U.PORTO

FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Robotics / Intelligent Robotics
Robot Learning

Luís Paulo Reis, Nuno Lau, David Simões, Armando Sousa

lpreis@fe.up.pt

Director of LIACC – Artificial Intelligence and Computer Science Lab.

Associate Professor at DEI/FEUP – Informatics Engineering Department, Faculty of Engineering of the University of Porto, Portugal

President of APPIA – Portuguese Association for Artificial Intelligence

Robot Learning - Motivation

Programming Robots is a hard task

- No high-level programming language
- Sensors and actuators are noisy
- Robotics is moving towards increasingly unstructured environments

If only **robots could learn** how **to perform tasks** by themselves…

$\Rightarrow$ **Optimization and Learning in Robotics**

Motivation

Table-Tennis

Robots

Mülling + Peters

Humans

We need **learning** and **adaptation** to improve robot skills!

Motivation

Optimization vs. Learning

- Learning approaches use Optimization
- However there are some **differences**:
  - **Optimization** is focused on **training set**, while **Learning** is focused on **general performance**
  - The **loss function** used in learning is not always the real (optimized) loss function
  - **Stopping criteria** may be quite different

Motivation

Challenges in Robot Learning

- Cost of experimentation
- Cost of failure
- Limited data
- Generalization
- Curse of dimensionality
- Real time requirements
- Changes in environment
- Changes in task specification

Motivation

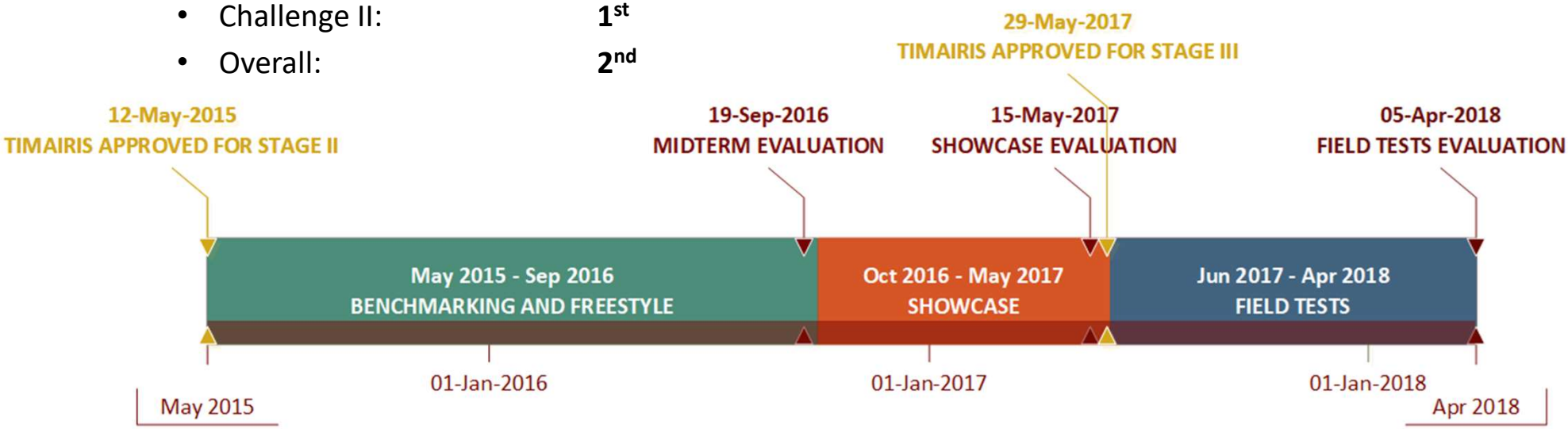
Learning in Robotics can be used for:

- Robot **Perception**
- Robot **Decision**
- Adapt **Human-Robot Interaction**
- Robot **Actuation** (Behaviors)
- Multi-robot **Coordination and Communication**

EuRoC - TIMAIRIS Team

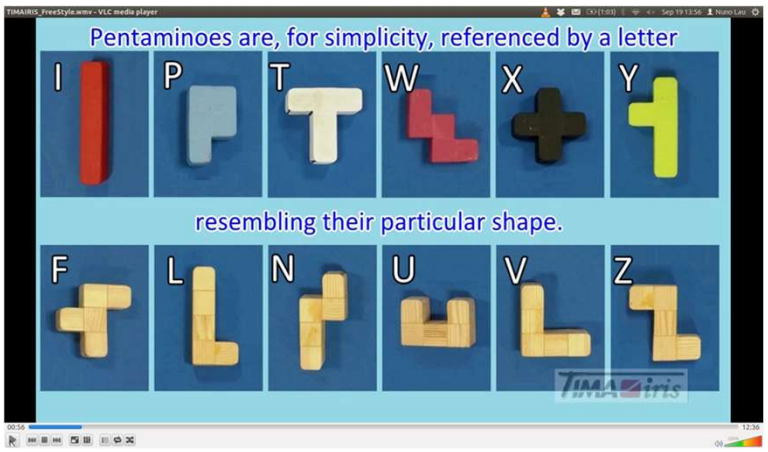
- Challenger: Universidade de Aveiro, Portugal
- End User: IMA S.p.A, Italy

- Stage I (Overall: 103 teams, Chall. II: 37 teams)
- Stage II Results (Challenge II: 5 teams)
  - Benchmarking: **1st**
  - Freestyle: **2nd**
  - Showcase: **1st**
- Stage III Results
  - Challenge II: **1st**
  - Overall: **2nd**

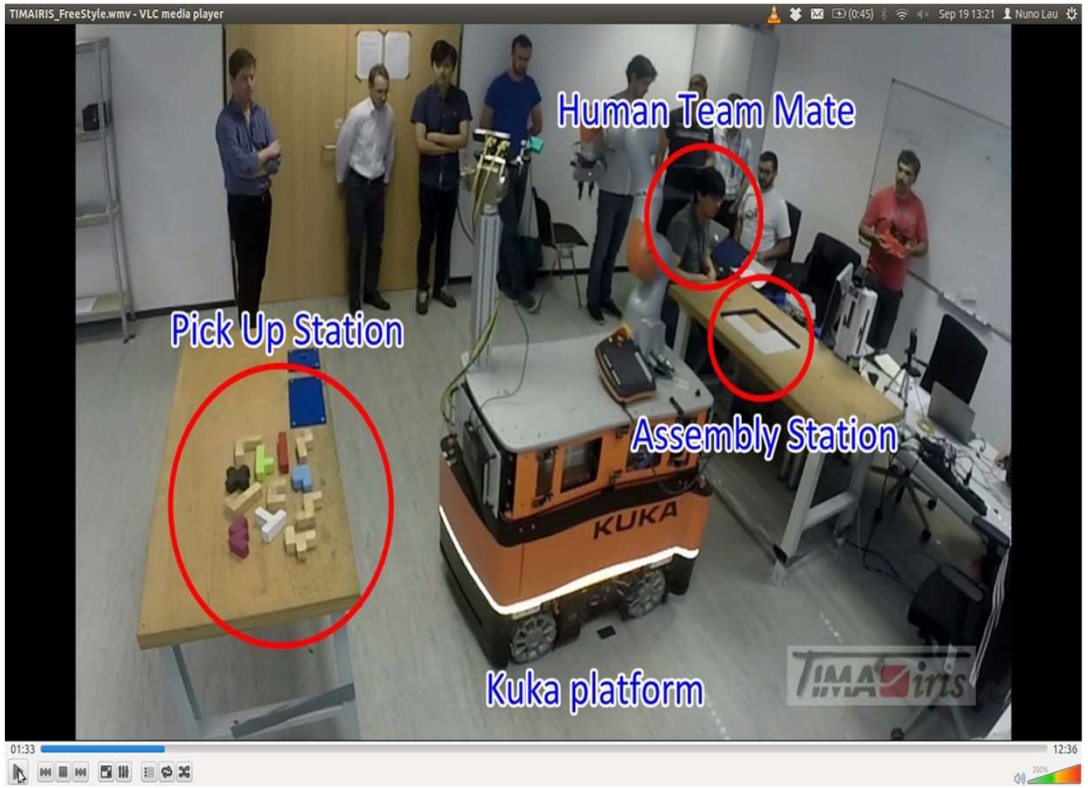


Gesture Recognition

Task: Assembling a puzzle cooperatively by a human and a robot (EuRoC Project)

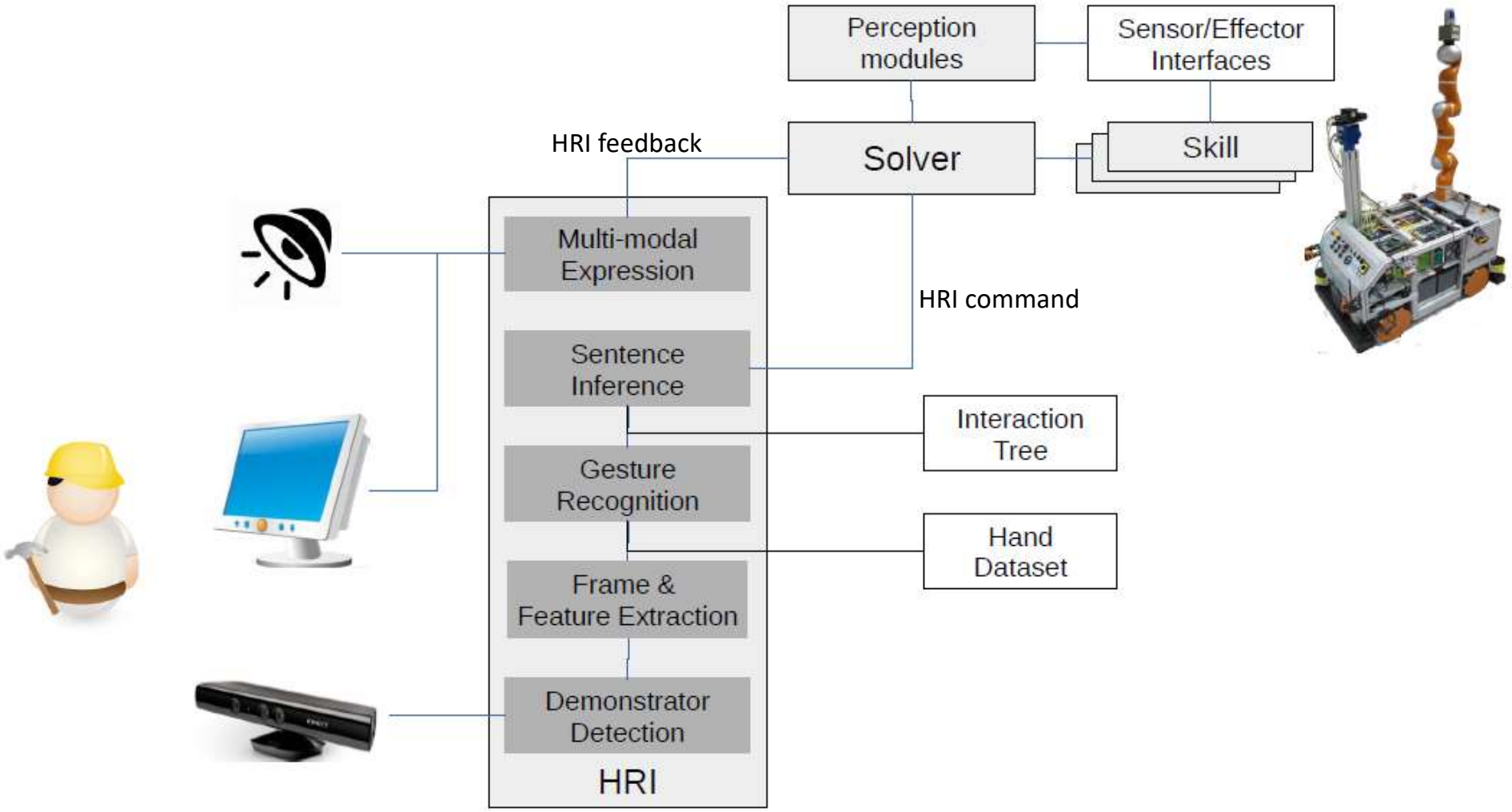


Set of 12 pentomino pieces



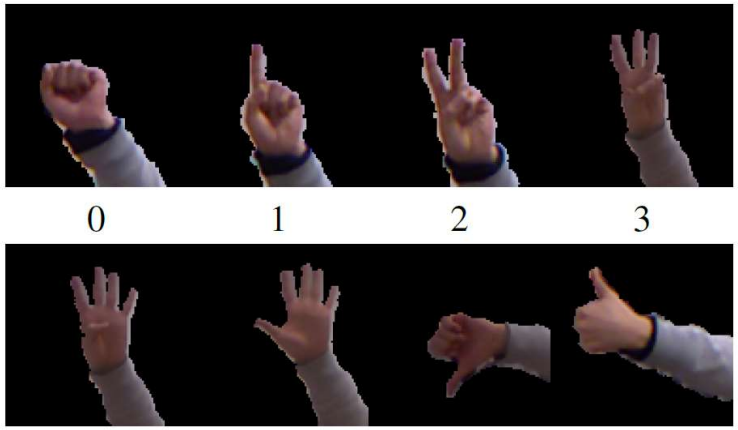
Task environment

Gesture Recognition



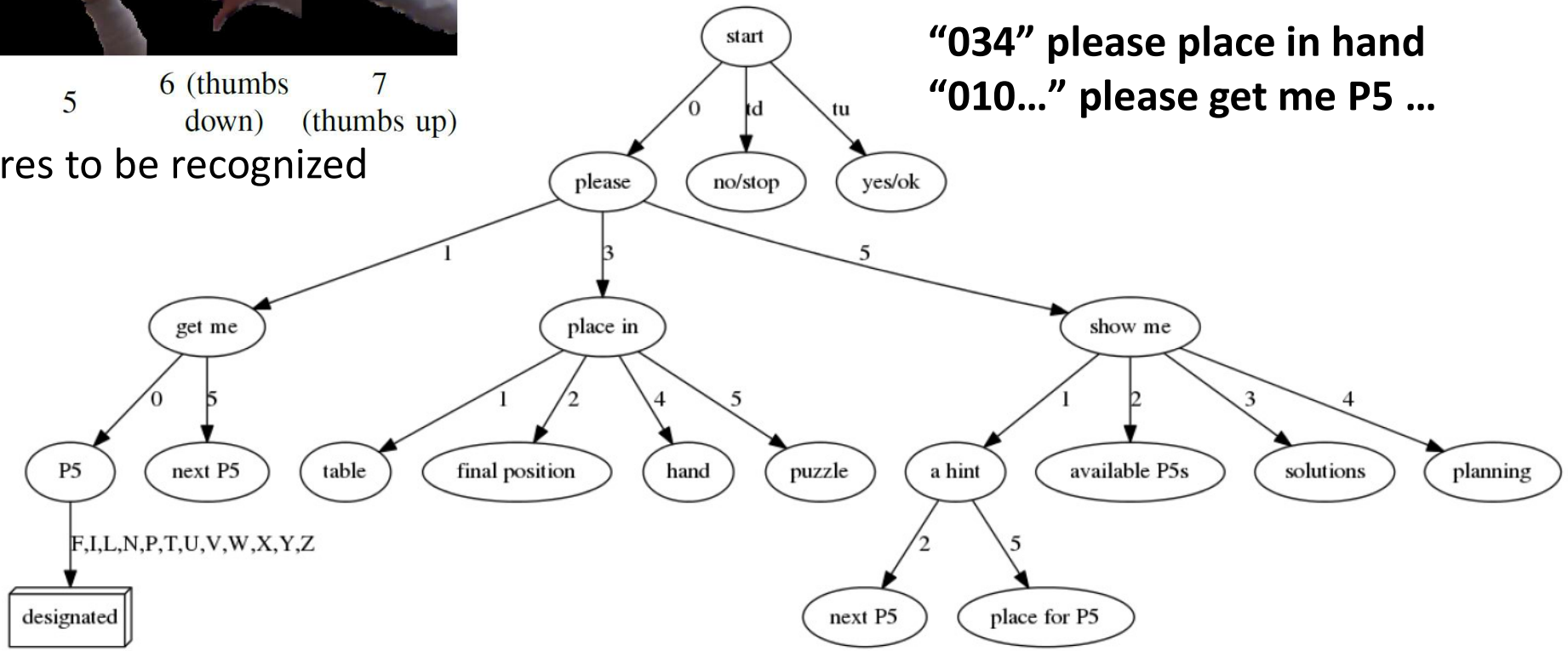
Human-Robot Interface architecture

Gesture Recognition



Gestures to be recognized

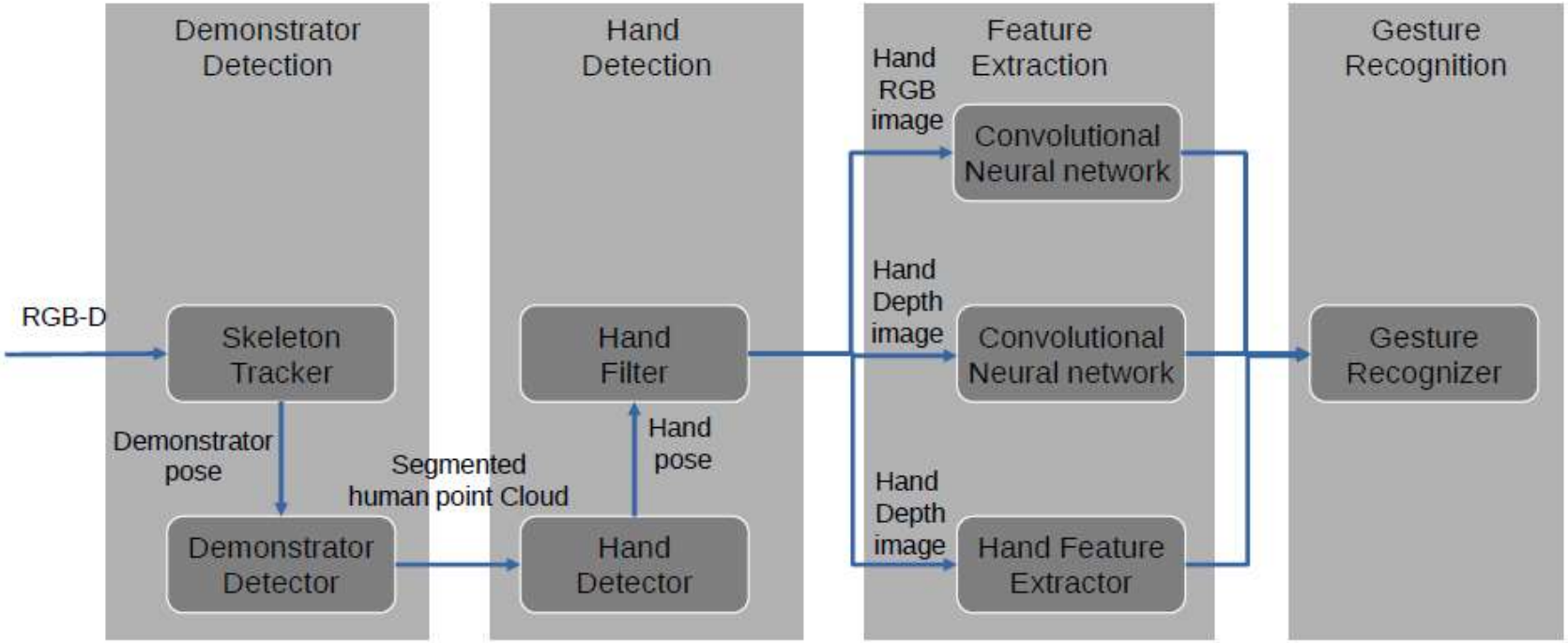
“034” please place in hand
“010…” please get me P5 …



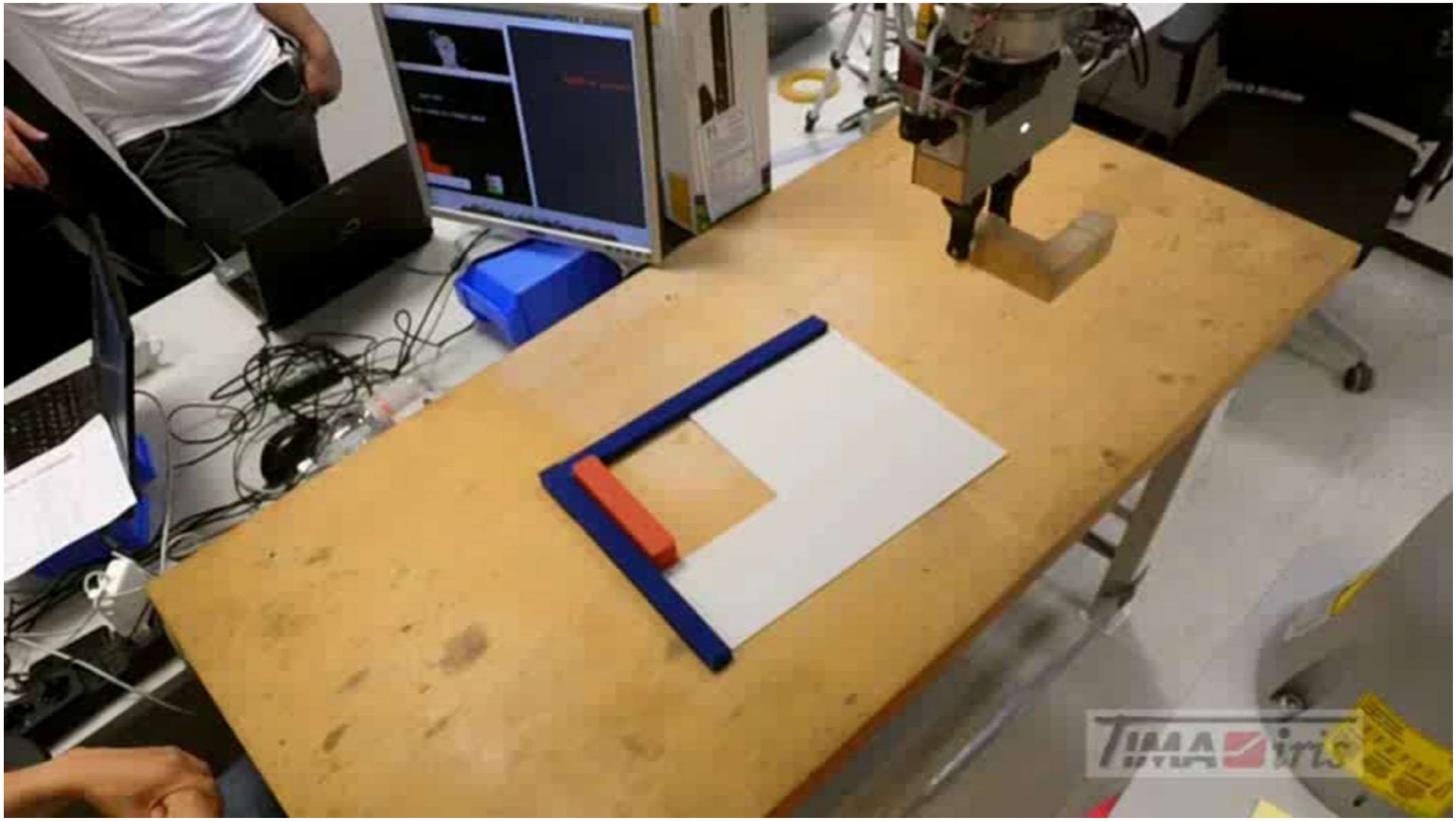
Interaction tree

Gesture Recognition

- **Task**: Recognize Gestures
- **Approach**:
  - 1st : Use Deep Learning
  - 2nd : Mix Deep Learning with other hand features



Gesture Recognition



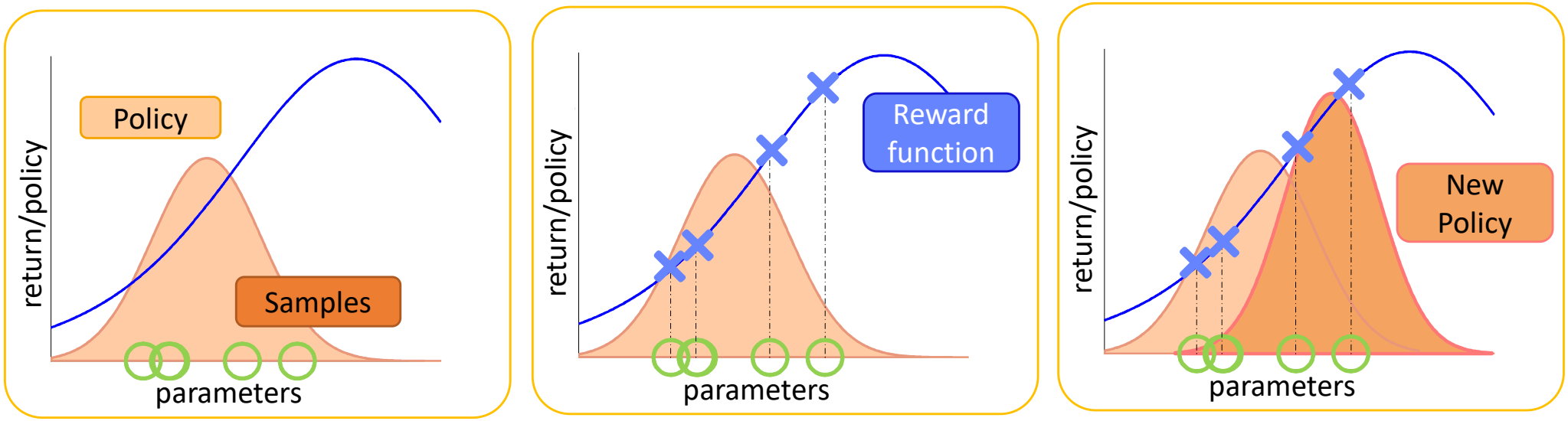
Lim, G.H. et al. Skill-based anytime agent architecture for European Robotics Challenges in realistic environments: EuRoC Challenge 2, Stage II — realistic labs, Robotics and Autonomous Systems, Vol. 120, 2019

Stochastic Search

$$\pi(\boldsymbol{w}) = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

- **Use Search-Distribution:** $\pi(\boldsymbol{w})$ $J_\pi = \int \pi(\boldsymbol{w}) R(\boldsymbol{w}) d\boldsymbol{w}$
- **Objective:** Find search distribution that maximizes

Explore **Evaluate** **Update**

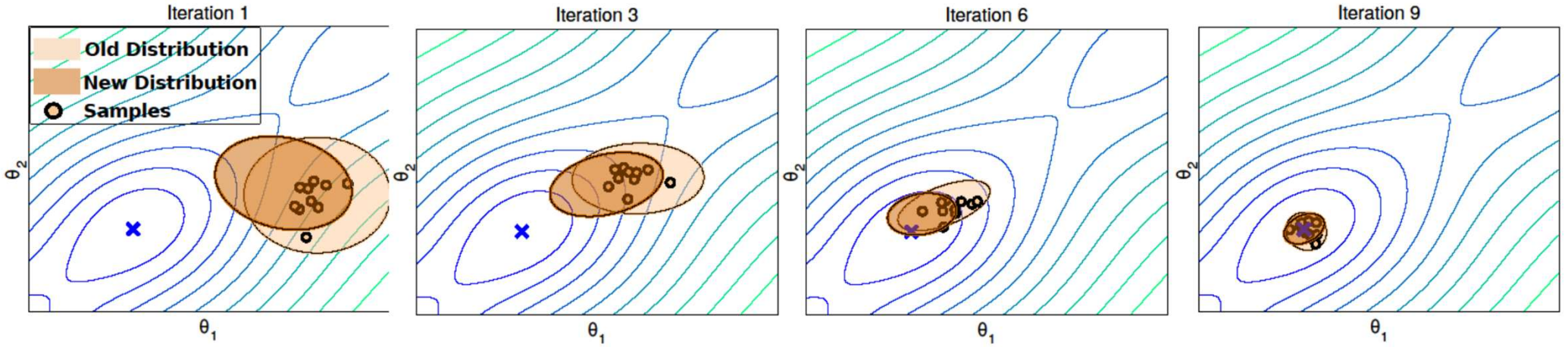


Stochastic Search

- **Use Search-Distribution:** $\pi(\boldsymbol{w}) = \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
  - **Mean:** Estimate of the maximum
  - **Covariance:** Direction to explore

- **Objective:** Find search distribution $\pi(\boldsymbol{w})$ that maximizes

$$J_\pi = \int \pi(\boldsymbol{w}) R(\boldsymbol{w}) d\boldsymbol{w}$$



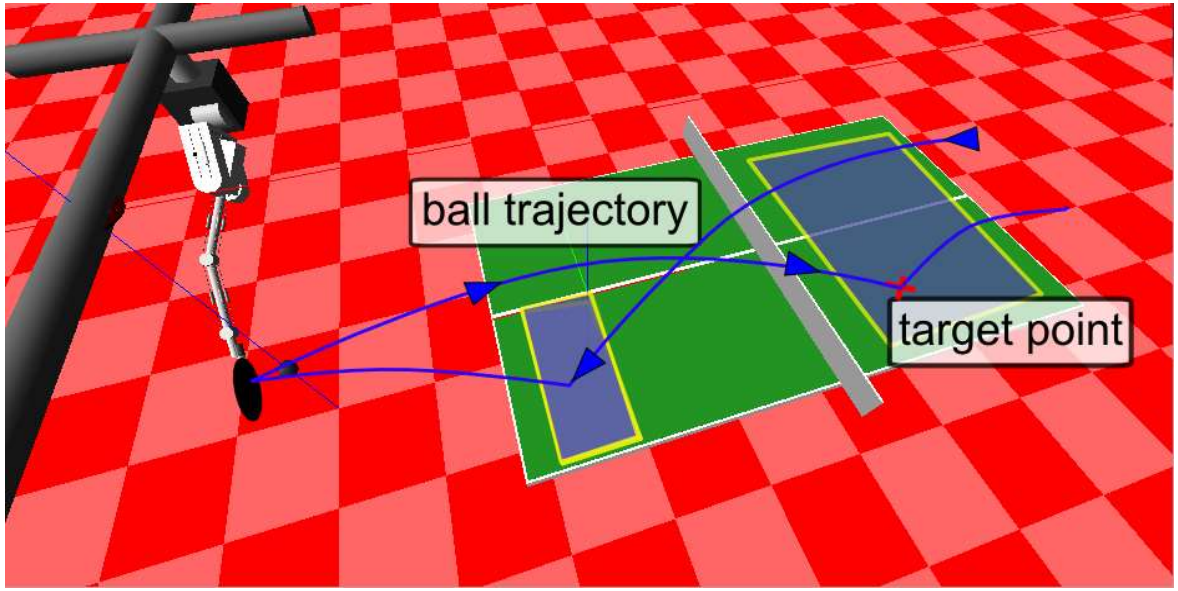
Contextual Stochastic Search

Goal: Adapt parameters w to different situations

- Different ball trajectories
- Different target locations

Introduce context vector s

- Continuous valued vector
- Characterizes environment and objectives of agent



Learn contextual search policy $\pi(w|s)$

Abdolmaleki, et. al, *Model-Based Relative Entropy Stochastic Search,* NIPS 2015
Abdolmaleki, A.et al., Contextual Direct Policy Search: With Regularized Covariance Matrix Estimation, Journal of Intelligent and Robotic Systems: Theory and Applications, 96 (2), pp. 141-157, 2019

Adaptation of Skills

Contextual distribution:

$$\pi(\boldsymbol{w}|\boldsymbol{s}) = \mathcal{N}(\boldsymbol{s}^T\boldsymbol{M} + \boldsymbol{m}, \boldsymbol{\Sigma})$$

Compatible Function Approximation:

$$R(\boldsymbol{s}, \boldsymbol{w}) \approx \boldsymbol{w}^T\boldsymbol{A}\boldsymbol{w} + \boldsymbol{s}^T\boldsymbol{B}\boldsymbol{w} + \boldsymbol{a}^T\boldsymbol{w} + a_0$$

Contextual distribution update:

1. Maximize **expected** return

$$\arg\max_{\pi} \mathbb{E}_{p(\boldsymbol{s})}\left[\int \pi(\boldsymbol{w}|\boldsymbol{s})R(\boldsymbol{s}, \boldsymbol{w})d\boldsymbol{w}\right]$$

2. Bound **expected** information loss

$$\text{s.t.:} \quad \mathbb{E}_{p(\boldsymbol{s})}\left[\mathrm{KL}\big(\pi(\cdot|\boldsymbol{s})||\pi_{\mathrm{old}}(\cdot|\boldsymbol{s})\big)\right] \leq \epsilon$$

3. Bound entropy loss

$$\underbrace{H(\pi_{\mathrm{old}}) - H(\pi)}_{\text{loss in entropy}} \leq \gamma$$

New distribution:

$$\pi(\boldsymbol{w}|\boldsymbol{s}) \propto \pi_{\mathrm{old}}(\boldsymbol{w}|\boldsymbol{s})^{\frac{\eta}{\eta+\omega}} \exp\left(\frac{R(\boldsymbol{s}, \boldsymbol{w})}{\eta + \omega}\right)$$

$$\propto \mathcal{N}(\boldsymbol{s}^T\boldsymbol{M}_{\mathrm{new}} + \boldsymbol{m}_{\mathrm{new}}, \boldsymbol{\Sigma}_{\mathrm{new}})$$

⇐ **Compatible Function Approximation**

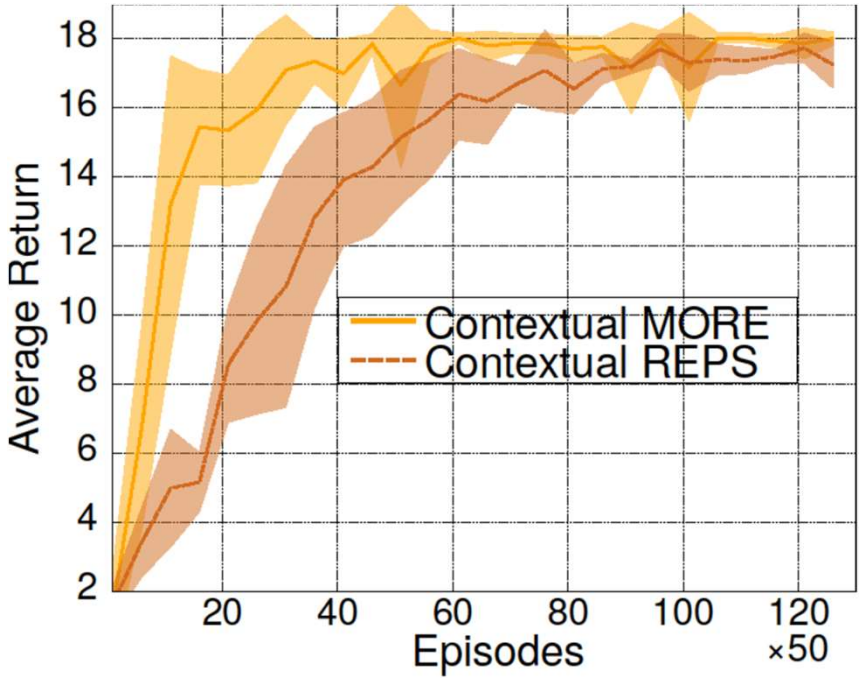
Adaptation of Skills: Table Tennis

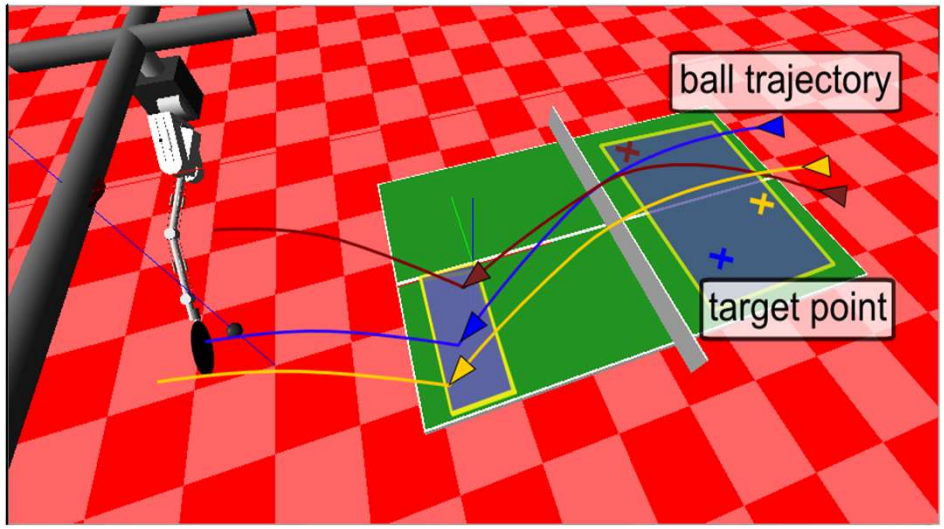
Contextual Stochastic Search:

- Context: Initial ball velocity

Reward:

- Hit ball
- Ball impacts at target position



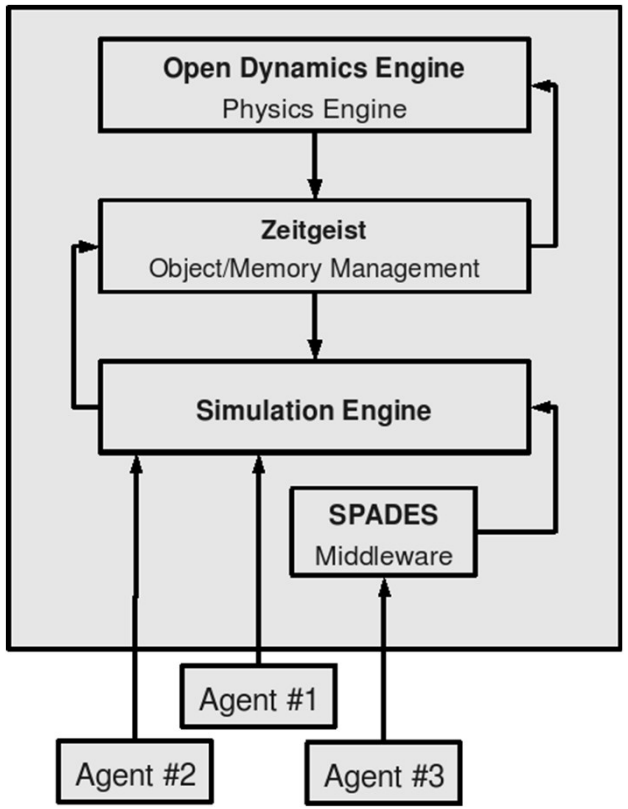


Skills Improvement:

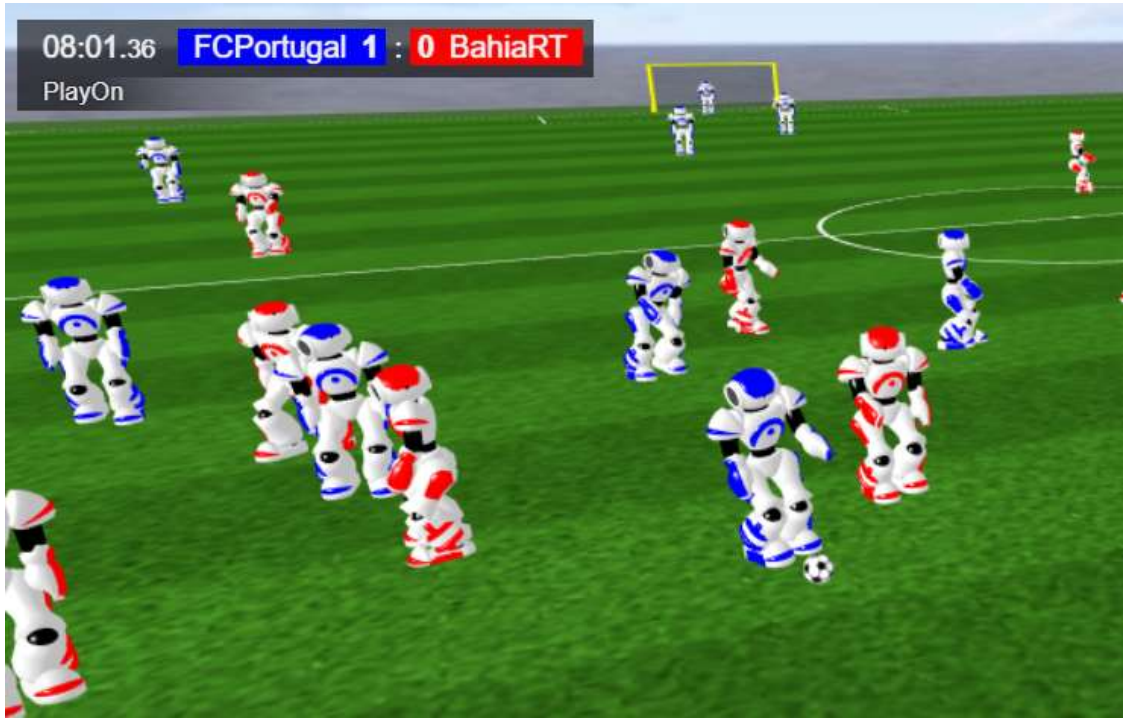
- ✓ Hot-start with imitation
- ✓ Continuous-valued decision making
- ✓ Low number of samples
- ✓ Adaptation

RoboCup Simulation 3D Simulator

Simspark server architecture



Simulation 3D game

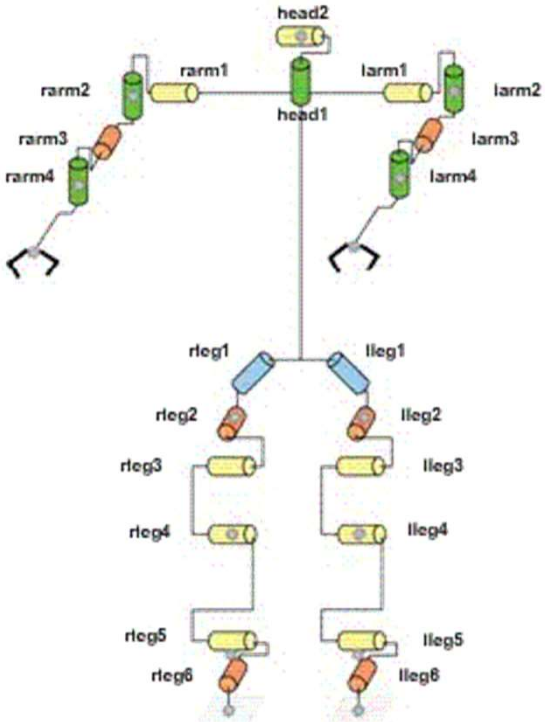


RoboCup 3D Simulated Agent

- **Body**: 22 degrees of freedom (DoFs), 57 cm height and 4.5 kg
- **Perceptors**: gyroscope; force sensors; DOFs; vision; hear
- **Effectors**: beam; DoFs; say

NAO (Aldebaran Robotics)

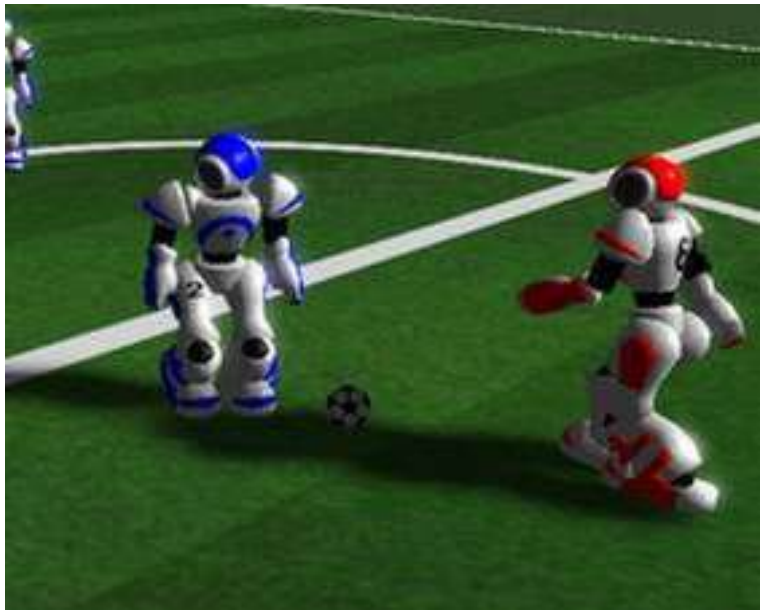
Simulated NAO (3D League)



Body structure

Skill Improvement: Controlled Kick

- **Task**
  - Develop a **kick with controlled kicking distance**
  - From 10 different positions in the soccer field (with distances ranging from 3m to 12m), kick the ball so that it stops in the center of the field
- **Classical approach**
  - Optimize for each distance
- **Contextual approach**
  - Optimize for all distances in a single process
  - Use all data to improve performance
  - Generalize for unknown contexts



Skill Improvement: Controlled Kick

Abbas Abdolmaleki et al. Learning a Humanoid Kick With Controlled Distance. RoboCup 2016: Robot World Cup XX, Springer, July 2016

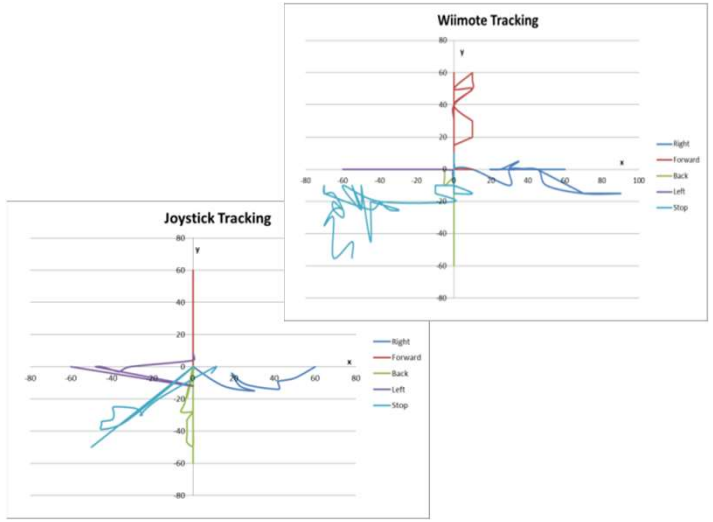
User Profiles and Adapted Interfaces

- **Users** of Intelligent Wheelchairs **have very different skills**
- **Command interface/language** provided for each user should be adapted to his/her capabilities

- **Task**
  - **Generate command language** adapted to each user for driving the IW
- **Approach**
  - **Optimize command language**
  - **User Profiling** provides relevant information

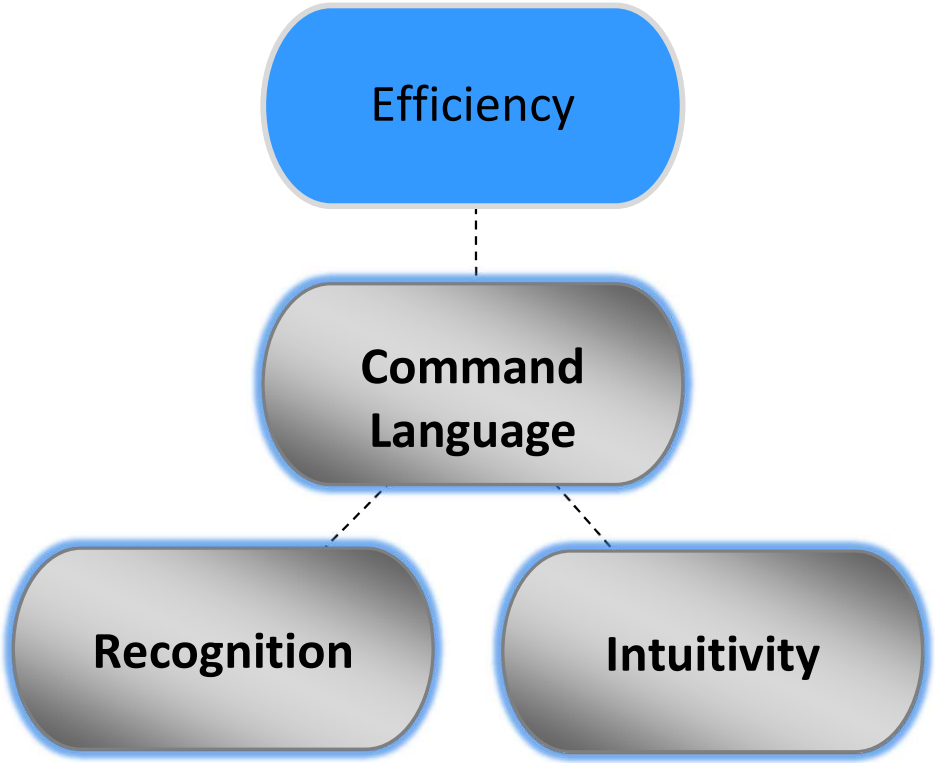
User Profiles and Adapted Interfaces

- **User Profiling Experiments**
  - 11 cerebral palsy users
  - Level IV (27.3%) and V (72.7%) GMFM
  - Voice Inputs
    - “Go”, “Front”, “Forward”, “Back”, “Right”, “Left”, “Turn”, “Spin” and “Stop”
  - Joystick and the Head Movements



User Profiles and Adapted Interfaces

- Command Language

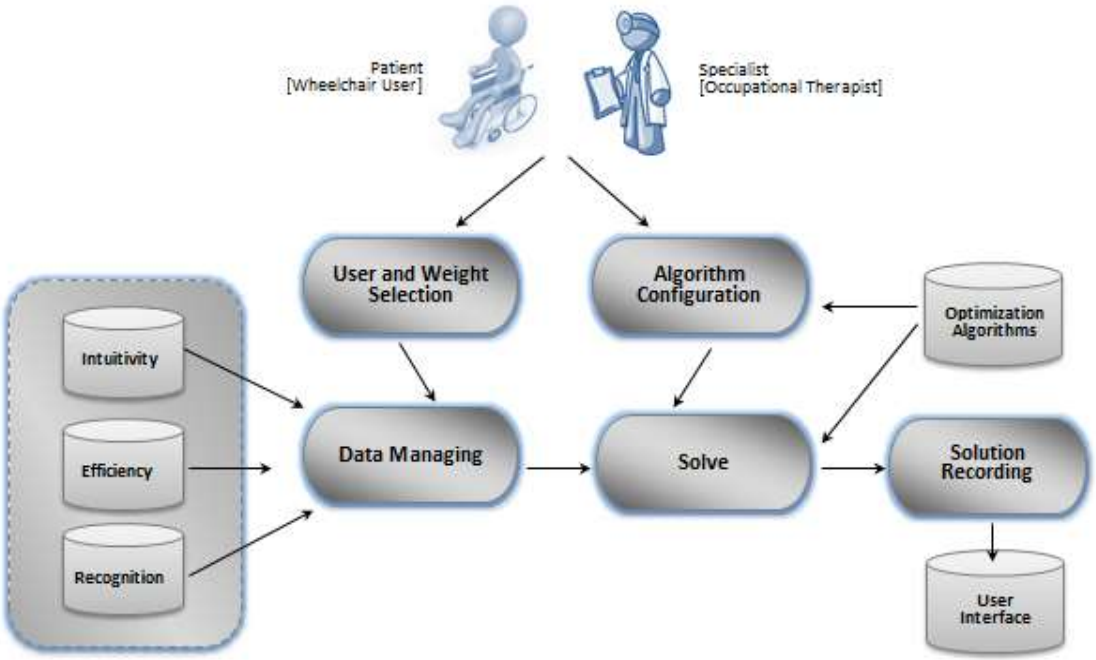


User Profiles and Adapted Interfaces

- Optimized Command Language

Maximizes the function composed by the total time efficiency, recognition and intuitiveness

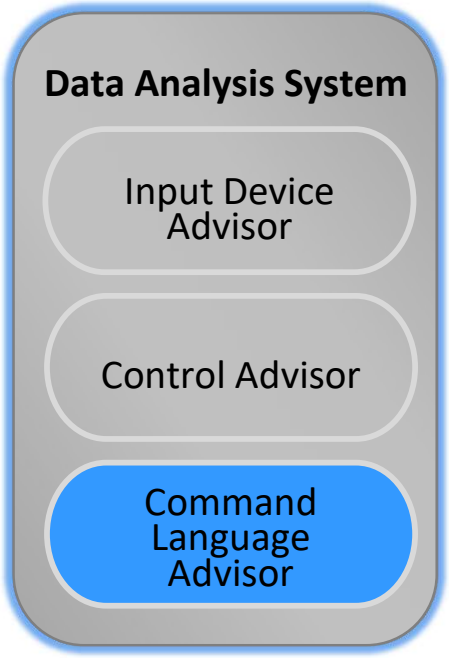
$$\arg\max_{T_{eff},T_{reg},T_{\text{int}}} (\alpha T_{eff} + \beta T_{reg} + \gamma T_{\text{int}})$$




```
(w_rec, w_time, w_intu) = weights; evaluation ← 0
for ncom = 1 to NC do
   recVal ← 1; timeVal ← 0;  intuVal ← 1
   for nseq = 1 to NS do
      inpDev ← inputDevice(solution[ncom][nseq])
      inp ← input(newSolution[ncom][nseq])
      if inpDev = NULL then break
      else
         recVal ← recVal * rec[inpDev][inp]
         timeVal ← timeVal + time[inpDev][inp]
         intuVal ← intuVal * intu[ncom][inpDev][inp]
      endif
   endfor
   evalComm ← w_rec* recVal + w_time*1/(timeVal+1)
                    + w_intu*intuVal
   evaluation ← evaluation  +  evalComm
endfor
return evaluation
```

User Profiles and Adapted Interfaces

- Command Language Advisor



Mean of DAS evaluation higher than mean of evaluation of the command language recommended by specialist (p value = 0.002)

Patient	Evaluation	Command Language for Patients: Forward	Left	Right	Back	Stop
P1						
Specialist	4.53	wiimote	joystick	joystick	joystick	joystick
IDAS	4.57	joystick	joystick	joystick	joystick	joystick
P2						
Specialist	4.18	joystick	joystick	joystick	joystick	voice (“stop”)
IDAS	4.85	joystick	joystick	joystick	joystick	voice (“go”)
P3						
Specialist	3.33	voice (“forward”)	wiimote	wiimote	joystick	voice (“stop”)
IDAS	4.51	wiimote	wiimote	wiimote	wiimote	voice (“go”)
P4						
Specialist	4.50	voice (“forward”)	joystick	joystick	joystick	voice (“stop”)
IDAS	4.60	joystick	joystick	joystick	joystick	voice (“stop”)
P5						
Specialist	4.14	voice (“front”)	wiimote	wiimote	joystick	voice (“stop”)
IDAS	4.40	wiimote	wiimote	voice (“turn”)	joystick	voice (“stop”)
P6						
Specialist	4.13	wiimote	joystick	joystick	joystick	joystick
IDAS	4.38	wiimote	wiimote	wiimote	wiimote	wiimote
P7						
Specialist	4.49	voice (“front”)	joystick	joystick	joystick	voice (“stop”)
IDAS	4.60	joystick	joystick	joystick	voice (“back”)	voice (“stop”)
P8						
Specialist	3.51	wiimote	joystick	joystick	joystick	joystick
IDAS	4.20	wiimote	wiimote	wiimote	wiimote	wiimote
P9						
Specialist	3.70	voice (“forward”)	wiimote	wiimote	joystick	voice (“stop”)
IDAS	4.75	joystick	joystick	joystick	joystick	joystick
P10						
Specialist	4.11	voice (“forward”)	voice (“left”)	voice (“right”)	voice (“turn”)	voice (“stop”)
IDAS	4.80	joystick	joystick	voice (“turn”)	joystick	voice (“go”)
P11						
Specialist	4.29	joystick	wiimote	wiimote	joystick	joystick
IDAS	4.30	wiimote	wiimote	wiimote	wiimote	wiimote

Brígida Mónica Faria, el al. A Methodology for Creating an Adapted Command Language for Driving an Intelligent Wheelchair. Journal of Intelligent & Robotic Systems, vol. 80, no. 3, December 2015

Multiagent Communication Learning

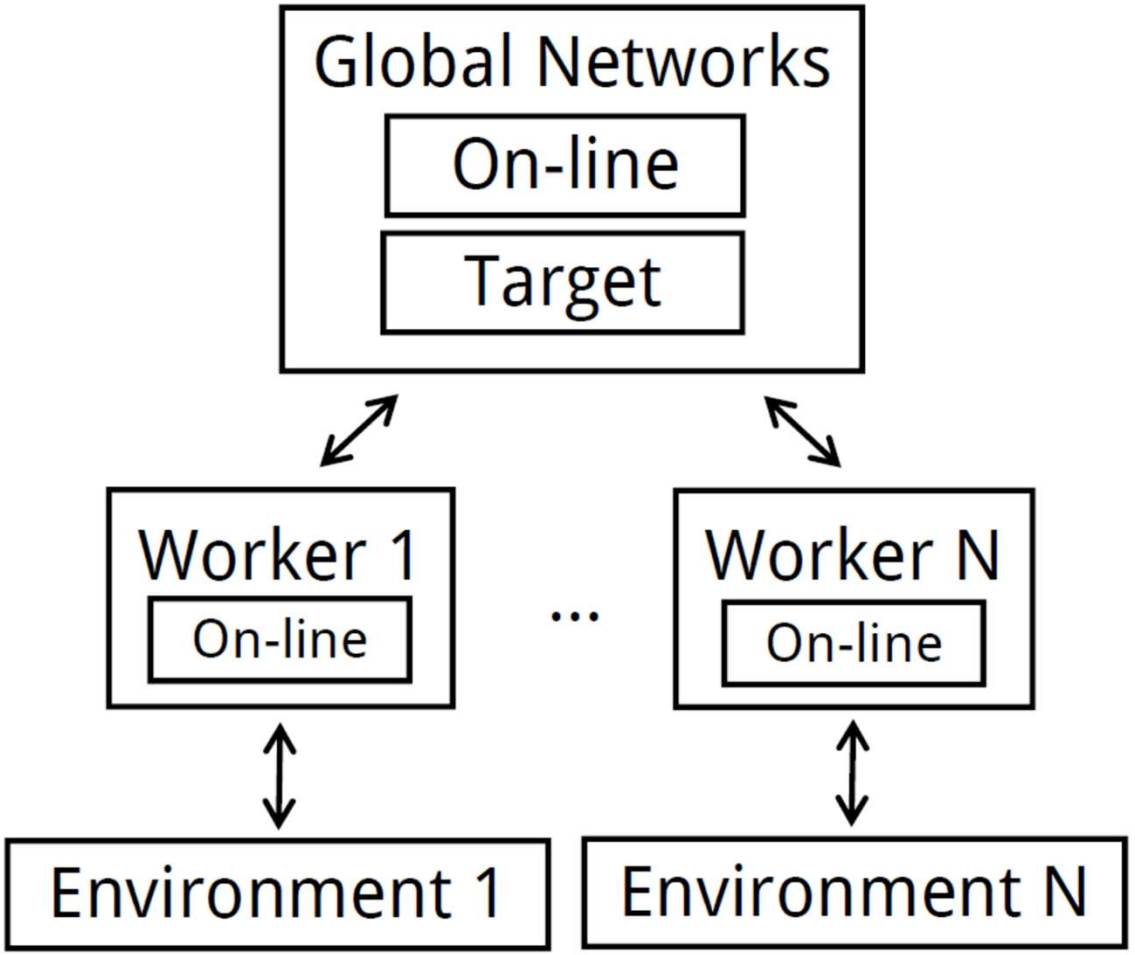
- Multiagent reward based learning challenges
  - Non static environment
  - Complexity exponential to number of agents
  - Learn how to use Communication

- **Task**
  - Learning Coordination and learning Communication from scratch

- **Proposal**
  - Asynchronous Advantage Actor Centralized-Critic with Communication (A3C3)

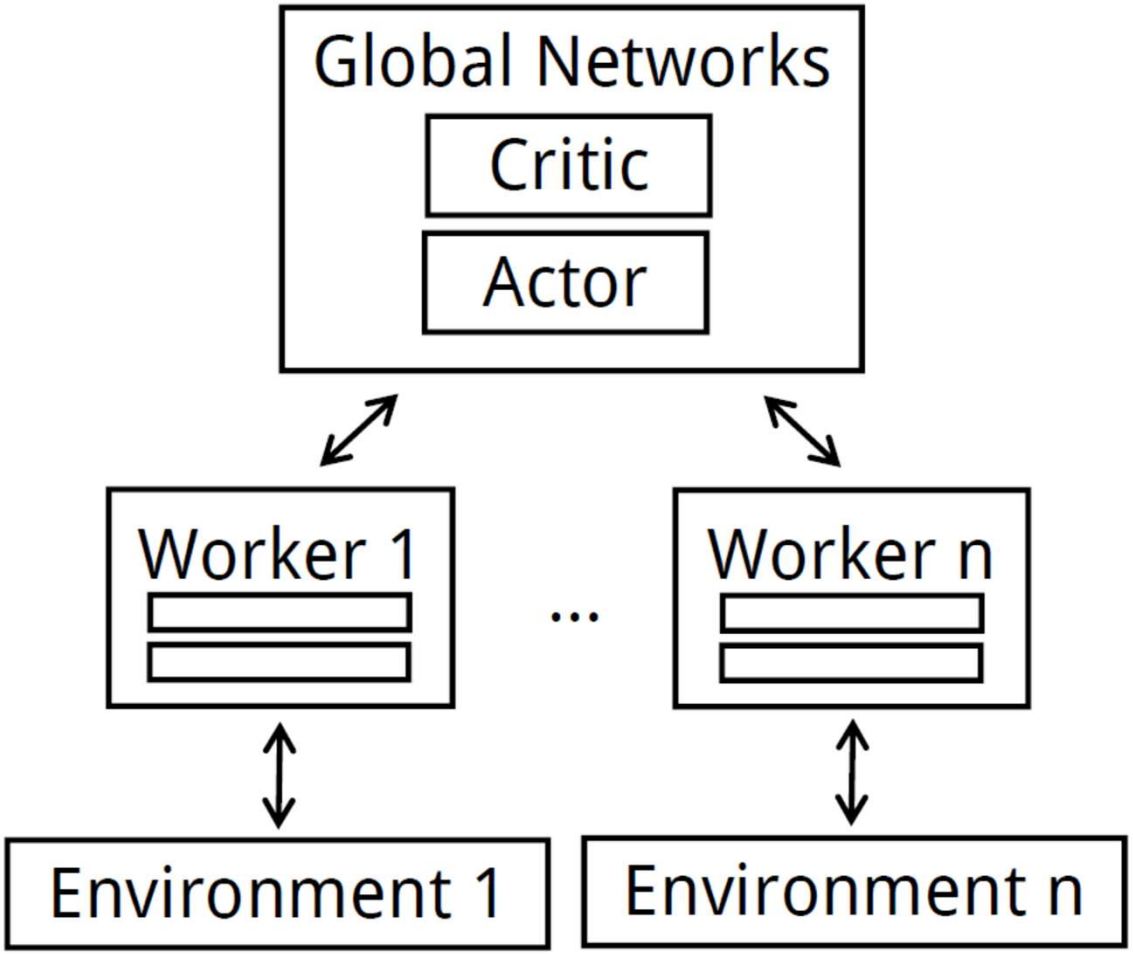
Multiagent Communication Learning

- Asynchronous n-step Q-Learning



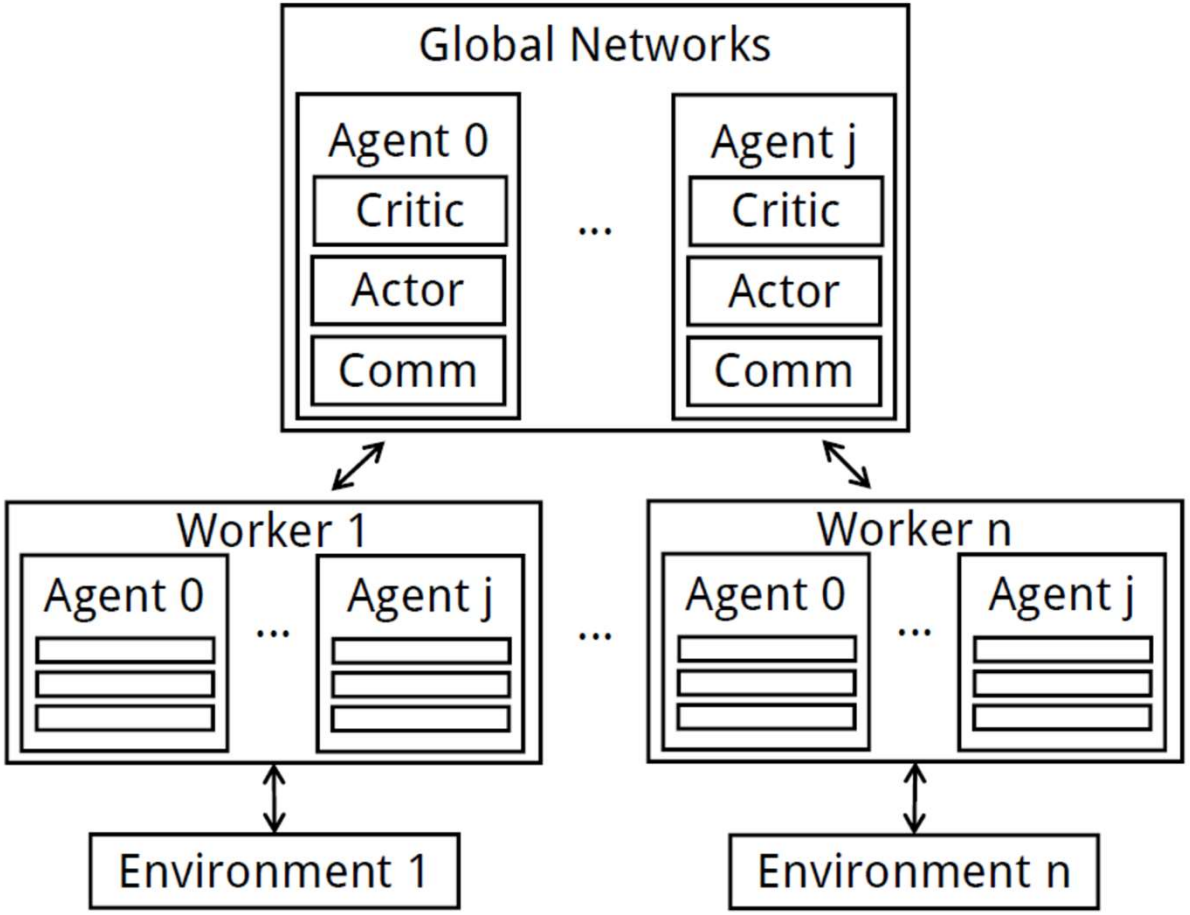
Multiagent Communication Learning

- A3C – Asynchronous Advantage Actor Critic



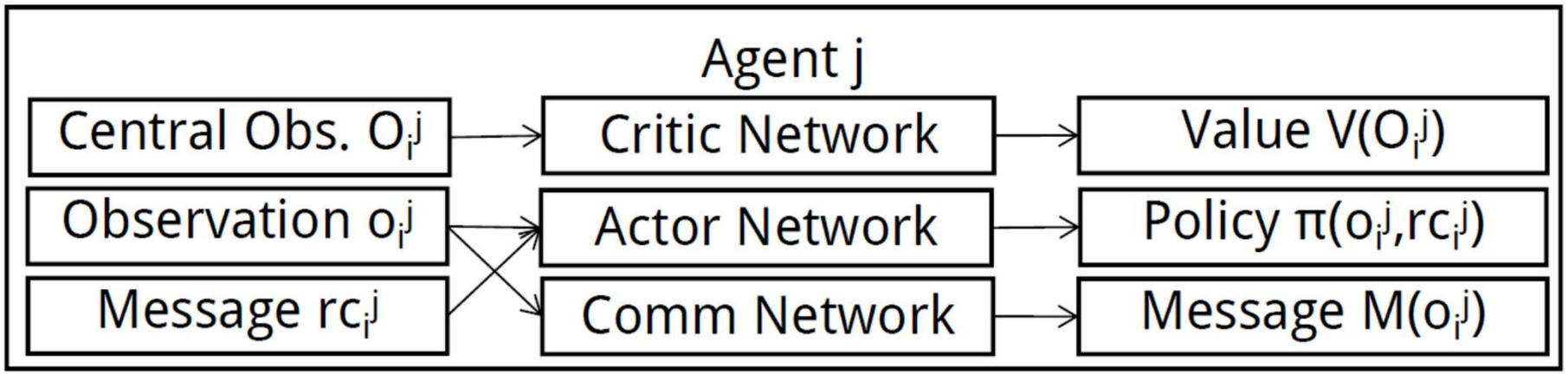
Multiagent Communication Learning

- **Proposal:** A3C3 – Asynchronous Advantage Actor Centralized-Critic with Communication



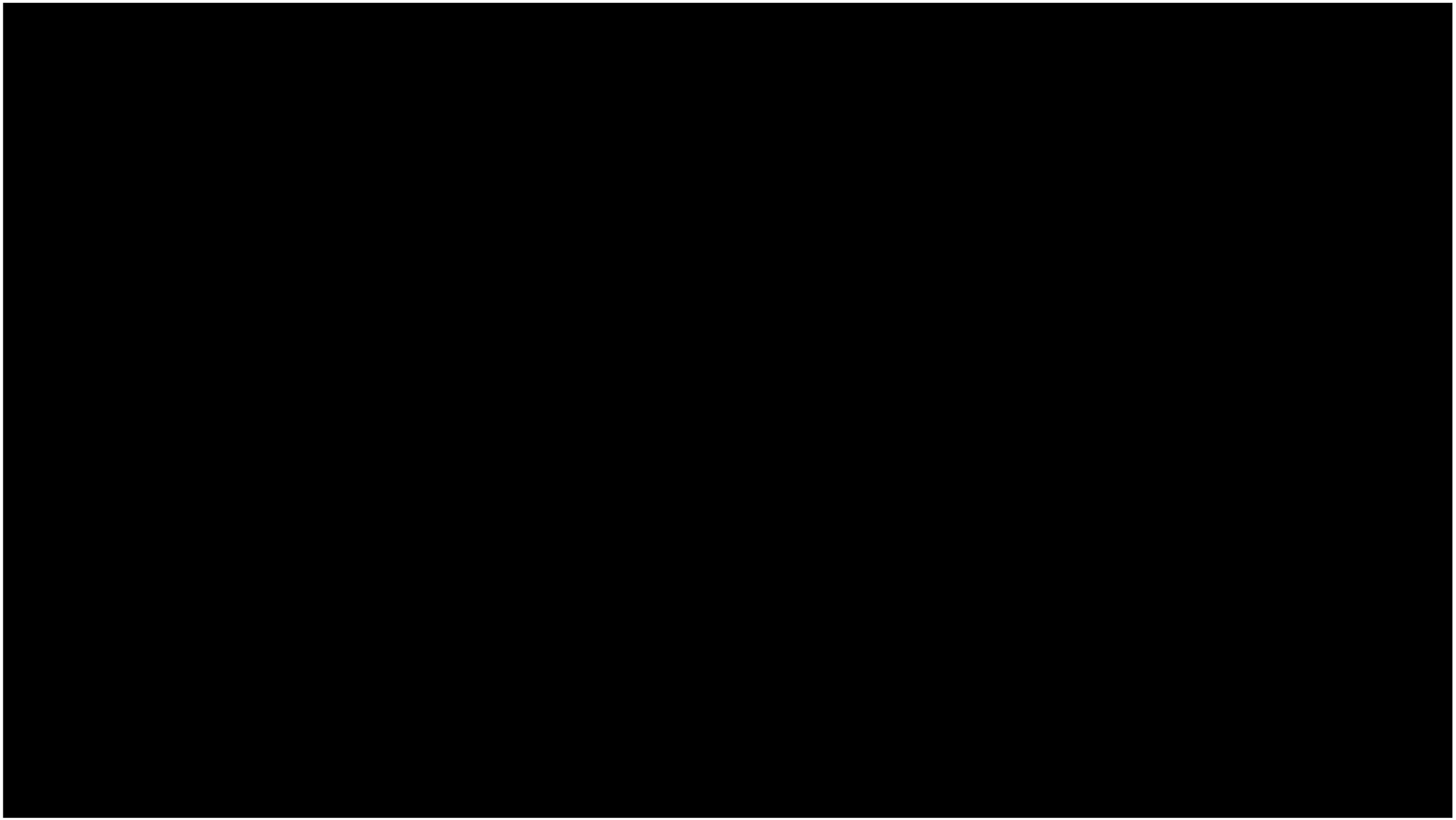
Multiagent Communication Learning

- **Proposal:** A3C3 – Asynchronous Advantage Actor Centralized-Critic with Communication



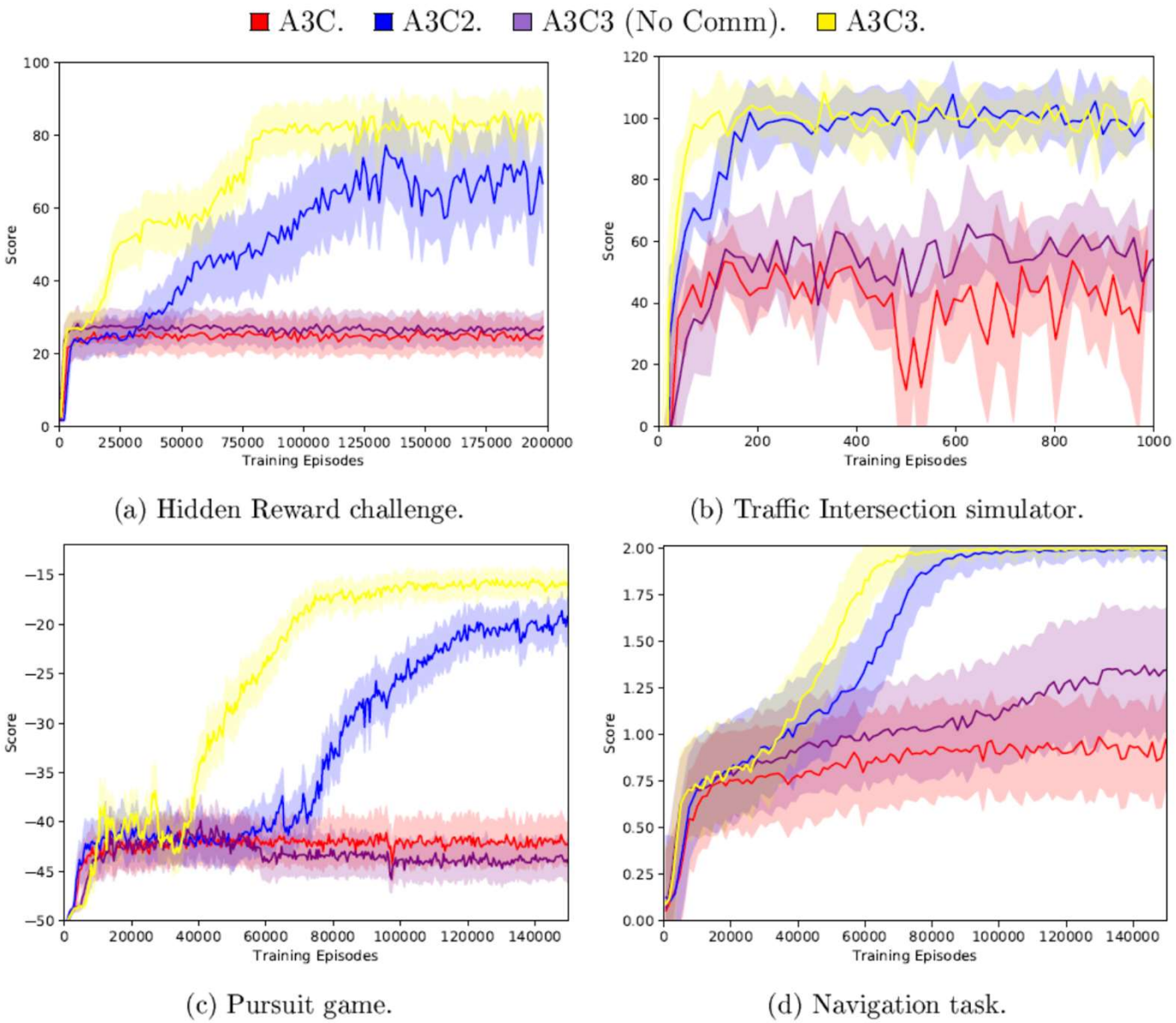
  - Centralized learning phase but distributed execution
  - Centralized Critic fed with complete state or aggregate of local observations
  - Communication at each time step learned *tabula rasa*

Multiagent Communication Learning



David Simões et al. Multi-agent deep reinforcement learning with emergent communication. In 2019 International Joint Conference on Neural Networks (IJCNN). IEEE, 2019.

Multiagent Communication Learning

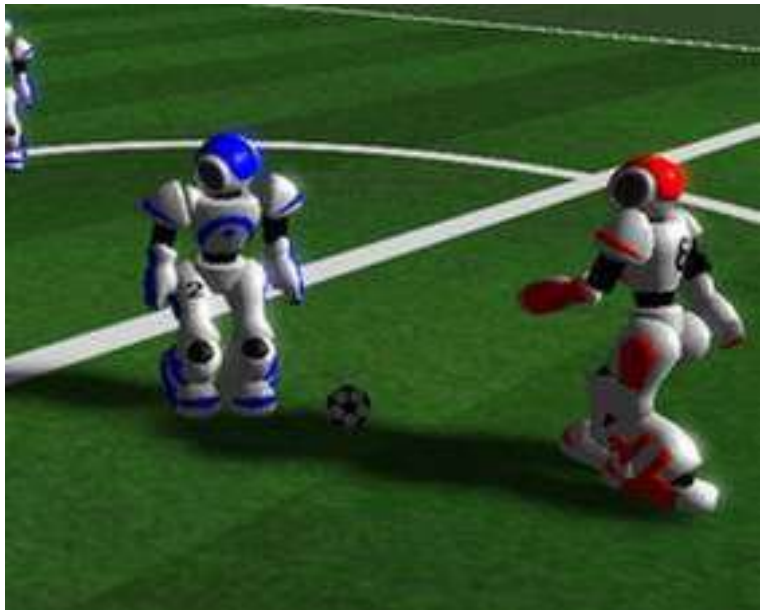


Multiagent Coordination

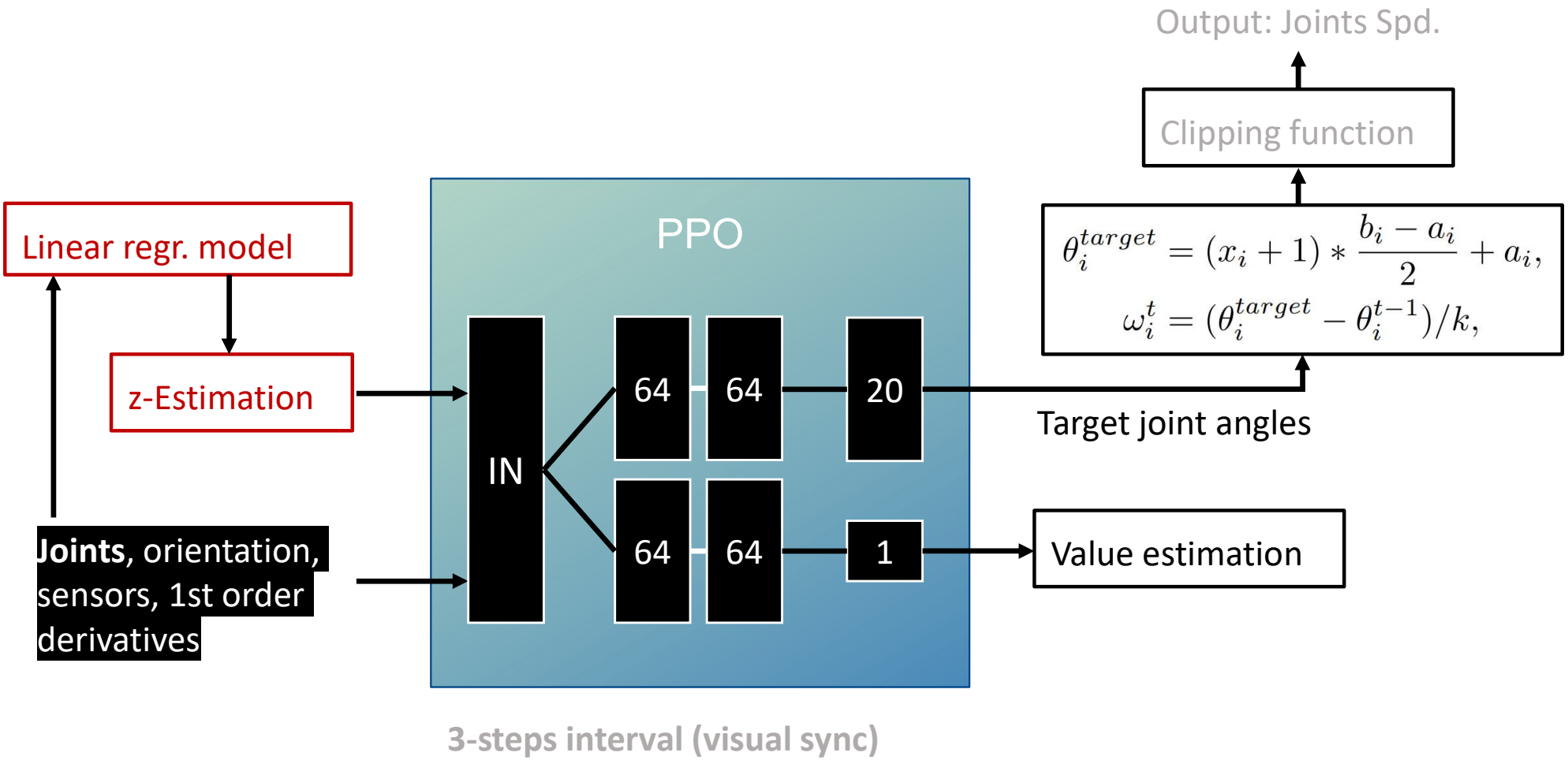
- Formation specification

Behavior Development

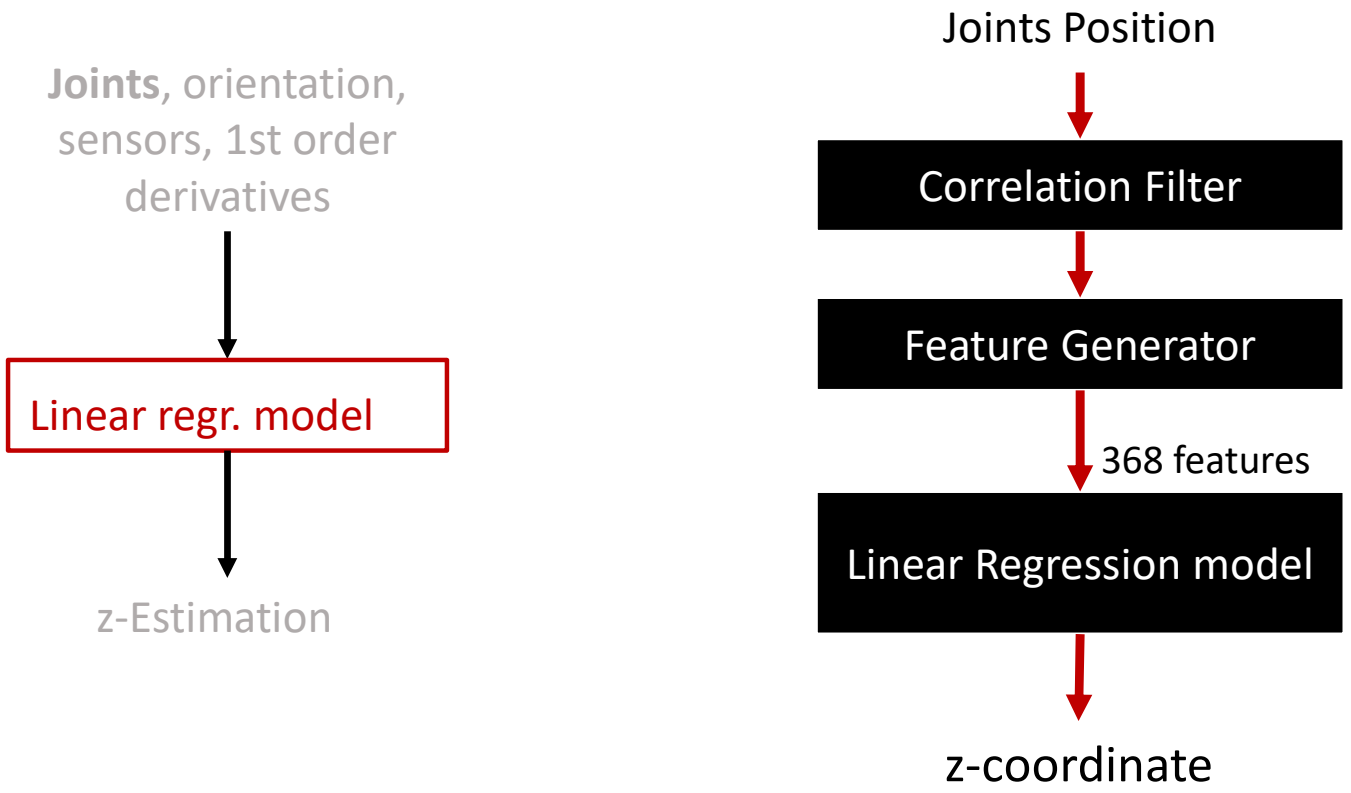
- **Task**
  - Generate fast walking engine for simulated humanoid robot
- **Classical Approach**
  - Model based Walk Engine (ZMP)
  - Optimize model parameters
- **Our Approach**
  - Deep RL: **PPO algorithm**
  - Feedback Walk Engine
  - Use large set of features
  - Initial training using ground-truth data
  - Estimate ground-truth using regression filter
  - Train with agent perceptions

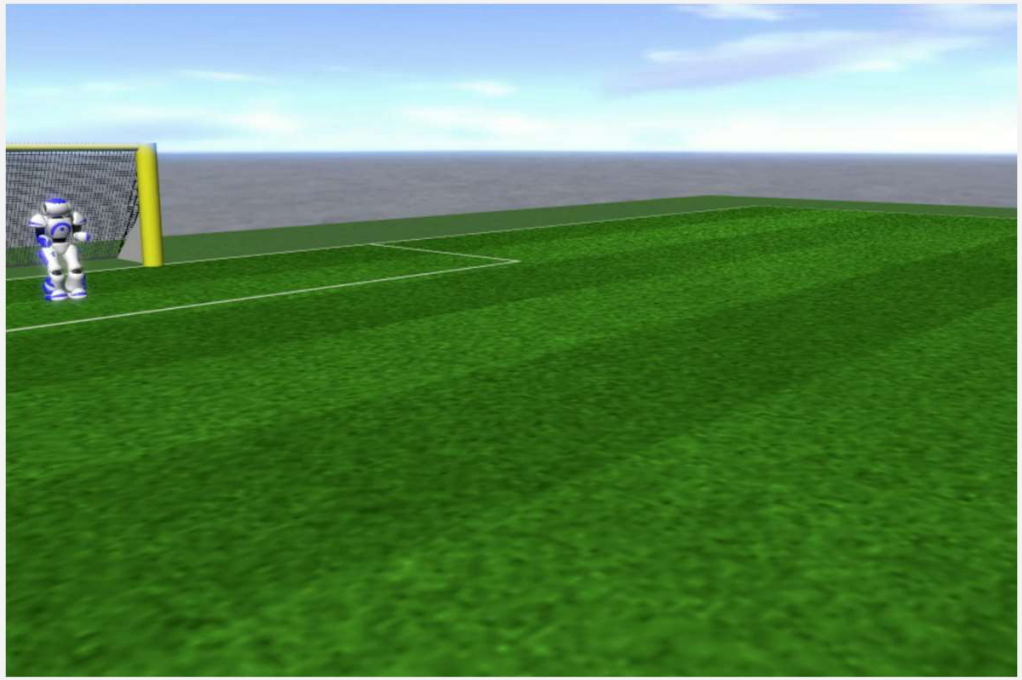


Behavior Development

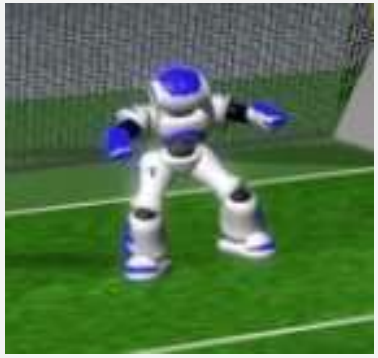


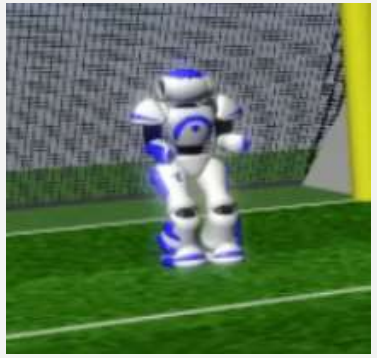
Estimation of Z-true value





sprint v1





sprint v2

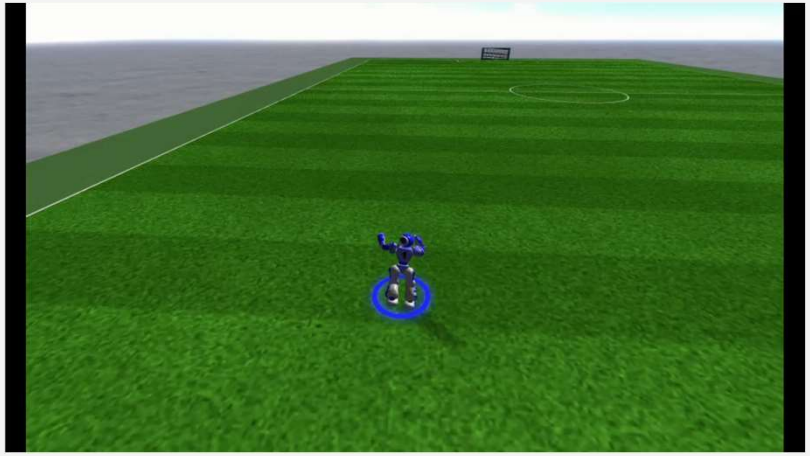
38

Abreu, M. et al., Learning low level skills from scratch for humanoid robot soccer using deep reinforcement learning, 19th IEEE International Conference on Autonomous Robot Systems and Competitions, ICARSC, 2019

Turn and sprint

Dribble

Run and kick



Agile Run

39

Abreu, M. et al., Learning low level skills from scratch for humanoid robot soccer using deep reinforcement learning, 19th IEEE International Conference on Autonomous Robot Systems and Competitions, ICARSC, 2019

Behavior Development

Abreu, M. et al., Learning low level skills from scratch for humanoid robot soccer using deep reinforcement learning, 19th IEEE International Conference on Autonomous Robot Systems and Competitions, ICARSC, 2019

Conclusions

- **Broad range of learning techniques applied to different areas of Robotics:**
  - **Perception**
  - **Behavior development (Value based, Contextual policy search, Deep Learning)**
  - **Adapting Human-Robot Interfaces**
  - **Coordination of Robot teams**
- **Learning can be applied to Robotics, but:**
  - **Data should be used as efficiently as possible**
  - **Take advantage of data structure**
  - **Combine different approaches, if needed**
  - **Use simulation/ground-truth in the first learning steps**

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Robotics / Intelligent Robotics
Robot Learning

Luís Paulo Reis, Nuno Lau, David Simões, Armando Sousa

lpreis@fe.up.pt

Director of LIACC – Artificial Intelligence and Computer Science Lab.

Associate Professor at DEI/FEUP – Informatics Engineering Department, Faculty of Engineering of the University of Porto, Portugal

President of APPIA – Portuguese Association for Artificial Intelligence