RoboCup MSL
Scientific Challenge 2021

Playing with humans: improving embedded perception
Playing with humans: the (near) future of RoboCup MSL

- A major step on the road to 2050
Playing with humans raises major challenges

- Improving robot skills: *(technical challenge)*
  - Require to be competitive with humans
- Playing without wireless communications: *(technical challenge)*
  - Except for the referee

**Require improving drastically embedded perception**

- No more shared perception
- Requires Robots/humans/landmarks identification and localisation
- Limited hardware computational capabilities.
What is an improved perception?

- Identification – Who?
  - Teammates / opponents / referee / landmarks

- Positioning – Where?
  - Teammates / opponents / referee / landmarks

- Postural communication: What?
  - What is «told» by gesture of teammates / opponents / referee / coach?
    - See excellent Tech United Technical Challenge 😊
How perception is implemented in the brain?

- Vision preprocessing
  - Close to CNN
- Identification and posture: Who/What?
  - Cortex ventral pathway
  - Mainly based on vision
- Positioning: Where?
  - Cortex dorsal pathway
  - Coupling « sensors »
    - Vision / Stereovision
    - Acoustic perception

Key bioinspired ideas

- Complementary perception organs
- Parallel processings
How to implement advanced perception in robots?

- **Combining multiple sensors and processings is a key**
  - Lidar: precision (angular and radial), but 1D
  - Camera: 2D, color but no radial precision
  - SLAM for sensor fusion

- **Processing efficiency is important: several task in parallel**
Who / What? Omnidirectional camera + deep learning

**Contribution: a new technical choice**

- 185° fisheye pointed vertically to the ground.
- No more problem of misalignment of camera and mirror.
  - No need for hardware calibration
- Excellent optical quality.

- Image transformed to panorama
  - Require a software autocalibration (instead of hardware calibration)
    - Finding fisheye center to have a straight panorama image.
Who / What? Omnidirectional camera + deep learning

- Followed by image segmentation using CNN
  - Well-known Yolo V5 algorithm on standard cameras.
  - Doesn’t work on fisheye images directly
    - CNN are rotation sensitive
- Contribution: first time applied to unwarped spherical images
  - Requires an extended training set
    - Nubot TS + fisheye panorama samples
Who / What? Omnidirectional camera + deep learning

- Experimental performances:
  - Detection of 4 classes: goal / robots / balls / humans
    - AUC: 87%
  - Latency: nearly real time
    - 34-40ms (Yolo V5 large model): 25 fps
    - 10-15ms (small model): more than 60 fps
  - GPU: 6% (GTX 1060) - CPU: 20% (mainly for fisheye to panorama)

- Pros:
  - Able to segment complex scenes
  - Able to distinguish human/robots color (an idea: shirt could be enlarge in the future for easier analysis, like in real soccer)
  - Excellent angular resolution

- Cons:
  - Poor radial resolution
  - Computationally expensive
Where? Lidar: the perfect complement to omni-camera

- Excellent radial resolution all around the robot: 1 cm
  - Impossible with omnicamera
  - 360° difficult with stereovision.
- Allow landmarks extraction with precision
  - Example: room corners
    - Limited to simple situations using geometric methods
Where ? Lidar : the perfect complement to omni-camera

- **Contribution**: a low-computational cost 1D Lidar deep learning algorithm
- Finding landmarks in a scene (for further SLAM). Here:
  - Goal posts
  - Room / stadium corners.
- Can be found by 2D image analysis using CNN: computationally expensive
Where? Lidar: the perfect complement to omni-camera

**Contribution**: a low-computational cost 1D lidar deep learning algorithm

- Dedicated labelling software
- Using a custom 2 layers CNN
- Optimized for:
  - Object angle and distance
  - Classification

\[
\text{Loss} = \lambda_0 \sum_{i=0}^{C_x} 1^{obj}_i (\theta_i - \hat{\theta}_i)^2 + (d_i - \hat{d}_i)^2 \\
+ \lambda_{Pr1} \sum_{i=0}^{C_x} 1^{obj}_i (Pr_i - \hat{Pr}_i)^2 \\
+ \lambda_{Pr2} \sum_{i=0}^{C_x} 1^{noobj}_i (Pr_i - \hat{Pr}_i)^2 \\
+ \sum_{i=0}^{C_x} 1^{obj}_i \sum_{c \in \text{classes}} (c_i - \hat{c}_i)^2
\]
Where? Lidar: the perfect complement to omni-camera

- Results for detection of posts / balls / robots / humans
  - AUC 85%: good detection
  - Allow an efficient positioning

- Efficiency:
  - Angular error stddev: 0.025 rad (1.5°)
  - Radial error stddev: <10 cm
  - **Low computation time**: <1ms
    - -> low latency
    - Allow several processings in parallel
Final perception step: merging data in a single world map

- Implementing standard EKF SLAM:
  - Simultaneous localisation and mapping (SLAM)
  - Based on Extended Kalman Filter
  - **Require to have reliable landmarks**

- Will be presented next year 😊
Playing with humans: improving embedded perception

- Bio-inspired combination of sensors and efficient processings is the key:
  - Contribution 1: an omni camera with software calibration only
  - Contribution 2: an application of Yolo V5 to unwarped spherical images (Who/What?)

- Shared on GitHub: https://github.com/iutgeiitoulon/RoboCup2020

Recent team publications related to RoboCup MSL or robots multi-agent perception:

- Conf: 2019 RoboCup Symposium: Modelling and Optimisation of a RoboCup MSL coilgun
Thanks for your attention

Questions?