification and implementation of a first version of the Synchronizer. It is parametrized by heuristics to manage events and data coming from various sensors, to define instant boundaries, and to cope with possible high level interruptions (preemptions).

Moreover, to facilitate the job of the synchronizer (to build the instants) and of the runtime engine (to wake up only the relevant automata), each automaton provides information about the awaited events at each state, i.e the events which may trigger transitions to a next state. The ADeL compiler has in charge to generate this information. In a first attempt, we computed statically all the awaited events in all states of an automaton. However, this approach implied to build the entire explicit automaton from an equation system, which was not realistic. Thus, this year we added specific equations to the equation systems of the operational semantics to compute the awaited events of each operator of the language. The information about next awaited events is now computed at runtine, when a state of the automaton is reached.

7.22. Probabilistic Activity Description Language

Participants: Elisabetta de Maria, Sabine Moisan, Jean-Paul Rigault.

Since the arrival of E. De Maria in the STARS team in September 2018, we work on the conception of a probabilistic framework for human behavior representation. The goal is to propose (i) a textual language for the description of activities which takes uncertainty into account; (ii) a formal probabilistic model to represent behaviors. Such a model will be tested and validated using experimental data coming from Alzheimer's patients. We will use temporal data resulting from different sensors and corresponding to patients playing with serious games. This will be the topic of T. L'Yvonnet's PhD starting in December. E. De Maria's main researches concern the investigation of the dynamic behavior of biological neuronal networks, using Leaky Integrate and Fire (LIF) neuronal networks, whose temporal dimension is crucial (the state of each neuron is computed taking into account not only present inputs but also past ones). This year, we used the PRISM language to model LIF neuronal networks as probabilistic reactive systems and we proposed an algorithm which aims at reducing the number of neurons and synaptical connections of these networks [42].

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

- *Toyota:* (Action Recognition System): This project runs from the 1st of August 2013 up to 2019. It aimed at detecting critical situations in the daily life of older adults living home alone. The system is intended to work with a Partner Robot (to send real-time information to the robot) to better interact with the older adult. The funding was 106 Keuros for the 1st period and more for the following years.
- *Gemalto:* This contract is a CIFRE PhD grant and runs from September 2018 until September 2021 within the French national initiative SafeCity. The main goal is to analyze faces and events in the invisible spectrum (i.e., low energy infrared waves, as well as ultraviolet waves). In this context models will be developed to efficiently extract identity, as well as event - information. These models will be employed in a school environment, with a goal of pseudo-anonymized identification, as well as event-detection. Expected challenges have to do with limited colorimetry and lower contrasts.
- *BluManta:* This contract is a CIFRE PhD grant and runs from August 2018 to August 2021. The aim is to develop an end-to-end 3D face analysis model, involving a unified deep neural network in charge of (a) creating a depth map, (b) extracting embeddings, (c) embeddings similarity estimation. This model will be targeted for high accuracy in tasks such as face authentication.
- *Kontron:* This contract is a CIFRE PhD grant and runs from April 2018 until April 2021 to embed CNN based people tracker within a video-camera.
- *ESI:* This contract is a CIFRE PhD grant and runs from September 2018 until March 2022 to develop a novel Re-Identification algorithm which can be easily set-up with low interaction.

9. Partnerships and Cooperations

9.1. National Initiatives

9.1.1. ANR

9.1.1.1. ENVISION

Program: ANR JCJC

Project acronym: ENVISION

Project title: Computer Vision for Automated Holistic Analysis of Humans

Duration: October 2017-September 2020.

Coordinator: Antitza Dantcheva (STARS)

Abstract: The main objective of ENVISION is to develop the computer vision and theoretical foundations of efficient biometric systems that analyze appearance and dynamics of both face and body, towards recognition of identity, gender, age, as well as mental and social states of humans in the presence of operational randomness and data uncertainty. Such dynamics - which will include facial expressions, visual focus of attention, hand and body movement, and others, constitute a new class of tools that have the potential to allow for successful holistic analysis of humans, beneficial in two key settings: (a) biometric identification in the presence of difficult operational settings that cause traditional traits to fail, (b) early detection of frailty symptoms for health care.

9.1.2. FUI

9.1.2.1. Visionum

Program: FUI

Project acronym: Visionum

Project title: Visonium.

Duration: January 2015- December 2018.

Coordinator: Groupe Genious

Other partners: Inria(Stars), StreetLab, Fondation Ophtalmologique Rothschild, Fondation Hospitalière Sainte-Marie.

Abstract: This French project from Industry Minister aims at designing a platform to re-educate at home people with visual impairment.

9.1.2.2. StoreConnect

Program: FUI

Project acronym: StoreConect.

Project title: StoreConnect.

Duration: September 2016 - September 2018.

Coordinator: Ubudu (Paris).

Other partners: Inria(Stars), STIME (groupe Les Mousquetaires (Paris)), Smile (Paris), Thevolys (Dijon).

Abstract: StoreConnect is an FUI project started in 2016 and will end in 2018. The goal is to improve the shopping experience for customers inside supermarkets by adding new sensors such as cameras, beacons and RFID. By gathering data from all the sensors and combining them, it is possible to improve the way to communicate between shops and customers in a personalized way. StoreConnect acts as a middleware platform between the sensors and the shops to process the data and extract interesting knowledge organized via ontologies.

9.1.2.3. ReMinAry

Program: FUI

Project acronym: ReMinAry.

Project title: ReMinAry.

Duration: September 2016 - September 2019.

Coordinator: GENIOUS Systèmes,

Other partners: Inria(Stars), MENSIA technologies, Institut du Cerveau et de la Moelle épinière, la Pitié-Salpêtrière hospital.

Abstract:This project is based on the use of motor imagery (MI), a cognitive process consisting of the mental representation of an action without concomitant movement production. This technique consists in imagining a movement without realizing it, which entails an activation of the brain circuits identical to those activated during the real movement. By starting rehabilitation before the end of immobilization, a patient operated on after a trauma will gain rehabilitation time and function after immobilization is over. The project therefore consists in designing therapeutic video games to encourage the patient to re-educate in a playful, autonomous and active way in a phase where the patient is usually passive. The objective will be to measure the usability and the efficiency of the reeducative approach, through clinical trials centered on two pathologies with immobilization: post-traumatic (surgery of the shoulder) and neurodegenerative (amyotrophic lateral sclerosis).

9.2. International Initiatives

9.2.1. International Initiatives

FER4HM

Title: Facial expression recognition with application in health monitoring

International Partner (Institution - Laboratory - Researcher):

Chinese Academy of Sciences (China) Institute of Computing Technology - Hu HAN

Duration: 2017 - 2019

Start year: 2017

See also: https://project.inria.fr/fer4hm/

The proposed research aims to provide computer vision methods for facial expression recognition in patients with Alzheimer's disease. Most importantly though, the work seeks to be part of a paradigm shift in current healthcare, in efficiently and cost effectively finding objective measures to (a) assess different therapy treatments, as well as to (b) enable automated human-computer interaction in remote large-scale healthcare- frameworks. Recognizing expressions in severely demented Alzheimer's disease (AD) patients is essential, since such patients have lost a substantial amount of their cognitive capacity [1-3], and some even their verbal communication ability (e.g., aphasia)2. This leaves patients dependent on clinical staff to assess their verbal and non-verbal language, in order to communicate important messages, as of discomfort associated to potential complications of the AD [9, 10]. Such assessment classically requires the patients' presence in a clinic, and time consuming examination involving medical personnel. Thus, expression monitoring is costly and logistically inconvenient for patients and clinical staff, which hinders among others large-scale monitoring. Approaches need to cater to the challenging settings of current medical recordings, which include continuous pose variations, occlusions, camera-movements, camera-artifacts, as well as changing illumination. Additionally and importantly, the (elderly) patients exhibit generally less profound facial activities and expressions in a range of intensities and predominantly occurring in combinations (e.g., talking and smiling). Both, Inria-STARS and CAS-ICT have already initiated research activities related to the here proposed topic. While both sides have studied facial expression recognition, CAS-ICT has explored additionally the use of heart rate monitoring sensed from a webcam in this context.

SafEE

Title: Safe Easy Environment

International Partner (Institution - Laboratory - Researcher):

Duration: 2018 - 2020

Start year: 2018

SafEE (Safe Easy Environment) investigates technologies for the evaluation, stimulation and intervention for Alzheimer patients. The SafEE project aims at improving the safety, autonomy and quality of life of older people at risk or suffering from Alzheimer's disease and related disorders. More specifically the SafEE project : 1) focuses on specific clinical targets in three domains: behavior, motricity and cognition 2) merges assessment and non pharmacological help/intervention and 3) proposes easy ICT device solutions for the end users. In this project, experimental studies will be conducted both in France (at Hospital and Nursery Home) and in Taiwan.

9.3. International Research Visitors

9.3.1. Visits to International Teams

Antitza Dantcheva visited Wael Abd-Almageed's laboratory at the Information Sciences Institute of the University of Southern California Viterbi School of Engineering in August 2018.

Antitza Dantcheva, Abhijit Das and Yaohui Wang visited the Institute of Computing Technology (ICT) at the Chinese Academy of Sciences (CAS) in August 2018.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. General Chair, Scientific Chair

Francois Brémond was a General Chair for the 3rd IEEE International Conference on Image Processing, Applications and Systems (IPAS 2018).

10.1.1.2. Member of Organizing Committees

- Abhijit Das organized 5th Sclera Segmentation Benchmarking Competition 2018, in conjunction with International Conference on Biometrics 2018.
- Abhijit Das organized 1st Thai Student Signature and Name component Recognition and Verification Competition 2018 in conjunction with ICFHR 2018.
- Antitza Dantcheva and Abhijit Das organized Recent Advances in Biometric Technology for Mobile Devices (RABTMD 2018) in conjunction with the 9th IEEE International Conference on Biometrics: Theory, Applications, and Systems BTAS 2018.
- Antitza Dantcheva and Abhijit Das were local organizing co-chairs of IEEE International Conference on Image Processing, Applications and Systems (IPAS 2018).
- Alexandra Konig was member of the organizing committee of RaPID-2018 workshop (Resources and ProcessIng of linguistic, para-linguistic and extra-linguistic Data from people with various forms of cognitive/psychiatric impairments) 8th of May 2018, Miyazaki, Japan.

10.1.2. Scientific Events Selection

10.1.2.1. Chair of Conference Program Committees

- Antitza Dantcheva was program Co-chair at the International Conference of the Biometrics Special Interest Group (BIOSIG) 2017 and 2018, September, 2018, Darmstadt, Germany.
- Alexandra Konig was program chair of 2nd Workshop on AI for Aging, Rehabilitation and Independent Assisted Living (ARIAL) @IJCAI'18 - In conjunction with 27th International Joint Conference on Artificial Intelligence and the 23rd European Conference on Artificial Intelligence,July 15, 2018, Stockholm, Sweden.

10.1.2.2. Member of Conference Program Committees

- Francois Brémond was program committee member of WACV18.
- Francois Brémond was a member of the AVSS Steering Committee for 2018
- Monique Thonnat was program committee member of the conference ICPRAM 2019
- Antitza Dantcheva was in the technical program committee of the IAPR International Conference on Biometrics (ICB) 2018
- Jean-Paul Rigault is a member of the *Association Internationale pour les Technologies à Objets* (AITO) which organizes international conferences such as ECOOP.

10.1.3. Reviews

- Francois Brémond was reviewer for many journals, such as IEEE Transactions on Circuits and Systems for Video Technology, for the journal Frontiers in Human Neuroscience, for the journal "revue Retraite et société" and Medical Engineering & Physics.
- Antitza Dantcheva was reviewer for a number of journals including IEEE Transactions on Information Forensics and Security (TIFS), IEEE Transactions on Biometrics, Behavior, and Identity Science (T-BIOM), IEEE Transactions on Circuits and Systems for Video Technology (TCSVT), IET Biometrics, Pattern Recognition letters, Pattern Recognition.
- Francois Brémond was reviewer for many conferences including : CVPR2018, ECCV2018, VOT2018, ICCV2018, WACV2018-19, ISG18.
- Monique Thonnat is a reviewer for the journal Artificial Intelligence in Medicine AIIM (Elsevier).
- Sabine Moisan was reviewer for the 10th ICAART'18 International Conference on Agents and Artificial Intelligence.

10.1.4. Member of Editorial Boards

- Francois Brémond has been handling editor of the international journal "Machine Vision and Application" since 2014 and editor of a PANORAMA special issue of Journal of Electronic Imaging Letters (JEI): Ultra Wide Context and Content Aware Imaging.
- Antitza Dantcheva has been in the Editorial Board of the Journal Multimedia Tools and Applications (MTAP) since 2017.

10.1.5. Invited Talks

Francois Brémond was invited by:

- Prof. Vasek Hlavac (Czech Technical University in Prague) to give a talk at the Computer Vision Winter Workshops CVWW in Cesky Krumlov, 5th February 2018.
- Derek J Collins (Huawei) to give a talk at the 3rd Annual Computer Vision / Video Intelligence Forum in Dublin, 0ct. 2018.
- Nicolas Padoy (University of Strasbourg) to give a talk at the workshop MCV at CVPR, June 2018
- Dr Lauren CAURO (City of Nice) to give a talk at the workshop "Ethique en santé connectée : La santé connectée, un progrès pour tous? ", 26 Oct 2018
- Christophe ROUSSEAU (University of Nice) to give a talk at the SophIA Summit : IA et vision, 9 Nov 2018

- Monique Thonnat was invited by INGER to give a talk on Monitoring People with Video Analysis at the Franco-Mexican workshop AI Technology Applications and Research on Frailty and Dementia, Mexico, 22-23 November 2018.
- A. Dantcheva. “Facial analysis: from soft biometrics to healthcare” at Information Sciences Institute (ISI) at the University of Southern California (USC), Los Angeles, USA, September 2018.
- A. Dantcheva. “Facial analysis: from biometrics to healthcare” at Institute of Computing Technology (ICT) at the Chinese Academy of Sciences (CAS), Beijing, China, August 2018.
- Monique Thonnat has been invited to give a talk on Activity Recognition for Neurocognitive Disorders at the International Symposium on Smart Healthcare and Age-Friendly by Taichung Veterans General Hospital, Taichung, Taiwan, 4 December 2018.
- Monique Thonnat has been invited by Taipei Medical University to give a talk on An Approach with serious Exergames for Assessment and Stimulation of Patients with Neurocognitive Disorders at Bioinformatics vs Medecine: The Elderly Care in the Information Era, The Needs and Responses, Taipei, Taiwan, 6 December 2018.
- Francois Brémond has been invited to give a talk on Activity Recognition for People Monitoring at the International Symposium on Smart Healthcare and Age-Friendly by Taichung Veterans General Hospital, Taichung, Taiwan, 4 December 2018.
- Francois Brémond has been invited by Taipei Medical University to give a talk on Activity Recognition to Monitor Older People at Bioinformatics vs Medecine: The Elderly Care in the Information Era, The Needs and Responses, Taipei, Taiwan, 6 December 2018.

10.1.6. Leadership within the Scientific Community

- Francois Brémond was a member of the Evaluation Committee (i.e. HCERES) of the research laboratory LIPADE from Paris Descartes University, 15 March 2018.
- Francois Brémond was a member of the Evaluation Committee for the professor position in computer science of Lyon University, April 24th and May 15th 2018.
- Francois Brémond was a Reviewer for the AME Programmatic Proposal ”Human-Robot Collaborative AI for Advanced Manufacturing and Engineering” on behalf of (A*STAR) Singapore, 30th May 2018.
- Francois Brémond was an expert for a research program at the Campus for Research Excellence and Technological Enterprise (CREATE) from the National University of Singapore (NUS), Oct 2018.
- Francois Brémond was the working group Leader for reviewing a new Inria Project Team Proposal: CHORALE
- Francois Brémond is part of the Advisory Board of the V4Design EU project for the Horizon 2020 framework, ICT-20-2017 Call, Tools for smart digital content in the creative industries, 2018-21
- Antitza Dantcheva serves in the Technical Activities Committee of the IEEE Biometrics Council since 2017
- Antitza Dantcheva serves in the EURASIP Biomedical Image & Signal Analytics (BISA) SAT 2018-2021
- Antitza Dantcheva is member of the European Reference Network for Critical Infrastructure Protection (ERNCIP), Thematic Group Extended Virtual Fencing - use of biometric and video technologies, since 2017
- Antitza Dantcheva is member of the European Association for Biometrics, since 2018

10.2. Teaching - Supervision - Juries

10.2.1. Supervision

- PhD: Guillaume Sacco, Serious video games in gerontological practice: application to relationships between physical activity and cognition, Thèses, Université Côte d'Azur, June 2018.
- PhD: F. NEGIN, Toward Unsupervised Human Activity and Gesture Recognition in Videos, Theses, Universit´e Cote d'Azur, Sep 2018.
- PhD: L. A. NGUYEN, Long-term people trackers for video monitoring systems, Theses, Université Côte d'Azur, July 2018.
- PhD: M. K. PHAN TRAN, Maintaining the engagement of older adults with dementia while interacting with serious game, Theses, Université Côte d'Azur, April 2017,

10.2.2. Juries

Francois Brémond was part of several PhD and HDR Juries :

- Nicolas Padoy, HDR, University of Strasbourg, 19 January 2018
- Renato Baptista, University of Luxembourg, 29 January 2018
- Nicolas Chesneau, Inria Lear, Grenoble, 23 February 2018
- Dinh Van Nguyen, UPMC, Paris, 2 May 2018
- Riadh Ksantini, HDR, SUP'COM Tunis, 10 September 2018
- Yiqiang Chen, LIRIS, University of Lyon, 12 October 2018
- Katy Blanc, University of Nice, 17 December 2018
- Mohamed Adel Benamara, LIRIS, University of Lyon, 19 December 2018

Monique Thonnat was member of the selection board of professor PU27 at ENIB, Brest.

Monique Thonnat is member of the scientific board of ENPC, Ecole Nationale des Ponts et Chaussées since June 2008.

11. Bibliography

Major publications by the team in recent years

[1] A. AVANZI, F. BRÉMOND, C. TORNIERI, M. THONNAT. *Design and Assessment of an Intelligent Activity Monitoring Platform*, in "EURASIP Journal on Applied Signal Processing, Special Issue on "Advances in Intelligent Vision Systems: Methods and Applications"", August 2005, vol. 2005:14, pp. 2359-2374

[2] H. BENHADDA, J. PATINO, E. CORVEE, F. BRÉMOND, M. THONNAT. *Data Mining on Large Video Recordings*, in "5eme Colloque Veille Stratégique Scientifique et Technologique VSST 2007", Marrakech, Marrocco, 21st - 25th October 2007

[3] B. BOULAY, F. BRÉMOND, M. THONNAT. *Applying 3D Human Model in a Posture Recognition System*, in "Pattern Recognition Letter", 2006, vol. 27, n^{o} 15, pp. 1785-1796

[4] F. BRÉMOND, M. THONNAT. *Issues of Representing Context Illustrated by Video-surveillance Applications*, in "International Journal of Human-Computer Studies, Special Issue on Context", 1998, vol. 48, pp. 375-391

[5] G. CHARPIAT. *Learning Shape Metrics based on Deformations and Transport*, in "Proceedings of ICCV 2009 and its Workshops, Second Workshop on Non-Rigid Shape Analysis and Deformable Image Alignment (NORDIA)", Kyoto, Japan, September 2009

[6] N. CHLEQ, F. BRÉMOND, M. THONNAT. *Advanced Video-based Surveillance Systems*, Kluwer A.P. , Hangham, MA, USA, November 1998, pp. 108-118

[7] F. CUPILLARD, F. BRÉMOND, M. THONNAT. *Tracking Group of People for Video Surveillance*, Video-Based Surveillance Systems, Kluwer Academic Publishers, 2002, vol. The Kluwer International Series in Computer Vision and Distributed Processing, pp. 89-100

[8] F. FUSIER, V. VALENTIN, F. BRÉMOND, M. THONNAT, M. BORG, D. THIRDE, J. FERRYMAN. *Video Understanding for Complex Activity Recognition*, in "Machine Vision and Applications Journal", 2007, vol. 18, pp. 167-188

[9] B. GEORIS, F. BRÉMOND, M. THONNAT. *Real-Time Control of Video Surveillance Systems with Program Supervision Techniques*, in "Machine Vision and Applications Journal", 2007, vol. 18, pp. 189-205

[10] C. LIU, P. CHUNG, Y. CHUNG, M. THONNAT. *Understanding of Human Behaviors from Videos in Nursing Care Monitoring Systems*, in "Journal of High Speed Networks", 2007, vol. 16, pp. 91-103

[11] N. MAILLOT, M. THONNAT, A. BOUCHER. *Towards Ontology Based Cognitive Vision*, in "Machine Vision and Applications (MVA)", December 2004, vol. 16, n^{o} 1, pp. 33-40

[12] V. MARTIN, J.-M. TRAVERE, F. BRÉMOND, V. MONCADA, G. DUNAND. *Thermal Event Recognition Applied to Protection of Tokamak Plasma-Facing Components*, in "IEEE Transactions on Instrumentation and Measurement", Apr 2010, vol. 59, n^{o} 5, pp. 1182-1191

[13] S. MOISAN. *Knowledge Representation for Program Reuse*, in "European Conference on Artificial Intelligence (ECAI)", Lyon, France, July 2002, pp. 240-244

[14] S. MOISAN. *Une plate-forme pour une programmation par composants de systèmes à base de connaissances*, Université de Nice-Sophia Antipolis, April 1998, Habilitation à diriger les recherches

[15] S. MOISAN, A. RESSOUCHE, J.-P. RIGAULT. *Blocks, a Component Framework with Checking Facilities for Knowledge-Based Systems*, in "Informatica, Special Issue on Component Based Software Development", November 2001, vol. 25, n^{o} 4, pp. 501-507

[16] J. PATINO, H. BENHADDA, E. CORVEE, F. BRÉMOND, M. THONNAT. *Video-Data Modelling and Discovery*, in "4th IET International Conference on Visual Information Engineering VIE 2007", London, UK, 25th - 27th July 2007

[17] J. PATINO, E. CORVEE, F. BRÉMOND, M. THONNAT. *Management of Large Video Recordings*, in "2nd International Conference on Ambient Intelligence Developments AmI.d 2007", Sophia Antipolis, France, 17th - 19th September 2007

[18] A. RESSOUCHE, D. GAFFÉ, V. ROY. *Modular Compilation of a Synchronous Language*, in "Software Engineering Research, Management and Applications", R. LEE (editor), Studies in Computational Intelligence, Springer, 2008, vol. 150, pp. 157-171, selected as one of the 17 best papers of SERA'08 conference

[19] A. RESSOUCHE, D. GAFFÉ. *Compilation Modulaire d'un Langage Synchrone*, in "Revue des sciences et technologies de l'information, série Théorie et Science Informatique", June 2011, vol. 4, n^{o} 30, pp. 441-471, http://hal.inria.fr/inria-00524499/en

[20] M. THONNAT, S. MOISAN. *What Can Program Supervision Do for Software Re-use?*, in "IEE Proceedings - Software Special Issue on Knowledge Modelling for Software Components Reuse", 2000, vol. 147, n^{o} 5

[21] M. THONNAT. *Vers une vision cognitive: mise en oeuvre de connaissances et de raisonnements pour l'analyse et l'interprétation d'images*, Université de Nice-Sophia Antipolis, October 2003, Habilitation à diriger les recherches

[22] M. THONNAT. *Special issue on Intelligent Vision Systems*, in "Computer Vision and Image Understanding", May 2010, vol. 114, n^{o} 5, pp. 501-502

[23] A. TOSHEV, F. BRÉMOND, M. THONNAT. *An A priori-based Method for Frequent Composite Event Discovery in Videos*, in "Proceedings of 2006 IEEE International Conference on Computer Vision Systems", New York USA, January 2006

[24] V. VU, F. BRÉMOND, M. THONNAT. *Temporal Constraints for Video Interpretation*, in "Proc of the 15th European Conference on Artificial Intelligence", Lyon, France, 2002

[25] V. VU, F. BRÉMOND, M. THONNAT. *Automatic Video Interpretation: A Novel Algorithm based for Temporal Scenario Recognition*, in "The Eighteenth International Joint Conference on Artificial Intelligence (IJCAI'03)", 9-15 September 2003

[26] N. ZOUBA, F. BRÉMOND, A. ANFOSSO, M. THONNAT, E. PASCUAL, O. GUERIN. *Monitoring elderly activities at home*, in "Gerontechnology", May 2010, vol. 9, n^{o} 2

Publications of the year

Doctoral Dissertations and Habilitation Theses

[27] F. NEGIN. *Toward Unsupervised Human Activity and Gesture Recognition in Videos*, Université Côte d'Azur, October 2018, https://hal.inria.fr/tel-01947341

Articles in International Peer-Reviewed Journals

[28] F. NEGIN, S. AGAHIAN, C. KÖSE. *Improving bag-of-poses with semi-temporal pose descriptors for skeleton-based action recognition*, in "Visual Computer", February 2018 [*DOI :* 10.1007/s00371-018-1489-7], https://hal.inria.fr/hal-01849283

[29] F. NEGIN, J. BOURGEOIS, P. ROBERT, F. BRÉMOND. *A Gesture Recognition Framework for Cognitive Assessment*, in "Gerontechnology", April 2018, vol. 17, n^{o} s [*DOI :* 10.4017/GT.2018.17.S.164.00], https://hal.inria.fr/hal-01849278

[30] F. NEGIN, P. RODRIGUEZ, M. KOPERSKI, A. KERBOUA, J. GONZÀLEZ, J. BOURGEOIS, E. CHAPOULIE, P. ROBERT, F. BRÉMOND. *PRAXIS: Towards automatic cognitive assessment using gesture recognition*, in "Expert Systems with Applications", September 2018, vol. 106, pp. 21 - 35 [*DOI :* 10.1016/J.ESWA.2018.03.063], https://hal.inria.fr/hal-01849275

[31] P. ROBERT, K. L. LANCTÔT, L. AGÜERA-ORTIZ, P. AALTEN, F. BRÉMOND, M. DEFRANCESCO, C. HANON, R. DAVID, B. DUBOIS, K. DUJARDIN, M. HUSAIN, A. KÖNIG, R. LEVY, V. MANTUA, D. MEULIEN, D. MILLER, H. J. MOEBIUS, J. RASMUSSEN, G. ROBERT, M. RUTHIRAKUHAN, F. STELLA, J. YESAVAGE, R. ZEGHARI, V. MANERA. *Is it time to revise the diagnostic criteria for apathy in brain disorders? the 2018 international consensus group*, in "European Psychiatry: The Journal of the Association of European Psychiatrists", 2018, vol. 17, n^{o} 54, pp. 71-76 [*DOI :* 10.1016/J.EURPSY.2018.07.008], https://hal.inria.fr/hal-01850396

[32] G. SACCO, M. THONNAT, G. B. SADOUN, P. ROBERT. *An approach with serious exergames for assessment and stimulation of patients with neurocognitive disorders*, in "Gerontechnology", April 2018, vol. 17, n^{o} Supplement, 150 p. , https://hal.inria.fr/hal-01838510

[33] D. TRAFIMOW, V. AMRHEIN, C. ARESHENKOFF, C. BARRERA-CAUSIL, E. BEH, Y. BILGIÇ, R. BONO, M. BRADLEY, W. BRIGGS, H. CEPEDA-FREYRE, S. CHAIGNEAU, D. CIOCCA, J. C. CORREA, D. COUSINEAU, M. R. DE BOER, S. DHAR, I. DOLGOV, J. GÓMEZ-BENITO, M. GRENDAR, J. GRICE, M. GUERRERO-GIMENEZ, A. GUTIÉRREZ, T. HUEDO-MEDINA, K. JAFFE, A. JANYAN, A. KARIMNEZHAD, F. KORNER-NIEVERGELT, K. KOSUGI, M. LACHMAIR, R. LEDESMA, R. LIMONGI, M. LIUZZA, R. LOMBARDO, M. MARKS, G. MEINLSCHMIDT, L. NALBORCZYK, H. T. NGUYEN, R. OSPINA, J. PEREZ-GONZALEZ, R. PFISTER, J. RAHONA, D. RODRÍGUEZ-MEDINA, X. ROMÃO, S. RUIZ-FERNÁNDEZ, I. SUAREZ, M. TEGETHOFF, M. TEJO, R. VAN DE SCHOOT, I. VANKOV, S. VELASCO-FORERO, T. WANG, Y. YAMADA, F. ZOPPINO, F. MARMOLEJO-RAMOS. *Manipulating the Alpha Level Cannot Cure Significance Testing*, in "Frontiers in Psychology", May 2018, vol. 9, https://hal.archives-ouvertes.fr/hal-01957088

International Conferences with Proceedings

[34] S. BAABOU, F. M. KHAN, F. BRÉMOND, A. BEN FRAD, M. AMINE FARAH, A. KACHOURI. *Tracklet and Signature Representation for Multi-shot Person Re-Identification.* , in "SSD 2018 - International Multi-Conference on Systems, Signals and Devices", Hammamet, Tunisia, March 2018, pp. 1-6, https://hal.inria.fr/hal-01849457

[35] A. DANTCHEVA, F. BRÉMOND, P. BILINSKI. *Show me your face and I will tell you your height, weight and body mass index*, in "International Coference on Pattern Recognition (ICPR)", Beijing, China, August 2018, https://hal.inria.fr/hal-01799574

[36] S. DAS, A. CHAUDHARY, F. BRÉMOND, M. THONNAT. *Where to Focus on for Human Action Recognition?*, in "WACV 2019 - IEEE Winter Conference on Applications of Computer Vision", Waikoloa Village, Hawaii, United States, January 2019, pp. 1-10, https://hal.inria.fr/hal-01927432

[37] A. DAS, A. DANTCHEVA, F. BRÉMOND. *Mitigating Bias in Gender, Age and Ethnicity Classification: a Multi-Task Convolution Neural Network Approach*, in "ECCVW 2018 - European Conference of Computer Vision Workshops", Munich, Germany, September 2018, https://hal.inria.fr/hal-01892103

[38] A. DAS, C. GALDI, H. HAN, R. RAMACHANDRA, J.-L. DUGELAY, A. DANTCHEVA. *Recent Advances in Biometric Technology for Mobile Devices*, in "BTAS'18, 9th IEEE International Conference on Biometrics: Theory, Applications and Systems", Los Angeles, United States, October 2018, https://hal.inria.fr/hal-01894140

[39] S. DAS, M. F. KOPERSKI, F. BRÉMOND, G. FRANCESCA. *Deep-Temporal LSTM for Daily Living Action Recognition*, in "15th IEEE International Conference on Advanced Video and Signal-based Surveillance", Auckland, New Zealand, November 2018, https://hal.inria.fr/hal-01896064

[40] S. DAS, K. SAKHALKAR, M. F. KOPERSKI, F. BRÉMOND. *Spatio-Temporal Grids for Daily Living Action Recognition*, in "11th Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP-2018)", Hyderabad, India, December 2018 [*DOI :* 10.1145/3293353.3293376], https://hal.inria.fr/hal-01939320

[41] S. DAS, M. THONNAT, K. SAKHALKAR, M. F. KOPERSKI, F. BRÉMOND, G. FRANCESCA. *A New Hybrid Architecture for Human Activity Recognition from RGB-D videos*, in "25th International Conference on MultiMedia Modeling", Thessaloniki, Greece, January 2019, https://hal.inria.fr/hal-01896061

[42] E. DE MARIA, D. GAFFÉ, A. RESSOUCHE, C. GIRARD RIBOULLEAU. *A Model-checking Approach to Reduce Spiking Neural Networks*, in "BIOINFORMATICS 2018 - 9th International Conference on Bioinformatics Models, Methods and Algorithms", Funchal Madeira, Portugal, January 2018, pp. 1-8, https://hal.archives-ouvertes.fr/hal-01638248

[43] A. GOEL, A. ABUBAKR, M. KOPERSKI, F. BRÉMOND, G. FRANCESCA. *Online temporal detection of daily-living human activities in long untrimmed video streams*, in "IEEE IPAS 2018", Nice, France, December 2018, https://hal.inria.fr/hal-01948387

[44] J. D. GONZALES ZUNIGA, T.-L.-A. NGUYEN, F. BRÉMOND. *Residual Transfer Learning for Multiple Object Tracking*, in "International Conference on Advanced Video and Signal-based Surveillance (AVSS)", Auckland, New Zealand, IEEE, November 2018, https://hal.inria.fr/hal-01928612

[45] F. M. KHAN, F. BRÉMOND. *Cross domain Residual Transfer Learning for Person Re-identification*, in "WACV 2019", Waikoloa Village, Hawaii, United States, January 2019, https://hal.inria.fr/hal-01947523

[46] N. LINZ, X. KLINGE, J. TRÖGER, J. ALEXANDERSSON, R. ZEGHARI, R. PHILIPPE, A. KÖNIG. *Automatic Detection of Apathy using Acoustic Markers extracted from Free Emotional Speech*, in "2ND WORKSHOP ON AI FOR AGING, REHABILITATION AND INDEPENDENT ASSISTED LIVING (ARIAL) @IJCAI'18", Stockholm , Sweden, July 2018, https://hal.inria.fr/hal-01850436

[47] I. SARRAY, A. RESSOUCHE, S. MOISAN, J.-P. RIGAULT, D. GAFFÉ. *A Synchronous Approach to Activity Recognition*, in "IEEE 12th International Conference on Semantic Computing (ICSC)", Laguna Hills, CA, United States, January 2018, https://hal.inria.fr/hal-01931315

[48] I. SARRAY, A. RESSOUCHE, S. MOISAN, J.-P. RIGAULT, D. GAFFÉ. *Semantic Studies of a Synchronous Approach to Activity Recognition*, in "International Conference on Software Engineering and Applications", Dubaï, United Arab Emirates, January 2018, https://hal.inria.fr/hal-01763511

[49] R. TRICHET, F. BRÉMOND. *LBP Channels for Pedestrian Detection*, in "WACV", Lake Tahoe, United States, March 2018, https://hal.inria.fr/hal-01849431

[50] J. TRÖGER, N. LINZ, A. KÖNIG, P. ROBERT, J. ALEXANDERSSON. *Telephone-based Dementia Screening I: Automated Semantic Verbal Fluency Assessment*, in "PervasiveHealth 2018 - 12th EAI International Conference on Pervasive Computing Technologies for Healthcare", New York , United States, May 2018 [*DOI :* 10.1145/NNNNNNN.NNNNNNN], https://hal.inria.fr/hal-01850406

[51] U. UJJWAL, A. DZIRI, B. LEROY, F. BRÉMOND. *Late Fusion of Multiple Convolutional Layers for Pedestrian Detection*, in "15th IEEE International Conference on Advanced Video and Signal-based Surveillance", Auckland, New Zealand, November 2018, https://hal.inria.fr/hal-01926073

[52] Y. WANG, A. DANTCHEVA, J.-C. BROUTART, P. ROBERT, F. BRÉMOND, P. BILINSKI. *Comparing methods for assessment of facial dynamics in patients with major neurocognitive disorders*, in "ECCVW 2018 - European Conference of Computer Vision Workshops", Munich, Germany, September 2018, https://hal.inria.fr/hal-01894162

[53] Y. WANG, A. DANTCHEVA, F. BRÉMOND. *From attribute-labels to faces: face generation using a conditional generative adversarial network*, in "ECCVW'18, 5th Women in Computer Vision (WiCV) Workshop in conjunction with the European Conference on Computer Vision", Munich, Germany, September 2018, https://hal.inria.fr/hal-01894150

[54] Y. WANG, A. DANTCHEVA, F. BRÉMOND. *From attributes to faces: a conditional generative network for face generation*, in "BIOSIG'18, 17th International Conference of the Biometrics Special Interest Group", Darmstadt, Germany, September 2018, https://hal.inria.fr/hal-01894144

Other Publications

[55] C. ABI NADER, N. AYACHE, V. MANERA, P. ROBERT, M. LORENZI. *Disentangling spatio-temporal patterns of brain changes in large-scale brain imaging databases through Independent Gaussian Process Analysis*, May 2018, vol. Revue d'Épidémiologie et de Santé Publique, n^{o} 66, S159 p. , 12ème Conférence Francophone d'Epidémiologie Clinique (EPICLIN) et 25èmes Journées des statisticiens des Centre de Lutte Contre le Cancer (CLCC), Poster [*DOI :* 10.1016/J.RESPE.2018.03.108], https://hal.archives-ouvertes.fr/hal-01826517

[56] L. ANTELMI, M. LORENZI, V. MANERA, P. ROBERT, N. AYACHE. *A method for statistical learning in large databases of heterogeneous imaging, cognitive and behavioral data*, 12e Conférence francophone d'Épidémiologie clinique 25e Journée des statisticiens des Centres de lutte contre le cancer, Elsevier, May 2018, vol. 66, n^{o} 3, S180 p. , EPICLIN 2018 - 12ème Conférence Francophone d'Epidémiologie Clinique / CLCC 2018 - 25èmes Journées des statisticiens des Centre de Lutte Contre le Cancer, Poster [*DOI :* 10.1016/J.RESPE.2018.03.306], https://hal.inria.fr/hal-01827389

References in notes

[57] M. ACHER, P. COLLET, F. FLEUREY, P. LAHIRE, S. MOISAN, J.-P. RIGAULT. *Modeling Context and Dynamic Adaptations with Feature Models*, in "Models@run.time Workshop", Denver, CO, USA, October 2009, http://hal.inria.fr/hal-00419990/en

[58] M. ACHER, P. LAHIRE, S. MOISAN, J.-P. RIGAULT. *Tackling High Variability in Video Surveillance Systems through a Model Transformation Approach*, in "ICSE'2009 - MISE Workshop", Vancouver, Canada, May 2009, http://hal.inria.fr/hal-00415770/en

[59] J. BUOLAMWINI, T. GEBRU. *Gender shades: Intersectional accuracy disparities in commercial gender classification*, in "Conference on Fairness, Accountability and Transparency", 2018, pp. 77-91

[60] A. CHAN-HON-TONG, C. ACHARD, L. LUCAT. *Deeply Optimized Hough Transform: Application to Action Segmentation*, in "ICIAP", 2013

[61] C. CHEN, K. GRAUMAN. *Efficient Activity Detection in Untrimmed Video with Max-Subgraph Search*, in "IEEE Trans. Pattern Anal. Mach. Intell.", 2017

[62] R. DAVID, E. MULIN, P. MALLEA, P. ROBERT. *Measurement of Neuropsychiatric Symptoms in Clinical Trials Targeting Alzheimer's Disease and Related Disorders*, in "Pharmaceuticals", 2010, vol. 3, pp. 2387-2397

[63] J. DENG, W. DONG, R. SOCHER, L.-J. LI, K. LI, L. FEI-FEI. *ImageNet: A large-scale hierarchical image database*, in "2009 IEEE Conference on Computer Vision and Pattern Recognition", 2009, pp. 248-255

[64] H. DING, S. K. ZHOU, R. CHELLAPPA. *Facenet2expnet: Regularizing a deep face recognition net for expression recognition*, in "Automatic Face & Gesture Recognition (FG 2017), 2017 12th IEEE International Conference on", IEEE, 2017, pp. 118–126

[65] K. HE, X. ZHANG, S. REN, J. SUN. *Deep Residual Learning for Image Recognition*, in "arXiv preprint arXiv:1512.03385", 2015

[66] K. HE, X. ZHANG, S. REN, J. SUN. *Deep Residual Learning for Image Recognition*, in "2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)", 2016, pp. 770-778

[67] M. HIRZER, C. BELEZNAI, P. M. ROTH, H. BISCHOF. *Person Re-identification by Descriptive and Discriminative Classification*, in "SCIA", 2011

[68] A. KARAKOSTAS, A. BRIASSOULI, K. AVGERINAKIS, I. KOMPATSIARIS, M. TSOLAKI. *The dem@ care experiments and datasets: a technical report*, in "arXiv preprint arXiv:1701.01142", 2016

[69] C. KÄSTNER, S. APEL, S. TRUJILLO, M. KUHLEMANN, D. BATORY. *Guaranteeing Syntactic Correctness for All Product Line Variants: A Language-Independent Approach*, in "TOOLS (47)", 2009, pp. 175-194

[70] A. KÖNIG, N. LINZ, J. TRÖGER, M. WOLTERS, J. ALEXANDERSSON, P. ROBERT, A. KONIG. *Fully Automatic Speech-Based Analysis of the Semantic Verbal Fluency Task*, in "Dementia and Geriatric Cognitive Disorders", June 2018, vol. 45, n^{o} 3-4, pp. 198 - 209, https://hal.inria.fr/hal-01850408

[71] M. MESHRY, M. E. HUSSEIN, M. TORKI. *Linear-time online action detection from 3D skeletal data using bags of gesturelets*, in "WACV", 2016

[72] S. MOISAN, J.-P. RIGAULT, M. ACHER, P. COLLET, P. LAHIRE. *Run Time Adaptation of Video-Surveillance Systems: A software Modeling Approach*, in "ICVS, 8th International Conference on Computer Vision Systems", Sophia Antipolis, France, September 2011, http://hal.inria.fr/inria-00617279/en

[73] H.-W. NG, V. D. NGUYEN, V. VONIKAKIS, S. WINKLER. *Deep learning for emotion recognition on small datasets using transfer learning*, in "Proceedings of the 2015 ACM on international conference on multimodal interaction", ACM, 2015, pp. 443–449

[74] L. M. ROCHA, S. MOISAN, J.-P. RIGAULT, S. SAGAR. *Girgit: A Dynamically Adaptive Vision System for Scene Understanding*, in "ICVS", Sophia Antipolis, France, September 2011, http://hal.inria.fr/inria-00616642/en

[75] R. ROMDHANE, E. MULIN, A. DERREUMEAUX, N. ZOUBA, J. PIANO, L. LEE, I. LEROI, P. MALLEA, R. DAVID, M. THONNAT, F. BRÉMOND, P. ROBERT. *Automatic Video Monitoring system for assessment*

of Alzheimer's Disease symptoms, in "The Journal of Nutrition, Health and Aging Ms(JNHA)", 2011, vol. JNHA-D-11-00004R1, http://hal.inria.fr/inria-00616747/en

[76] K. SIMONYAN, A. ZISSERMAN. *Very Deep Convolutional Networks for Large-Scale Image Recognition*, in "CoRR", 2014, vol. abs/1409.1556

[77] G. VAQUETTE, A. ORCESI, L. LUCAT, C. ACHARD. *The DAily Home LIfe Activity Dataset: A High Semantic Activity Dataset for Online Recognition*, in "FG 2017", May 2017

[78] T. WANG, S. GONG, X. ZHU, S. WANG. *Person Re-identification by Video Ranking*, in "ECCV", 2014

[79] Z.-H. ZHOU. *A brief introduction to weakly supervised learning*, in "National Science Review", 2017