# INF6804 Computer Vision
# H2024 – Practical Assignment 3
# Detection and visual object tracking

---

**Objectives:**

- Allow the student to learn about object detection and/or tracking methods for image sequences.

**Submission:**

- *All your source code* (we should be able to run your tests)
- A report (*.pdf format* of 8 to 12 pages with font size of 10)
- Submit before April 12, 5:00 PM, on Moodle –*late submissions will not be accepted*
- *You must also submit your report on TurnItIn*
  - Register at www.turnitin.com using the info available on Moodle!

**References:**

- See course notes on Moodle (Chapter 4)

## Presentation

In this assignment, you will have to propose a method capable of detecting and tracking multiple objects of interest through a video sequence. The objects in question are identified via a bounding box and an unique identifier associated to each detected objects. The objects must be re-located and keep the same identifier as previously.

The objects of interest for this Assignment are cups. We provide on Moodle a video sequence for which only the first bounding boxes are available in the *init.txt*. You will also find in the *'readme.txt'* file the instructions to follow for submission. Your final results will be evaluated on this sequence after the submission deadline using our "secret" groundtruth. Note that the outcome of this evaluation does not influence your grade in this assignment. As always, the presentation of the method, its strengths and weaknesses, and your experiments are the most important parts of this work. The evaluation will focus on the *HOTA* metric, which involves the calculation of the detection accuracy *DetA*, the association accuracy *AssA*, and the localization accuracy *LocA* scores between your predictions and the groundtruth of the sequence provided on Moodle. You are encouraged to evaluate your method using these different scores, including the calculation of *HOTA(0)* with a localization threshold $\alpha$ of **5%**. To ensure the performance of your method during its development, you must evaluate it using another public dataset (e.g., MOT17 or MOT20).

For the method itself, you can once again implement your own solution, or use/modify an existing one. The choice of the solution must reflect the challenges present in the video sequence provided on Moodle.

In your report, you have to include the following elements (marked on 20 pts):

1. Detailed description of the solution (3 pts) :

   In your own words, describe in details your chosen method. What are its strengths/weaknesses?

2. Identification of the challenges for the Moodle sequence (3 pts) :

   Identify the challenges specific to the video sequence provided on Moodle for the detection and tracking of the objects of interest. In other words, what makes this sequence hard to process?

3. Justification of the solution w.r.t. identified challenges (2 pts) :

   Justify the choice of the proposed solution with respect to the challenges identified above. Which problems would your solution solve easily? Which might still cause some issues, and why?

4. Description of the implementation used (2 pts) :

   Describe the implementation of the proposed method. If you did not write all the code yourself, where does it come from? Did it require modifications? Otherwise, from which papers or websites did you inspire yourself to write it? In all cases, what are the primary parameters of your methods? How did you set their values?

5. Experimentation results for validation tests (3 pts) :

   Provide the evaluation results from your validation experiments on the other public dataset (e.g. MOT17 or MOT20). Use a proper format for their presentation — tables, figures, ...

6. Discussion of results (4 pts) :

   How do you explain the difference between the results obtained with *HOTA* and *HOTA(0)*. Explain the strengths/weaknesses of your method depending on the different scores calculated to obtain the *HOTA* metric. Among the challenges identified in the second point, which challenges seem to be resolved by your solution? Which are not? Can we expect a similar performance for the sequence on Moodle?

7. Readability and completeness (3 pts) :

   In addition to the content, the format must be conscientious and complete.

During the lab periods, do not hesitate to ask questions — we can help you with any technical issue if you are working on Windows/Linux, or if you are coding in C/C++, Python or Matlab.

You will be penalized by 50% of the total grade if you do not hand in your code. Also, if your report is not submitted to TurnItIn, it will not be graded. The order of presentation for the topics listed above does not matter, as long as they are all present.

## Resources

Datasets:

- MOT Benchmark (https://motchallenge.net/)

Evaluation metrics:

- HOTA metric (https://arxiv.org/pdf/2009.07736.pdf)
- TrackEval (https://github.com/JonathonLuiten/TrackEval)

Vision libraries:

- OpenCV (https://docs.opencv.org/4.0.0/d9/df8/tutorial_root.html)
- scikit-image (https://scikit-image.org/docs/stable/auto_examples/index.html)

Deep learning frameworks:

- PyTorch (https://pytorch.org/tutorials/)
- Tensorflow (https://www.tensorflow.org/tutorials)

Python:

- Guide (https://wiki.python.org/moin/BeginnersGuide/Programmers)
- NumPy (https://docs.scipy.org/doc/numpy/user/quickstart.html)

Matlab:

- Guide (http://www.mathworks.com/help/pdf_doc/matlab/getstart.pdf)
- Cheatsheet (http://web.mit.edu/18.06/www/Spring09/matlab-cheatsheet.pdf)