



**Offer #2022-05291**

## **PhD Position F/M Computer Vision / Deep Learning for human behavior monitoring**

**Contract type :** Fixed-term contract

**Level of qualifications required :** Graduate degree or equivalent

**Fonction :** PhD Position

### **About the research centre or Inria department**

The Inria Sophia Antipolis - Méditerranée center counts 34 research teams as well as 7 support departments. The center's staff (about 500 people including 320 Inria employees) is made up of scientists of different nationalities (250 foreigners of 50 nationalities), engineers, technicians and administrative staff. 1/3 of the staff are civil servants, the others are contractual agents. The majority of the center's research teams are located in Sophia Antipolis and Nice in the Alpes-Maritimes. Four teams are based in Montpellier and two teams are hosted in Bologna in Italy and Athens. The Center is a founding member of Université Côte d'Azur and partner of the I-site MUSE supported by the University of Montpellier.

### **Context**

Inria, the French National Institute for computer science and applied mathematics, promotes "scientific excellence for technology transfer and society". Graduates from the world's top universities, Inria's 2,700 employees rise to the challenges of digital sciences. With its open, agile model, Inria is able to explore original approaches with its partners in industry and academia and provide an efficient response to the multidisciplinary and application challenges of the digital transformation. Inria is the source of many innovations that add value and create jobs.

#### **Team**

The STARS research team combines advanced theory with cutting edge practice focusing on cognitive vision systems.

**Team web site :** <https://team.inria.fr/stars/>

#### Scientific context

STARS group works on automatic video monitoring and human behavior understanding for health applications. The Deep Learning 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/gestures using conventional cameras. However, there are scientific challenges in people tracking when dealing with real word scenes: cluttered scenes, handling wrong and incomplete person segmentation, handling static and dynamic occlusions, low contrasted objects, moving contextual objects (e.g. chairs), similar appearance of clothes among different people ...

Multiple Object Tracking (MOT) is a fundamental task that aims at associating the same objects across multiple frames in a video clip. A robust and accurate MOT algorithm is indispensable in broad applications, such as people monitoring and video surveillance. An end-to-end MOT algorithm can be divided into three different but closely related tasks; single frame detection of objects, short term tracking and long-term tracking of said objects, the latter two are usually merged together into a problem commonly known as data association. This gave rise to the dominant paradigm in MOT, tracking-by-detection, which first obtains bounding boxes by detection frame by frame, and then generates trajectories by associating the same objects between frames. While these tasks are part of the same MOT problem, they are often treated apart, either trained separately or the data association step is not a deep learning-based approach which hinders the whole process.

On top of the aforementioned issue of separated training, short-term tracking and long-term tracking have the same objective (data association) but they have different inputs. Short-term tracking deals with per frame feature representation of an object and long-term tracking needs to deal with a historic feature representation that encapsulates the myriad of changes of an object across a larger frame span. In other words, we need a memory that tracks said changes, that is differentiable and can back-propagate the information all the way up to the detection task.

### **Assignment**

This work consists in designing efficient long-term People Monitoring for instance by Joint Detection and Tracking algorithms. One potential approach could use differentiable Memory Banks to build a Deep Learning memory-based architecture that can be trained to learn a feature representation of a tracklet. Therefore, the main difference with respect to the current state-of-the-art is that this MemoryTracker will be conceived to mitigate the loss of information from training separately both detection, short term tracking and long-term tracking tasks. Designing an efficient memory-based architecture is far from evident. Indeed, the first challenge is to be able to infer dense representations (i.e. tracklet vectors). To do so, we propose the use of ROI-alignment from the pipeline of deformable DETR detector. We also can take advantage of joint detection and short-term tracking by using 3D CNNs, this can allow us to have temporal and spatial information that is not available with vanilla 2D CNNs. The use of 3DCNNs can output more reliable tracklets over a small number of frames and use that information to better update the MemoryBank.

In addition to allowing a truly end-to-end pipeline, the MemoryTracker could overcome the batch training problem by storing the tracklet feature vector with an intra-batch loss and an out-of-batch loss. Both losses could be based on triplet loss functions that depend on the current input sequence (intra batch) and the following sequences (out-of-batch). However, while the features of the current frames are given to the detection pipeline, the features of the previous frames are given to the MemoryBank.

To validate the work, we will assess the proposed algorithms on video-monitoring applications and homecare videos from Nice Hospital and from public places, such as the ones in MOT20  
<https://motchallenge.net/data/MOT20/>.

A state of the art, bibliography and scientific references are available at the following URL, do not hesitate to log in: <http://www-sop.inria.fr/members/Francois.Bremond/>

Schedule:

1<sup>st</sup> year:

- Study the limitations of existing DL People Tracking algorithms.
- Proposing a new approach for People Tracking using Joint Detection and Tracking.

2<sup>nd</sup> year:

- Start to Improve the proposed DL People Tracking approach.
- Writing papers

3<sup>rd</sup> year:

- Evaluate, improve and optimize proposed DL People Tracking approach.
- Writing papers and PhD manuscript.

## Main activities

The Inria STARS team is seeking for a Ph.D. researcher with strong background in computer vision, deep learning and machine learning.

The candidate is expected to conduct research related to the development of computer vision algorithms for video understanding.

**Main activities:**

- Analyze the requirements of doctors and patients/end-users and Study the limitations of existing solutions.
- Propose a new algorithm for detecting the behaviors of patients/end-users
- Evaluate and optimize proposed algorithm on the targeted video datasets
- Oral presentation and Write reports
- Submit a scientific paper to a conference

## Skills

Candidates must hold a Master's degree or equivalent in Computer Science or a closely related discipline by the start date.

The candidate must be grounded in computer vision basics and have solid mathematical and programming skills.

With theoretical knowledge in Computer Vision, OpenCV, Mathematics, Deep Learning (PyTorch, TensorFlow), and technical background in C++ and Python programming, and Linux.

The candidate must be committed to scientific research and substantial publications.

In order to protect its scientific and technological assets, Inria is a restricted-access establishment. Consequently, it follows special regulations for welcoming any person who wishes to work with the institute. The final acceptance of each candidate thus depends on applying this security and defense procedure.

## Benefits package

- Subsidized meals
- Partial reimbursement of public transport costs
- Leave: 7 weeks of annual leave + 10 extra days off due to RTT (statutory reduction in working hours) + possibility of exceptional leave (sick children, moving home, etc.)
- Possibility of teleworking (after 6 months of employment) and flexible organization of working hours
- Professional equipment available (videoconferencing, loan of computer equipment, etc.)
- Social, cultural and sports events and activities
- Access to vocational training
- Social security coverage
- Supplementary social protection

## Remuneration

- Duration: 36 months
- Location: Sophia Antipolis, France
- Gross Salary per month: 2051€ per month (year 1 & 2) and 2158€ per month (year 3)

## General Information

- **Theme/Domain** : Stochastic Methods and Models  
Scientific computing (BAP E)
- **Town/city** : Sophia-Antipolis
- **Inria Center** : [Centre Inria d'Université Côte d'Azur](#)
- **Starting date** : 2022-11-01
- **Duration of contract** : 3 years
- **Deadline to apply** : 2022-10-12

## Contacts

- **Inria Team** : [STARS](#)
- **PhD Supervisor** :  
Brémond François / [Francois.Bremond@inria.fr](mailto:Francois.Bremond@inria.fr)

## About Inria

Inria is the French national research institute dedicated to digital science and technology. It employs 2,600 people. Its 200 agile project teams, generally run jointly with academic partners, include more than 3,500 scientists and engineers working to meet the challenges of digital technology, often at the interface with other disciplines. The Institute also employs numerous talents in over forty different professions. 900 research support staff contribute to the preparation and development of scientific and entrepreneurial projects that have a worldwide impact.

## The keys to success

- Essential qualities in order to fulfil this assignment are feeling at ease in an environment of scientific dynamics and wanting to learn and listen.
- Passionate about innovation, willing to go for a PhD thesis in the field of Computer Vision and Machine Learning.

Languages: English

- Relational skills: team work
- Other valued appreciated: leadership

**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.

## Instruction to apply

Before applying, it is strongly recommended that you contact the Scientific manager beforehand.

### Defence Security :

This position is likely to be situated in a restricted area (ZRR), as defined in Decree No. 2011-1425 relating to the protection of national scientific and technical potential (PPST). Authorisation to enter an area is granted by the director of the unit, following a favourable Ministerial decision, as defined in the decree of 3 July 2012 relating to the PPST. An unfavourable Ministerial decision in respect of a position situated in a ZRR would result in the cancellation of the appointment.

### Recruitment Policy :

As part of its diversity policy, all Inria positions are accessible to people with disabilities.

