

# Benchmarking Intelligent Service Robots through Scientific Competitions: the RoboCup@Home approach



Luca locchi

Sapienza University of Rome, Italy

---

# Motivation

- Benchmarking Domestic Service Robots
  - Complex Integrated Systems
  - Human-Robot Interaction
  - Large variety of tasks
  - Evaluating integrated AI



# About RoboCup@Home

- Starts in 2006
  - 8 international competitions
  - Many regional competitions
- Largest competition for domestic and service robots



# Large variety of tasks



# Benchmarking Domestic Service Robots

## Functional benchmarking

- Usually based on data set collection and off-line processing

## Difficulties in benchmarking DSR

- Human involved
- Real environments
- Integration of several capabilities coming from different research fields
- Large variety of tasks



# Robotic scientific competitions

- DARPA Challenges
- RoboCup Soccer, Rescue, @Home, @Work
- AAI / ICRA / IROS robot competitions
- RoboCup Junior, Eurobot
- RoCKIn

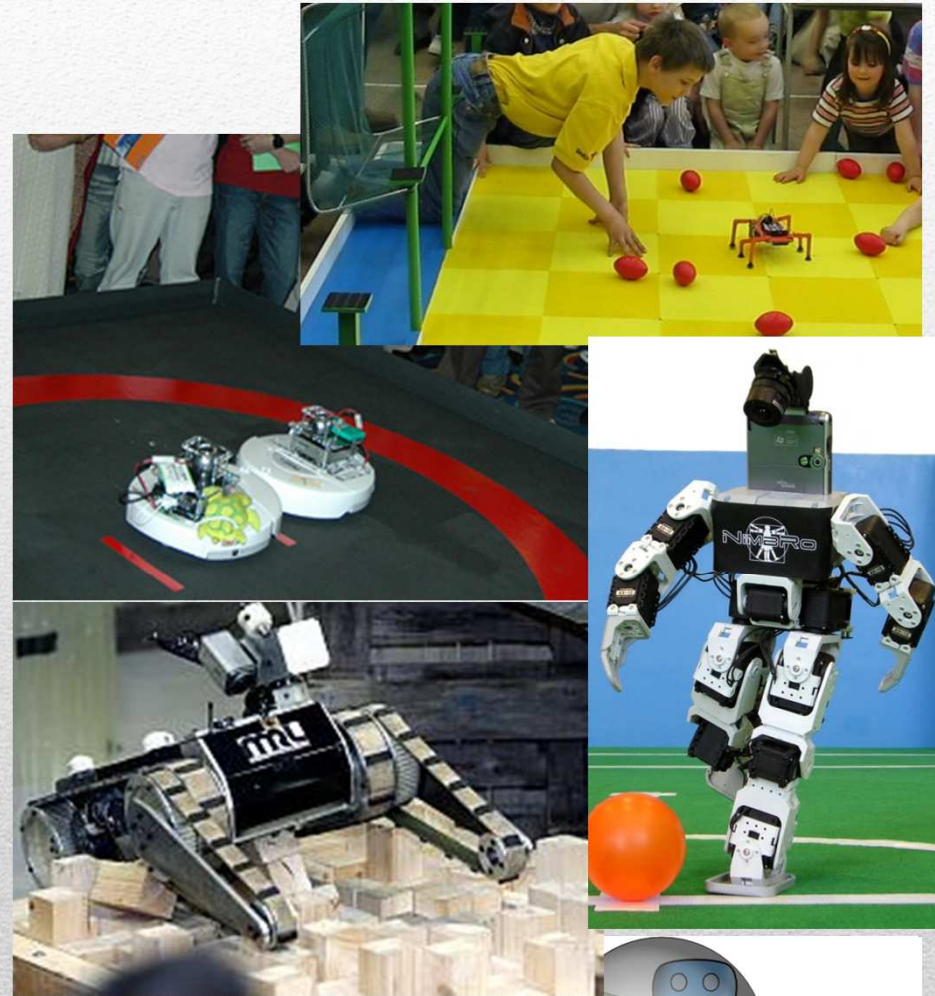
## Advantages of Competitions

- Set up of common test-beds
- Attractive for many teams (research groups)
- Collaboration and knowledge sharing
- Evolution over time



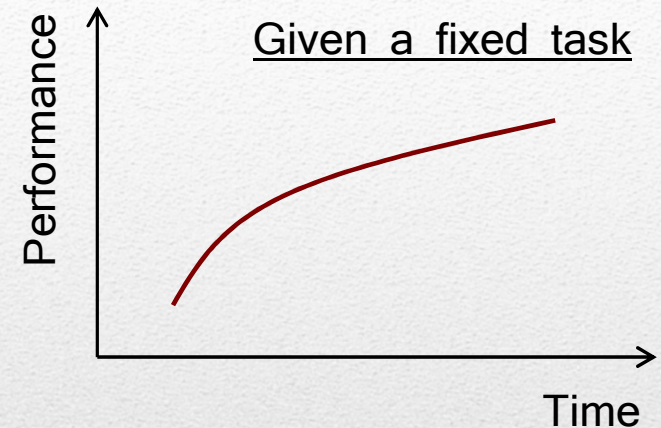
# Observations from other Robot Competitions

- Little HRI involved
- Limited application orientation
- No real world environment
- Very specific rules and regulations for robots and environment
- Often requires many resources (special environment, many robots)
- **Danger of developing towards local optima**

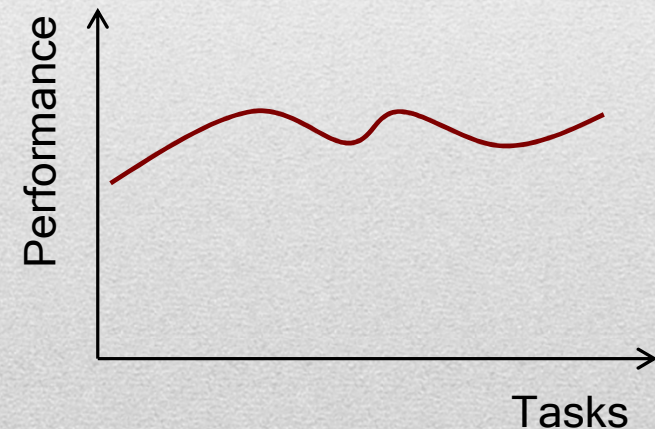


# Local Optima in Benchmarking

Fixed task +  
improving performance over time  
**Local optimum (overfitting)**



Set of changing tasks +  
maintaining performance over time  
**Global optimum**





# RoboCup@Home

(Difficulties in) Benchmarking DSR +  
(Advantages of) Benchmarking through competitions =

---

## RoboCup@Home

RoboCup@Home competition allows for testing DSRs in **many integrated tasks** (*not single functionalities*) in real or realistic environments with the interaction of **external users** (*not developers of the system under test*).



# RoboCup@Home approach

Integrated system benchmarking of DSR:

- Realistic/real environments
- Definition of many tests related to desired functionalities and evaluated by external users
- Changing tests over the years to keep performance "constantly good"
- Statistical evaluation for measuring league progresses

Can this approach be applied also to evaluate the development of a single "medium-term" project?



# RoboCup@Home Scenario and Concepts

- Autonomous robots
- Human-Robot Interaction
- Non-standardised realistic domestic environment and real public areas
- Many tests related to desired functionalities
- Changing tests over the years to keep performance "constantly good"
- Statistical evaluation for measuring league progresses

# Current focus of RoboCup@Home

## Functional abilities:

- Navigation
- Mapping
- Person recognition
- Person tracking
- Object recognition
- Object manipulation
- Speech recognition
- Gesture recognition
- Cognition



# Current focus of RoboCup@Home

## System properties:

- Ease of use
- Fast calibration and setup
- Natural and multi-modal interaction
- Attractiveness and ergonomics of the robot
- Adaptivity and general intelligence
- Robustness
- General applicability



# Implementation of RoboCup@Home

## General rules

- 2 stages with different focus
  - Stage 1 for basic tasks
  - Stage 2 for more complex, integrated tasks
- High level of uncertainty in the environment (no standardization)
- Only natural interaction allowed
- Very short setup time (usually 1 minute)
- Partial score system for tests



# Stage 1

- **Robot Inspection & Poster:** Autonomous registration to the competition, TC inspection, team poster
- **Follow me:** Lead the robot quickly on a path through an external scenario
- **Cocktail Party:** Deliver drinks to people in the apartment
- **Clean up:** Clean up a room in the apartment
- **Emergency Situation:** React to an unknown emergency situation
- **Technical Challenge:** Furniture-type Object perception
- **Open Challenge:** Present and demonstrate most important (scientific) achievements



## Stage 2

- **Enduring General Purpose Service Robot:** Solve multiple tasks not known beforehand upon request
- **Restaurant:** Mapping and serving drinks and food in a real unknown restaurant
- **Demo Challenge:** Demonstration of health care abilities (e.g., elder, children)
- **Finals:** Open demonstration with external jury evaluation + Exec evaluation





# Implementation of RoboCup@Home

|  | Navigation                         | Mapping | Person<br>Recogn. | Person<br>Tracking | Object<br>Recogn. | Object<br>Manipul. | Speech/<br>Gesture<br>Recogn. | Cognition |
|--|------------------------------------|---------|-------------------|--------------------|-------------------|--------------------|-------------------------------|-----------|
| <b>Follow Me</b>                                 | 49 %                               |         | 6 %               | 39 %               |                   |                    | 6 %                           |           |
| <b>Clean Up</b>                                  |                                    |         |                   |                    |                   |                    |                               |           |
| <b>Cocktail<br/>Party</b>                        |                                    |         |                   |                    |                   |                    |                               |           |
| <b>Emergency<br/>Situation</b>                   | Defined by the Technical Committee |         |                   |                    |                   |                    |                               |           |
| <b>General<br/>Purpose<br/>Service<br/>Robot</b> |                                    |         |                   |                    |                   |                    |                               |           |
| <b>Restaurant</b>                                |                                    |         |                   |                    |                   |                    |                               |           |
| <b>Open<br/>Challenge</b>                        |                                    |         |                   |                    |                   |                    |                               |           |
| <b>Demo<br/>Challenge</b>                        | Defined by the teams               |         |                   |                    |                   |                    |                               |           |
| <b>Final</b>                                     |                                    |         |                   |                    |                   |                    |                               |           |

# Test evolution: 'follow me' example

**2007:** proof of concept, special markers on the walker allowed

**2008:** walker known, but no special markers

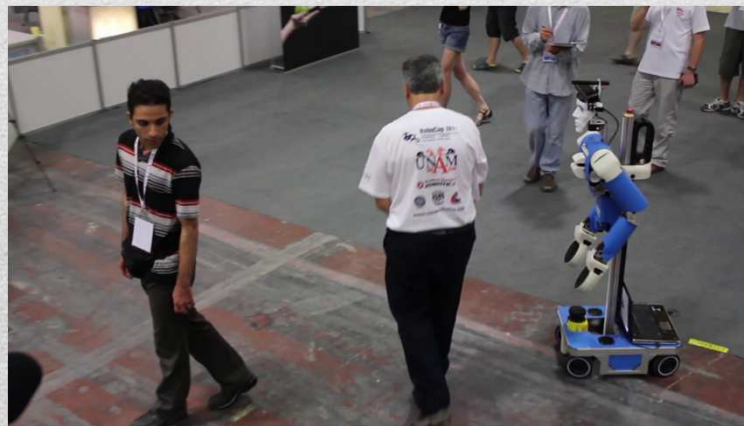
**2009:** walker unknown

**2010:** outside the arena (in the RoboCup venue)

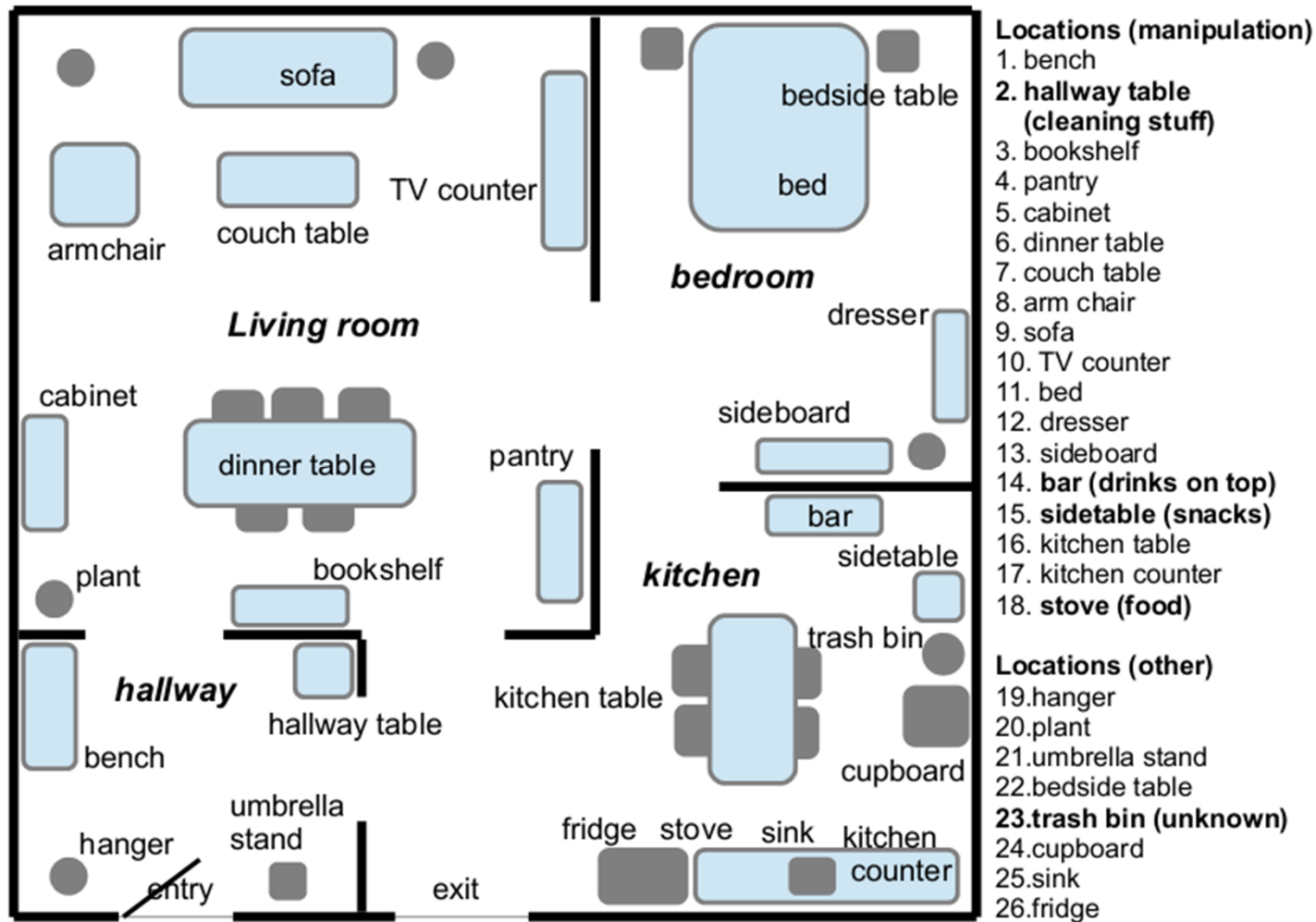
**2011:** pre-defined interferences (people passing between walker and robot)

**2012-2013:** crowded and complex environment  
(changing floor through an elevator)

**future:** public environment with crowd and unpredictable interferences



# Apartment, People and Objects



# Person names

| Standard Names |           |
|----------------|-----------|
| Michael        | Jessica   |
| Christopher    | Ashley    |
| Matthew        | Brittany  |
| Joshua         | Amanda    |
| Daniel         | Samantha  |
| David          | Sarah     |
| Andrew         | Stephanie |
| James          | Jennifer  |
| Justin         | Elizabeth |
| Joseph         | Lauren    |

# Objects

|   |   |                                      |    |   |                                   |    |  |  |    |   |                                     |    |   |                                      |
|---|---|--------------------------------------|----|---|-----------------------------------|----|--|--|----|---|-------------------------------------|----|---|--------------------------------------|
| 1 |    | Name: Deodorant<br>Category: Cleanin | 6  |    | Name: Beer bot<br>Category: Drink | 11 |    | Name: Chocolate m<br>Category: Drinks  | 16 |    | Name: Tomato Sau<br>Category: Food  | 21 |    | Name: Garlic sauce<br>Category: Food |
| 2 |    | Name: tooth pas<br>Category: Cleanin | 7  |    | Name: Fanta<br>Category: Drink    | 12 |    | Name: Energy drink<br>Category: Drinks | 17 |    | Name: Peanut butt<br>Category: Food | 22 |    | Name: Chocolate<br>Category: Snacks  |
| 3 |    | Name: Cleaner<br>Category: Cleanin   | 8  |    | Name: Beer can<br>Category: Drink | 13 |    | Name: Orange juice<br>Category: Drinks | 18 |    | Name: Chicken noc<br>Category: Food | 23 |    | Name: Cookies<br>Category: Snacks    |
| 4 |   | Name: Fresh dis<br>Category: Cleanin | 9  |   | Name: Coke<br>Category: Drink     | 14 |   | Name: Milk<br>Category: Drinks         | 19 |   | Name: Marmalade<br>Category: Food   | 24 |   | Name: Drops<br>Category: Snacks      |
| 5 |  | Name: Sponge<br>Category: Cleanin    | 10 |  | Name: Seven u<br>Category: Drink  | 15 |  | Name: Apple juice<br>Category: Drinks  | 20 |  | Name: Veggie Nood<br>Category: Food | 25 |  | Name: Crackers<br>Category: Snacks   |

# Object categories and default locations

| Category       | Locations     |
|----------------|---------------|
| Drinks         | Bar           |
| Cleaning stuff | Hallway table |
| Food           | Stove         |
| Snacks         | Side table    |
| Unknown        | Trash bin     |

# Benchmarking Robot Cognition: General Purpose Service Robot

- The test is about how much the robot can understand and reason about the environment and its task
- **No predefined task**
- Task goals are randomly generated at runtime
- Task goals can include multiple objects/locations, underspecified objects/locations and wrong information
- GPSR incorporates the abilities tested in all previous tests.



# Benchmarking Robot Cognition: General Purpose Service Robot

**Task goal is not predefined !**

Given a set of known objects, known locations and known persons,  
execute a **randomly generated task** from a set of templates.



## Category 1

Move to ⟨LOCATION1⟩, get ⟨OBJECT⟩ and put it at ⟨LOCATION2⟩  
Go to ⟨LOCATION1⟩, find ⟨PERSON⟩ and follow him/her  
...

## Category 2

Bring me a drink  
Look for a person in the apartment  
...

## Category 3

Find an ⟨OBJECT⟩ in ⟨LOCATION⟩  
(but there is not such an object in that location)  
...



# Evaluation of the League

- Year by year statistical analysis to:
  - Measure overall performance
  - Drive developments
  - Plan for rule changes



# Score system

Each test includes a set of the functional abilities

| Ability             | 2008 | 2009 | 2010  | 2011  | 2012  |
|---------------------|------|------|-------|-------|-------|
| Navigation          | 40%  | 33%  | 21.5% | 20.5% | 17.5% |
| Mapping             | 3%   | 3%   | 8%    | 8%    | 10.5% |
| Person Recognition  | 10%  | 12%  | 12.5% | 13%   | 7%    |
| Person Tracking     | 6%   | 4%   | 3%    | 3.5 % | 6.5%  |
| Object Recognition  | 13%  | 17%  | 8.5%  | 7%    | 9%    |
| Object Manipulation | 18%  | 17%  | 15%   | 16.5% | 17.5% |
| Speech Recognition  | 7%   | 8%   | 15%   | 11.5% | 13.5% |
| Gesture Recognition | 3%   | 6%   | 3.5%  | 4%    | 2.5%  |
| Cognition           | -    | -    | 13%   | 16%   | 16%   |

Distribution of functional abilities over tests evolves over time allowing for proper analysing and planning.

# Score system

Example from Follow me 2012 test

|          |     | Navigation | Object Recognition | Person Recognition | Person Tracking | Object Manipulation | Speech recognition | Gesture recognition | Mapping  | Cognition |             |
|----------|-----|------------|--------------------|--------------------|-----------------|---------------------|--------------------|---------------------|----------|-----------|-------------|
| CP1      | 300 | 0.5        |                    |                    | 0.5             |                     |                    |                     |          |           |             |
| CP2      | 300 | 0.3        |                    | 0.2                | 0.3             |                     | 0.1                | 0.1                 |          |           |             |
| CP3      | 300 | 0.5        |                    |                    | 0.5             |                     |                    |                     |          |           |             |
| Complete | 100 | 1          |                    |                    |                 |                     |                    |                     |          |           |             |
|          |     | <b>490</b> | <b>0</b>           | <b>60</b>          | <b>390</b>      | <b>0</b>            | <b>30</b>          | <b>30</b>           | <b>0</b> | <b>0</b>  | <b>1000</b> |

# Evaluation 2006-2012

| Ability             | 2008 [%]       | 2009 [%]           | 2010 [%]           | 2011 [%]           | 2012 [%]           |
|---------------------|----------------|--------------------|--------------------|--------------------|--------------------|
| Navigation          | 40 / 25        | 47 / 40            | 33 / 20            | 61 / 26            | 52 / 23            |
| Mapping             | 100 / 44       | 100 / 92           | 21 / 10            | 33 / 10            | 10 / 4             |
| Person Recognition  | 32 / 15        | 69 / 37            | 57 / 23            | 48 / 16            | 62 / 15            |
| Person Tracking     | 100 / 81       | 100 / 69           | 100 / 72           | 100 / 76           | 62 / 33            |
| Object Recognition  | 29 / 8         | 39 / 23            | 6 / 1              | 35 / 10            | 56 / 20            |
| Object Manipulation | 3 / 1          | 48 / 23            | 29 / 8             | 49 / 21            | 73 / 27            |
| Speech Recognition  | 87 / 37        | 89 / 71            | 50 / 38            | 76 / 59            | 90 / 56            |
| Gesture Recognition | 0 / 0          | 0 / 0              | 62 / 26            | 100 / 49           | 88 / 37            |
| Cognition           | -              | -                  | 17 / 3             | 68 / 24            | 32 / 8             |
| <b>Average</b>      | <b>41 / 21</b> | <b>61.5 / 44.4</b> | <b>41.6 / 22.4</b> | <b>63.3 / 32.5</b> | <b>58.2 / 24.8</b> |

Best/average score of the finalist teams.

# Evaluation 2006-2012

| Measure                   | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 |
|---------------------------|------|------|------|------|------|------|------|
| Number of teams           | 12   | 11   | 14   | 18   | 24   | 19   | 18   |
| Total amount of tests     | 66   | 76   | 86   | 127  | 164  | 141  | 108  |
| Percentage of succ. tests | 17%  | 36%  | 59%  | 83%  | 74%  | 73%  | 58%  |
| Avg. succ. tests p. team  | 1.0  | 2.5  | 4.9  | 7.3  | 6.3  | 6.5  | 4.2  |

Performance metrics of the RoboCup@Home league over the years

Performance do not always increase because of changes in the rules (major changes in 2008, 2010, 2012).

**Good: we are not going towards a local optimum!!!**



# Your 3-years project on intelligent robots

| 2013   | Navigation | Mapping | Person<br>Recogn. | Person<br>Tracking | Object<br>Recogn. | Object<br>Manipul. | Speech/<br>Gesture<br>Recogn. | Cognition |
|--------|------------|---------|-------------------|--------------------|-------------------|--------------------|-------------------------------|-----------|
| Test 1 |            |         |                   |                    |                   |                    |                               |           |
| 2014   | Navigation | Mapping | Person<br>Recogn. | Person<br>Tracking | Object<br>Recogn. | Object<br>Manipul. | Speech/<br>Gesture<br>Recogn. | Cognition |
| Test 2 |            |         |                   |                    |                   |                    |                               |           |
| Test 3 |            |         |                   |                    |                   |                    |                               |           |
| ...    |            |         |                   |                    |                   |                    |                               |           |
| 2015   | Navigation | Mapping | Person<br>Recogn. | Person<br>Tracking | Object<br>Recogn. | Object<br>Manipul. | Speech/<br>Gesture<br>Recogn. | Cognition |
| Test 1 |            |         |                   |                    |                   |                    |                               |           |
| Test 2 |            |         |                   |                    |                   |                    |                               |           |
| Test 3 |            |         |                   |                    |                   |                    |                               |           |
| ...    |            |         |                   |                    |                   |                    |                               |           |

**"The main outcome of my project is general applicability"**



# RoboCup@Home Community Resources



- Web site (information and rules)
- @Home Wiki (> 50 teams active worldwide) HW/SW/Papers...
- Mailing lists (active rule discussion)

[www.robocupathome.org](http://www.robocupathome.org)



# Scientific Achievements

- Speech understanding in noisy environments
- Speaker localization for following human guides
- Detecting and tracking human operators using laser and RGBD cameras
- Detecting, learning and recognizing objects
- Complex two-hands object manipulation

**Demonstrated within an integrated system**





# Future directions of RoboCup@Home

- More and more tests in the real world
- Improved cognitive and social skills
  - language skills
  - social behaviors
- **Improved safety and security**
- **Human-robot cooperation**
- Inter-team robot-robot cooperation
- Keep improving the adaptive benchmarking



# Conclusions

- **Benchmarking methodology based on the definition of several variable tests**
- RoboCup@Home can drive the development of effective intelligent robots
- Statistical analysis can drive fast achievements of the league.
- Research groups can use RoboCup@Home to develop, test, evaluate and disseminate DSR solutions.



**Thank you for your attention**

**Questions?**

**[www.robocupathome.org](http://www.robocupathome.org)**

