# COMPUTER VISION

## Sport video analysis

### Introduction

Sport video analytics relies on computer vision systems to identify and track players on sports teams and generate semantic information about the game, such as understanding the boundaries of the playing field and the context of a game action (e.g., checking offside situations). Such information provides a very interesting input to reporters or team coaches, who analyze how individual players move during the game and identify patterns in their behavior.

The goal of this project is to develop a computer vision system that can analyze video footage of various sport events (e.g., basket, soccer), providing high-level information such as player detection and team classification, recognizing team strategies and playing field boundaries. In more detail, the system to be developed should be capable of:

1. Localizing using a bounding box all the players in the image inside the playing field;
2. Segmenting the playing field and players from the rest of the image;
3. Classifying all the detected players according to their corresponding team.

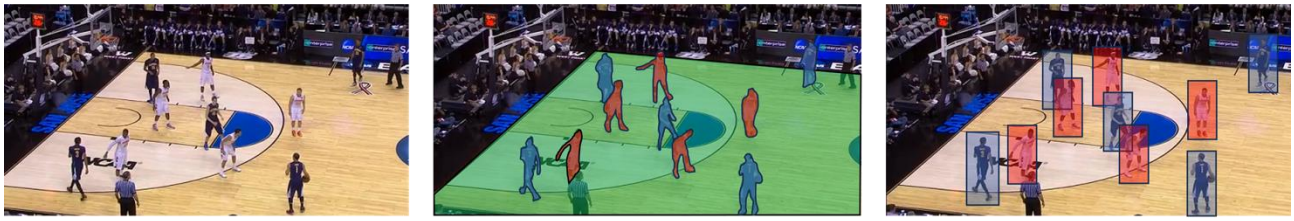

*Figure 1: Example of the system outputs. From left to right: input image, players and playing field segmentation, players detection and recognition (each player team is associated to a different class).*

As case studies for the development of the required system, various types of team sports are considered, like soccer, basketball, volleyball, hockey, and rugby. The system to be developed should be robust to all such different scenarios. In particular:

- it should recognize all the players within the playfield and their corresponding team; in some scenes, one of the players might be dressed in a different color with respect to the other team members (e.g., goalkeeper) - in such case, you are allowed not to associate a team to that player;
- it should ignore any non-player person (e.g., referees, spectators) outside the playing field;
- it should segment different types of playing field, considering different geometries and colors.

To assess the robustness of your system a benchmark dataset with annotations is provided at the following link:

https://drive.google.com/file/d/1L-P6ZAu1zgdvVtOH8hnLK68fSOmtr8K0/view?usp=sharing

## Performance measurement

For measuring the system performance, you should have a look and understand the following metrics:

- The mean Average Precision (mAP) - https://learnopencv.com/mean-average-precision-map-object-detection-model-evaluation-metric/
- The mean Intersection over Union (mIoU) - https://towardsdatascience.com/metrics-to-evaluate-your-semantic-segmentation-model-6bcb99639aa2

Such metrics shall be used to evaluate the sport analysis system as follows:

- For players localization, the mean Average Precision (mAP) calculated at IoU threshold 0.5;
- For player and playing field segmentation segmentation, the mean Intersection over Union (mIoU) metric, that is the average of the IoU computed for each class (background, team A, team B and playfield).

The metrics mentioned above need to compare the output of your system against the ground truth, namely what is considered "the truth" for each output image. Ground truth annotations are already included in the dataset provided for the final evaluation of your system. **Please note that the ground truth cannot be used as an input of your algorithm**.

The ground truth is stored in a set of files, one file for each input image. The information representing the ground truth (e.g., the rectangle enclosing the player or the pixels belonging to each player in the image) is expressed in such a file based on the following standard:

- For player detection: every player is found inside a rectangle defined by 5 parameters [x, y, width, height, team ID], where (x,y) are the top-left corner coordinates and width and height are the bounding box main dimensions; the $5^{th}$ parameter is the player's team ID; such parameters are listed in a row, one player per row;
- For player and playing field segmentation: a grayscale mask where to each pixel is assigned the corresponding category ID (background, team A, team B and playfield).

Feel free to add more test images, taken from the internet or acquired using your camera. In case you use additional images for the development of your system, you must include those images and related annotations (or a link to them) in your final delivery. If you use additional images, you are free to define your own standard. If you agree with other groups to share the ground truth collection, be sure to share the standard. The organization of the ground truth collection is completely free; if you wish, you can use the dedicated section in the moodle forum.

## Project delivery

The project must be developed in C++ with the OpenCV library. The only allowed exception is the usage of Python code for the deep learning part if you decide to exploit this family of techniques. **The project cannot be developed using machine learning / deep learning only; in particular you are not allowed to rely on state-of-the-art people detector or human pose estimator**.

You need to deliver your project including:

- All the source code (both C++ and Python), considering:
  - That each file must be written by **only one group member**;
  - **Each file must indicate its author in a comment at the beginning**;
- CMake configuration files (the use of CMake is mandatory);
- A report (no page limit) presenting your approach and **the performance measurement on the dataset provided and linked above**. You shall report **the metrics and the output images** for every element in the dataset.

When delivering your project, you should **clearly identify** the contribution of each member in terms of ideas, implementation, tests and performance measurement. You can organize the work as you prefer: you are not forced to assign one specific step to each group member. Please also include **the number of working hours** per person in the report. This is needed for a monitoring on our side of the effort requested – the evaluation will not depend at all on the number of working hours, but on the quality of the result.

You should include in your report the results of the player detection, player and playing field segmentation and team classification (both output images and metrics values) for the test images in the provided benchmark dataset.

**It is not allowed to include code that was not written by any of the group members.**