Journée GT2 Robotique Marine/Sous-marine

Deep Learning for Underwater Gesture Identification From Airbone Training

Prepared by: Bilal GHADER

Advised by:
- Vincent HUGEL (Direction)
- Éric WATELAIN (Co-Direction)
- Claire DUNE (Encadrement)
- Jeremy Nicola (Encadrement NotiloPlus)

Problem statement

How to allow an intuitive interaction between the drone and the diver?

Objectives

- **Relative positioning of the robot with respect to the diver**
- **Replacing the current solution (manual remote control) by gesture classification**
- Positioning within a team of divers
- surveillance of the team to detect any incident and offer assistance

Diver and robot, on top of the diver we can see the ceramic of the remote

The remote-control system, with the remote control and the ceramic acoustic emitter

State of the art: diver gesture classification

Different entries

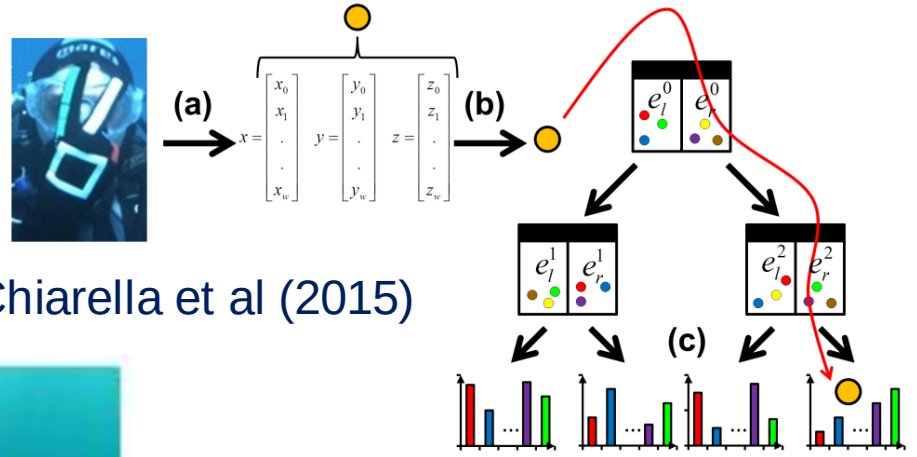
- Monocular-Stereo
- Specifically built gloves
- SONAR images

Classification

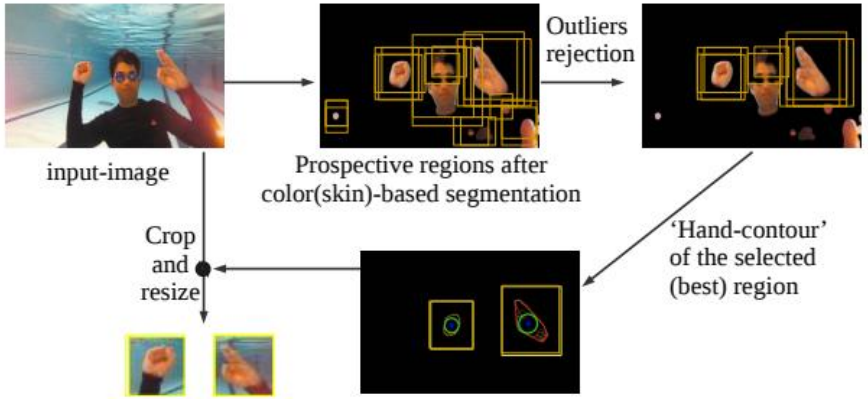
- Classical ML algorithms
- DeepLearning Networks

Hand / extraction

- skin-color segmentation
- Specifically designed gloves with colored features
- Deep Learning with region proposal phase

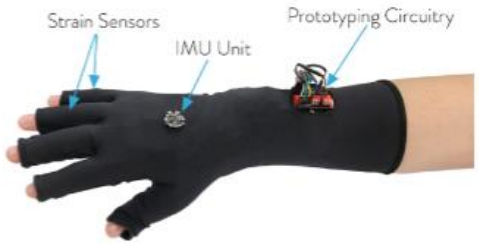


Chiarella et al (2015)



Islam et al (2018)

Rodd-Cowney and Jenkins (2019)



Nađ et al. (2019)

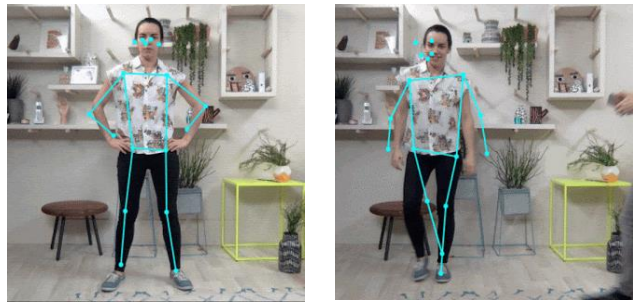
Diver detection and human pose estimation

Diver detection:

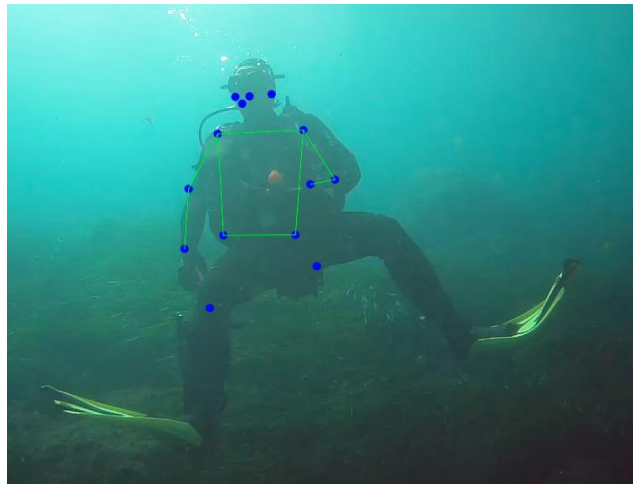
- We aim to test different diver detection techniques to test and see their limitation

Diver skeleton:

- We test different skeleton detection techniques (3D and 2D) to estimate to which extent they work underwater
- There is a need to adapt airborne solution to underwater environment



PoseNet demo on human in airborne environment



PoseNet applied on a data of human diver where the diver is in a vertical direction

Diver recognition network by Islam et al (2019), on the left we can observe a confusion of two divers for one

Underwater databases

Caddy Dataset

EU FP7 project "Cognitive autonomous diving buddy (CADDY). object classification, segmentation and human pose estimation tasks.

1. **Hand gesture:**
stereo camera recordings (≈10 K) in different environmental conditions.
2. **Position:**
stereo footage (≈12.7 K) of divers free-swimming, synchronized IMU located throughout the diver's suit (*DiverNet*), as ground-truth for human pose and tracking methods.

ScubaNet

EU FP7 project "Cognitive autonomous diving buddy (CADDY). object classification, segmentation and human pose estimation tasks.

1. **Diver detection:**
1000 annotated diver images with labels for Head, Hands and Body(Diver).
2. **Gesture:**
5 subjects, multiple gestures (24) not complete. Monocular video sequences.

VDD-C: Video Diver Detection Dataset

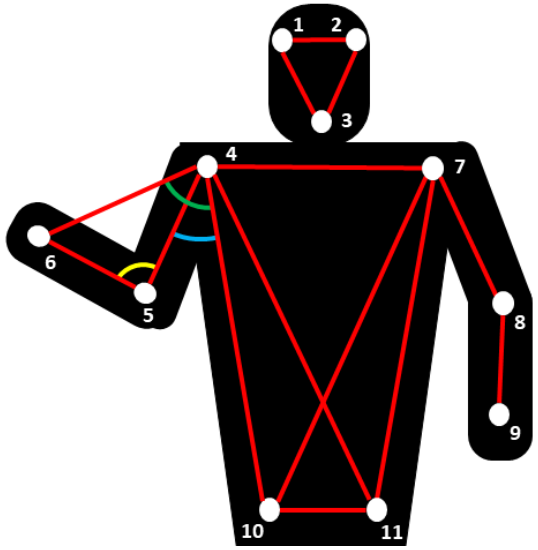
105,000 fully annotated images of divers**, drawn from videos taken in pool and field environments**

photographer bias in photo-based datasets means that video datasets have more realistic translations, rotations of divers.

Gesture classification

Objective: classify 8 diver gesture

- Without hands
- Learning with temporal trajectories data of upper body parts
- LSTM-based neural network classifier
- Network entry: (2 axis angle + 1 angle)
- (9 parameters per arm)



Different tracked angles in our classification method

Calculations

$u = (10,11),\ v = (10,4)$

$w = u \times v$
$v' = w \times u$

Frame (u, v', w) defined as a basis

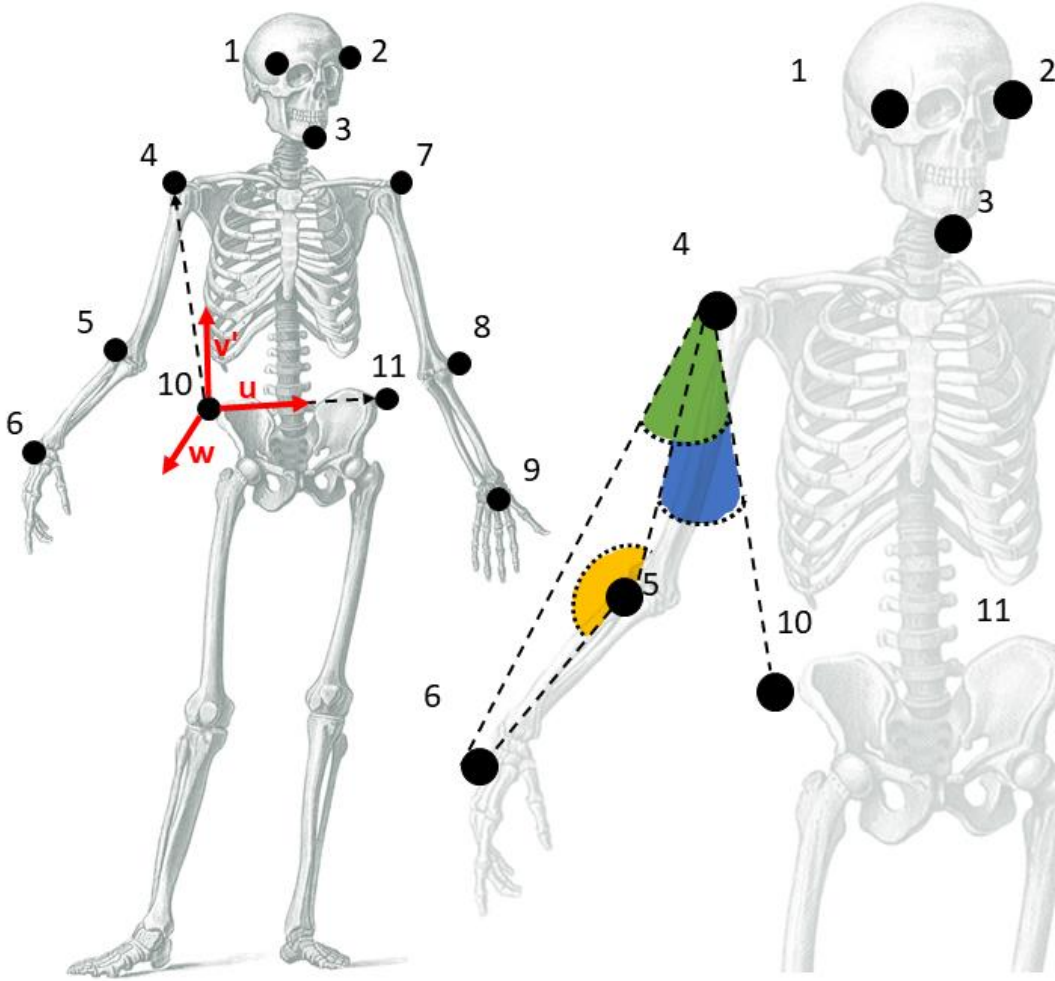
$$^{b}R_{w} = [u\ |v'|\ w\ ]$$

Rotate all the point at all time step using this

$$\sin\theta = \frac{|x \times y|}{|x|\,|y|},\ \cos\theta = \frac{x \cdot y}{|x|\,|y|}$$

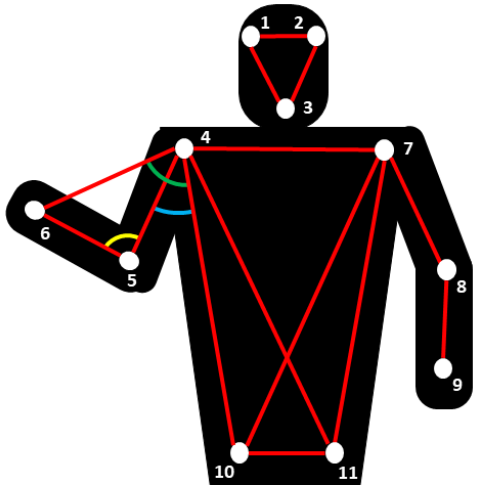
$$\theta = arctan2\ (\sin\theta\ , \cos\theta\ )$$

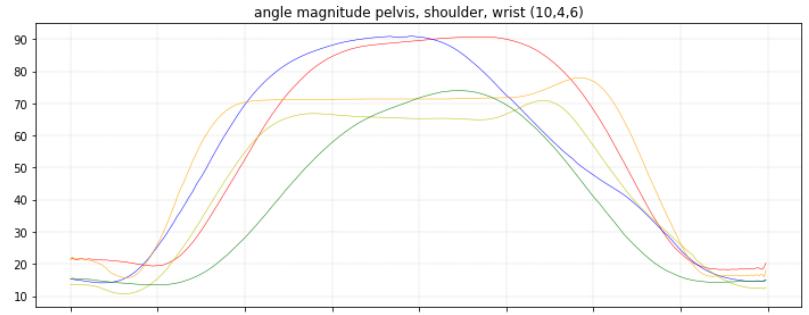
Renormalize each signal to 400 timesteps

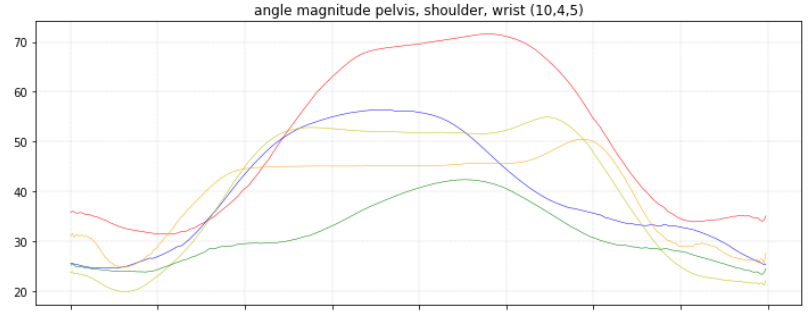


Gesture classification

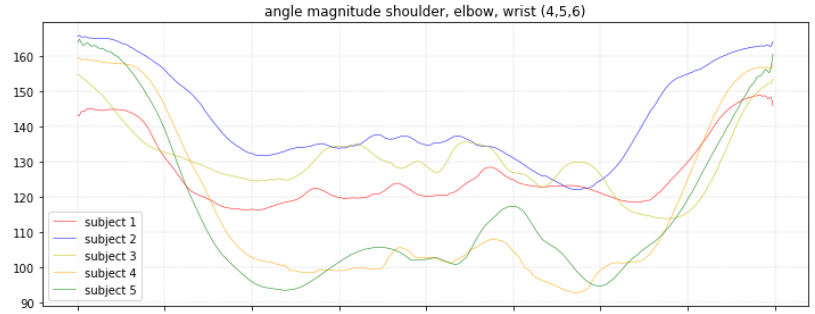
- 2 different gesture signal frames with more than one subject
- Data was treated using the Qualisys SDK
- We are only displaying the magnitude value of the angles
- Temporal data

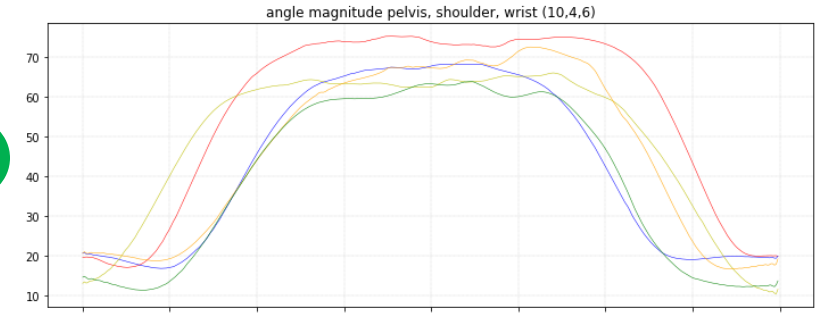






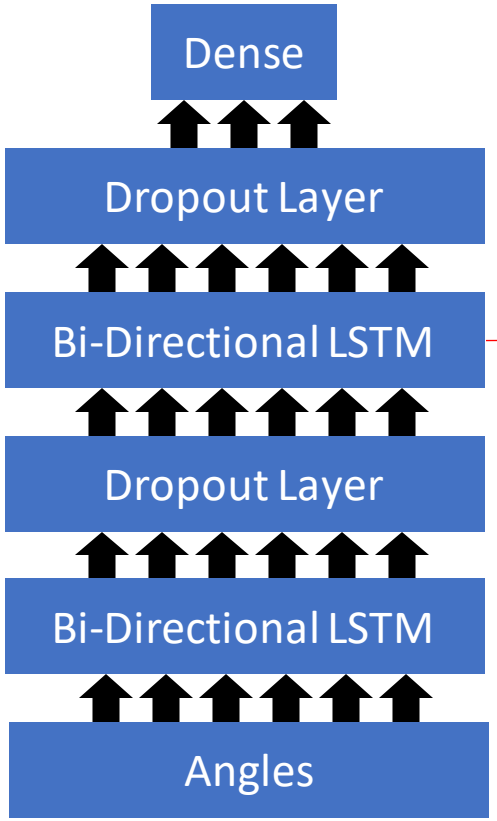
Go up

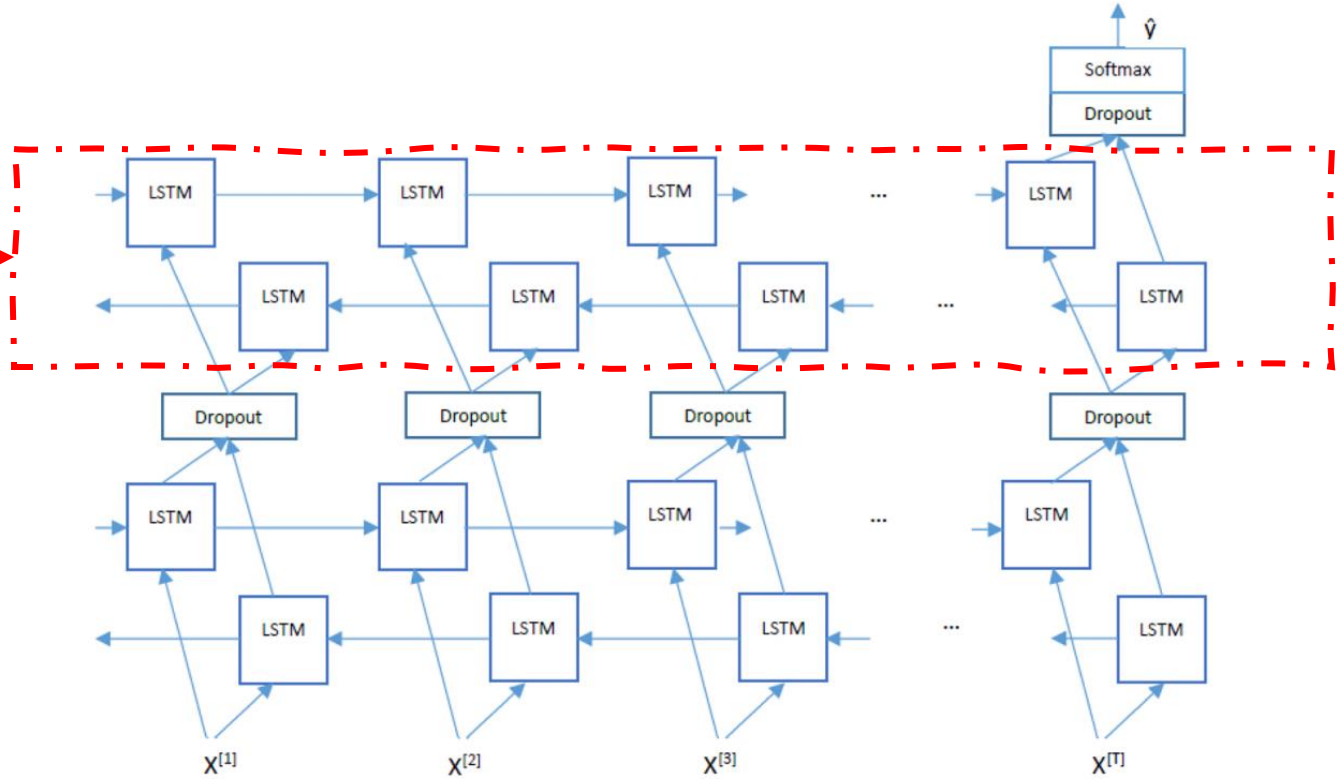




Not well

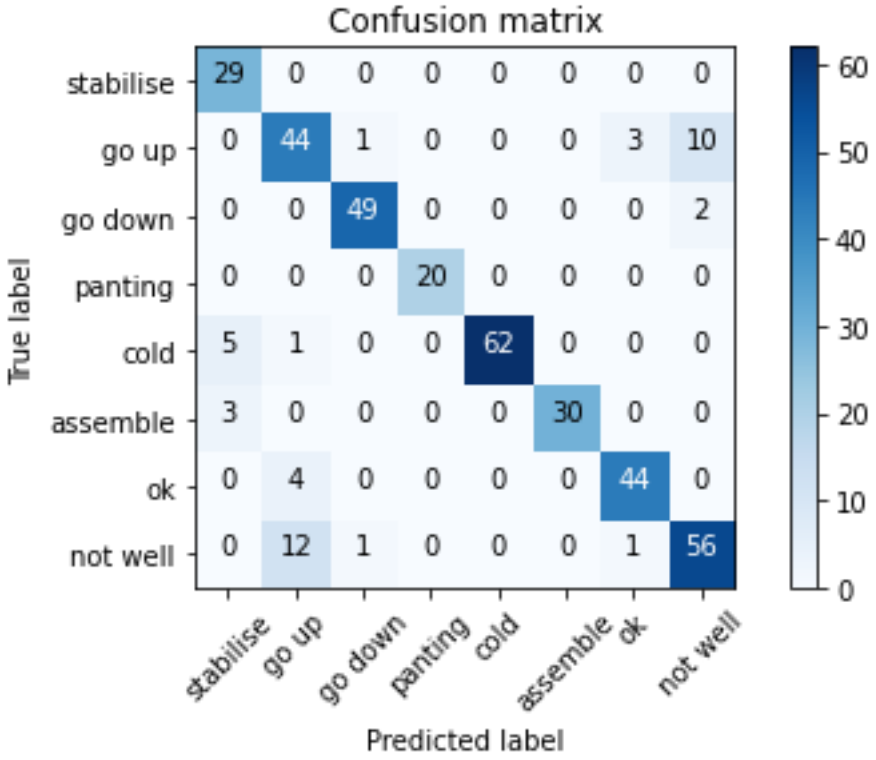
Gesture classification

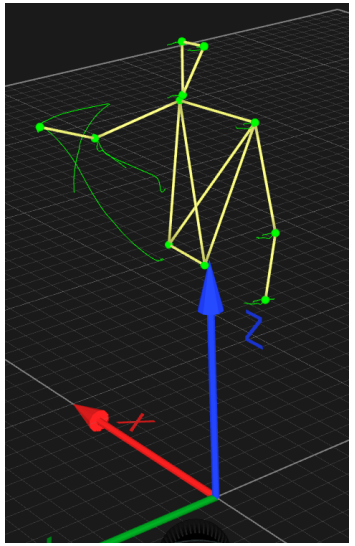


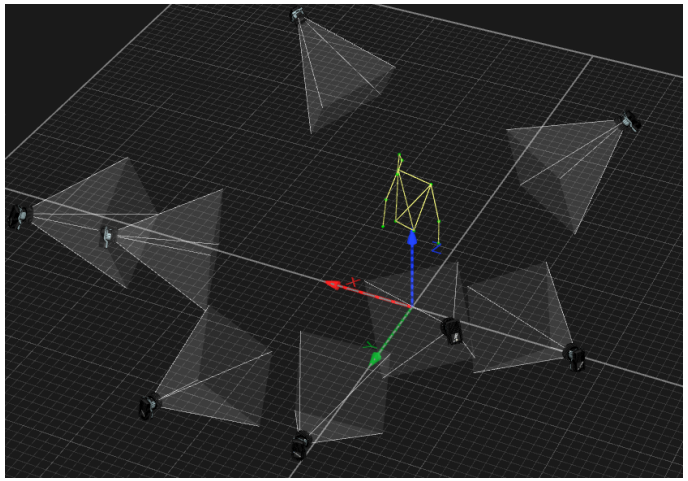


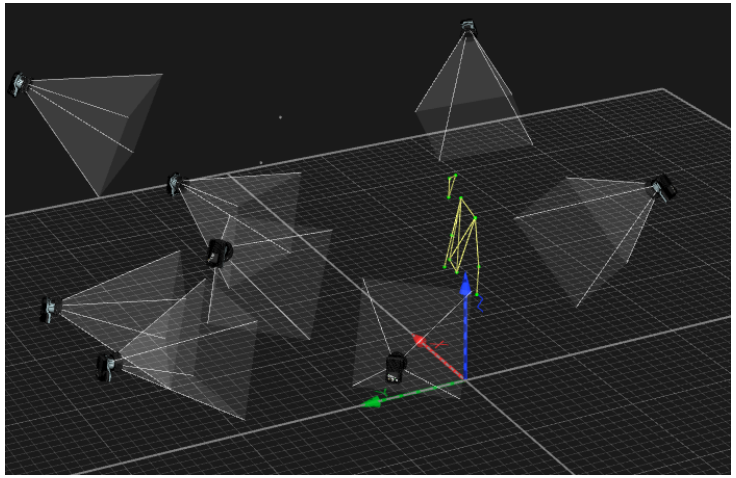
Gesture classification

- 17 different subjects
- Different diving levels (non initiated to professional)
- 8 different gestures with 10 repetitions each (20 for 2 hands)









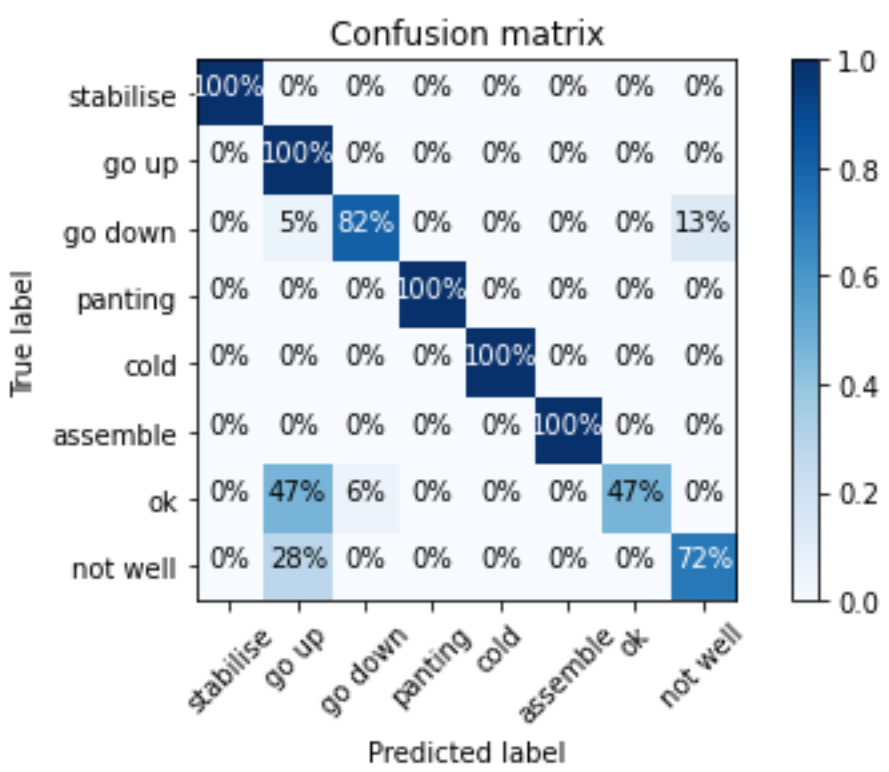
Gesture classification – Interpersons

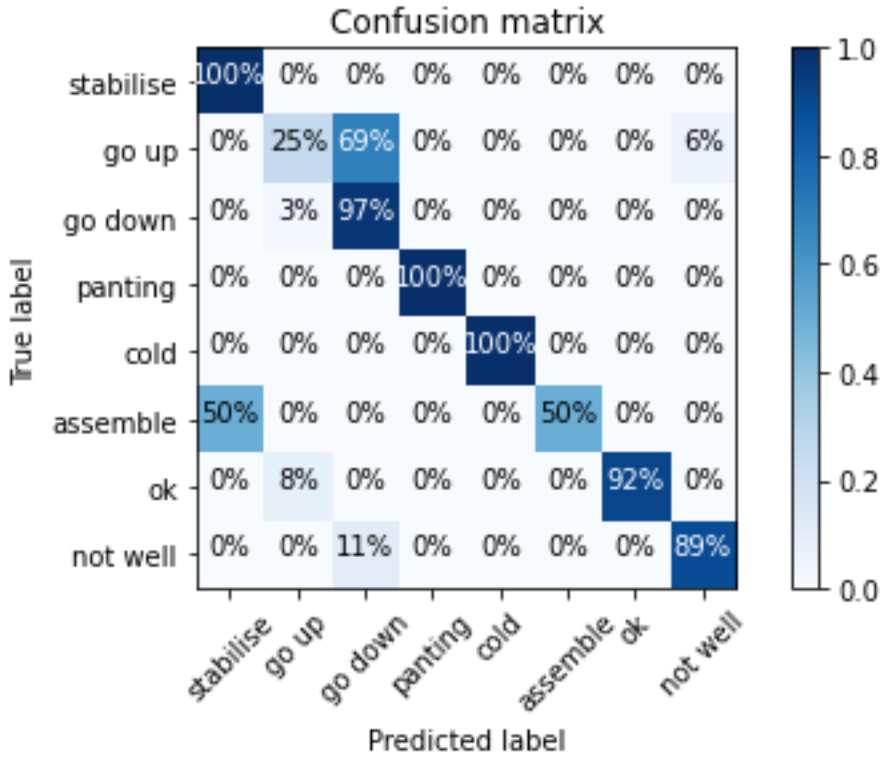
Interpersonal Transferability

Avoid over fitting over specific subject.

Trained with subject A, then without subject B respectively.

The validation matrix is on the subject not present in the training data





Confusion matrix of inference performed on subject A (left) and B (right). All the data related to the subject was removed from the initial training dataset and the network was retrained. The inference was then performed on the subject. This was carried out on both subjects A and B

Future Works

- **Diver gesture Classification**

  Verify the transferability of our method to other sensors

  - Aquatic Qualisys Data
  - Different Camera setup
  - RGBD Realsense
  - Monocular/stereo Images

- **Classification in real-time directly in the data flux**

- **Detect one diver and its orientation**

  The aim is not only to detect the diver (as it is done) but to also detect its orientation to allow the drone to position itself in front of the diver.

  - Detecting the mask (like hand detection)
  - Detection of the head?

Thank you for you attention !