**Assignment**

**2018-00712 - [STARS] How to integrate ontology description within CNN networks for Human Activity Recognition in untrimmed videos**

**Contract type**: Public service fixed-term contract  
**Level of qualifications required**: Graduate degree or equivalent  
**Other valued qualifications**: master  
**Fonction**: PhD Position  
**Level of experience**: Recently graduated

---

**Context**

Several investigations have been carried out to model activities of daily living (ADLs) to monitor older adults at home. Most systems have been developed using either simple sensor data (wearable sensors, touch sensors, RFID tags) or camera information to recognize ADLs in a home environment. However, existing work has either focused on simple activities in real-life scenarios, or the recognition of more complex (in terms of visual variabilities) activities in hand-clipped videos with well-defined temporal boundaries. We still lack research on methods that can retrieve several instances of complex activity in a continuous – untrimmed – video (and multimodal) flow of data. Existing methods that perform in online scenarios that can reason about the temporal and composite relations that characterize complex activities generally cannot handle uncertainty and tend to underperform in real life scenarios. Moreover, they have difficulties to distinguish similarly looking activities. On the other hand, methods that can handle uncertainty tend to ignore the temporal and composite relations of activities and learn short-term activity models directly from pixel data. Hence, latter model cannot recognize long-term or composed activities. 

For instance, Deep Convolutional Neural Network CNN algorithms have been applied with great success to images (and short videos), related to monitoring applications such as People Detection and Posture, Gesture and Action Recognition algorithms: DeeperCut (http://pose.mpipi-inf.mpg.de/). In addition, current state-of-the-art algorithms focus on some specific actions (with low intra class variation) like for instance “chopping”. Hence, more generic actions like “cooking” can mean either “chopping” or “mixing”. Current methods do not perform well on distinguishing similar looking activities, like laying down and falling down. Typical situations that we would like to monitor are Eating and drinking (how much? how often?) or Cooking (detect behavior that might lead to dangerous situations or non-completion of the task).

The system we want to develop will help senior people and their relatives to feel more comfortable at their home, since scene understanding intends to help at recognizing potentially dangerous situations and reporting to caregivers if necessary.

To support this work, we have a full team of researchers specialized in human behaviors, from expert in people detection and tracking, machine learning, up to medical doctors specialized in behavioral disorders. In fact, STARS group has been working on automatic video understanding since 1994. The “SUP” (“Scene Understanding Platform”) Platform developed in STARS, detects mobile objects, tracks their trajectory and recognizes related behaviors predefined by experts. This platform contains several techniques for the detection of people and for the recognition of human postures and gestures of one person using conventional cameras. We have access to large cohorts of patients and are able to collect video datasets, dedicated to behavioral disorders, such as the ones induced by dementia. We have also large storage resource and a large GPU farm, from which 28 GPU nodes are dedicated to STARS team.

---

**General Information**

- Theme/Domain: Vision, perception and multimedia interpretation  
- Information system (BAP E)  
- Town/city: Sophia Antipolis  
- Inria Center: CRI Sophia Antipolis - Méditerranée  
- Starting date: 2018-09-01  
- Duration of contract: 3 years  
- Deadline to apply: 2018-05-12

**Contacts**

- Inria Team: STARS  
- Recruiter:  
  Bremond François  
  francois.bremond@inria.fr

**The keys to success**

To be passionate about innovation and willing to go beyond the state-of-the-art.

**Conditions for application**

- Defence Security:  
- Conditions for application: As part of its diversity policy, all Inria positions are accessible to people with disabilities.

---

**Warning**: you must enter your e-mail address in order to save your application to Inria. Applications must be submitted online on the Inria website. Processing of applications sent from other channels is not guaranteed.
In this work, we would like to go beyond Deep Learning by taking advantage of CNN for pose estimation or short action detection and embedded them into an ontology based framework for long term activity recognition to address complex human behaviors. Typical pipeline can include CNNs for pose-estimation and body part classification depending on the action to detect. New 3D skeleton extractors can be used to get the action recognition process more independent from the camera viewpoint. Short temporal aspect of the actions can be handled through HMM, RNN or LSTM. Object classifier can be employed to better identify carried objects or interaction with the environment. The objective of these three steps is to extract meaningful mid-level features that can be further processed thanks to an ontology based reasoning engine. The ontology will be provided by the user to let him/her describe the targeted activities to be recognized. A challenge will be to propose an approach to leverage the knowledge acquisition process, in both parts CNN processing and ontology based reasoning. Another challenge is to design a method that can process in real-time and in a continuous/online manner an untrimmed video (long video containing several actions) and so to detect automatically the beginning and end of the targeted action.

This work will be conducted within the Cobtek team from Nice Hospital, who is specialized in clinical trials for older adults with dementia.

The evaluation of proposed frameworks and models should be performed on public datasets which contain everyday activities like Cooking Composite, Breakfast, DAHLIA (https://ieeexplore.ieee.org/document/7961782/) and domain-specific datasets like CHU (Nice Hospital – RGBD), ICP and GAADRD datasets [Kuehne et al, 2014; Rohrbach et al, 2015; Crispim-Junior et al, 2016].

There is a possibility of conducting first an internship, before the PhD thesis.

Main activities

1st year: Study the limitations of existing activity recognition algorithms. Depending on the targeted human activities to recognize, data collection might need to be carried out.
Propose an original algorithm that addresses current limitations of the state-of-the-art.
Evaluate the studied algorithms on benchmarking datasets,
Write a paper

2nd year: Design a novel approach for activity recognition
Investigation of feasibility/appropriateness of the proposed approach in practical situations
Propose an algorithm to address model learning task in (semi-)supervised settings
Write a paper
Write the PhD manuscript outline.

3rd year: Validate the novel approach
Optimize proposed algorithm for real-world scenarios and untrimmed videos.
Write a paper and PhD manuscript

Skills

Strong background in C++/Python programming languages, linux

Knowledge on the following topics is a plus:
Machine learning,
Deep Neural Networks frameworks,
Probabilistic Graphical Models,
Computer Vision,
and Optimization techniques

Benefits package

- Subsidised catering service
- Partially-reimbursed public transport
- Social security
- Paid leave
- Flexible working hours
- Sports facilities

Remuneration

Duration: 36 months
Location: Sophia Antipolis, France
Gross Salary per month: 1982€ brut per month (year 1 & 2) and 2085€ brut/month (year 3)