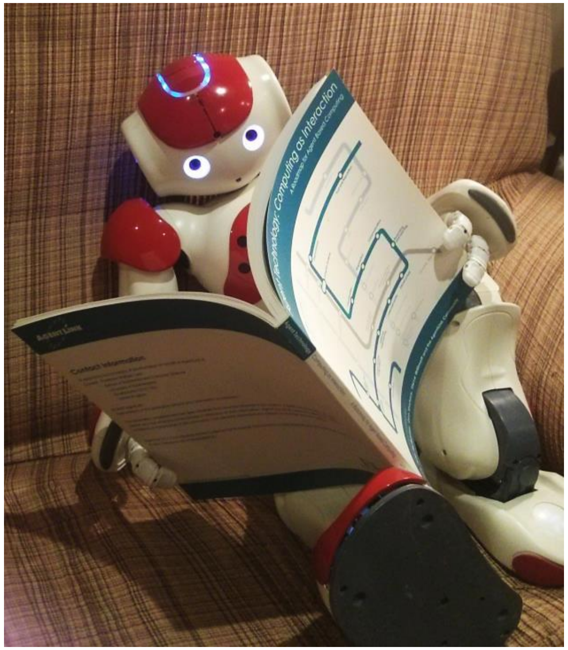
The International Joint Conference on Neural Networks

GESTURE BASED HUMAN MULTI-ROBOT INTERACTION

Gerard Canal, Cecilio Angulo, and Sergio Escalera

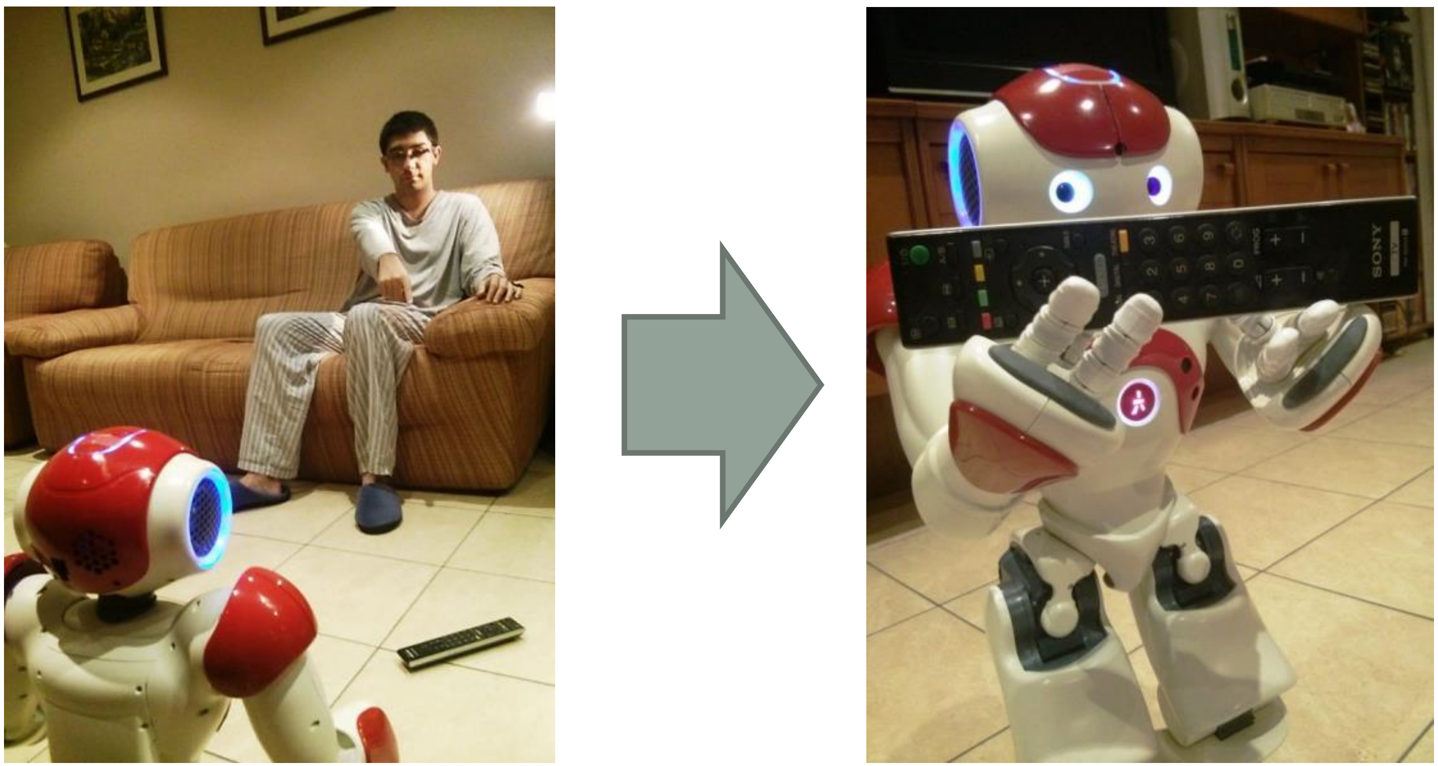
Introduction

- Nowadays robots are able to perform many useful tasks.
- Most of the human communication is non-verbal.
- HRI research on a gesture-based interaction system.



Motivation

- Elderly or handicapped person case.



Outline

- Goals
- Resources
- System overview
- Gesture Recognition
- HRI methods
- Results: Gesture recognition performance
- Results: User evaluation
- Conclusions
- Future work

Goals

- Design of a system *easy* to use and *intuitive*.
- *Real time*, therefore, *fast* response.
  - *Static* and *dynamic* gestures recognition.
  - *Accuracy* in pointing at the location.
  - Allowing the robot to respond in an intuitive manner.
  - Solving *ambiguous* situations.

Goals

- Design of a system *easy* to use and *intuitive*.
- *Real time*, therefore, *fast* response.
  - *Static* and *dynamic* gestures recognition.
  - *Accuracy* in pointing at the location.
  - **Allowing the robot to respond in an intuitive manner.**
  - Solving *ambiguous* situations.

Goals – System set up

Allowing the robot to respond in an intuitive manner.

- Vision sensor too large to be carried by the robot.
- DARPA Grand Challenge idea of a driving humanoid.

Hardware resources

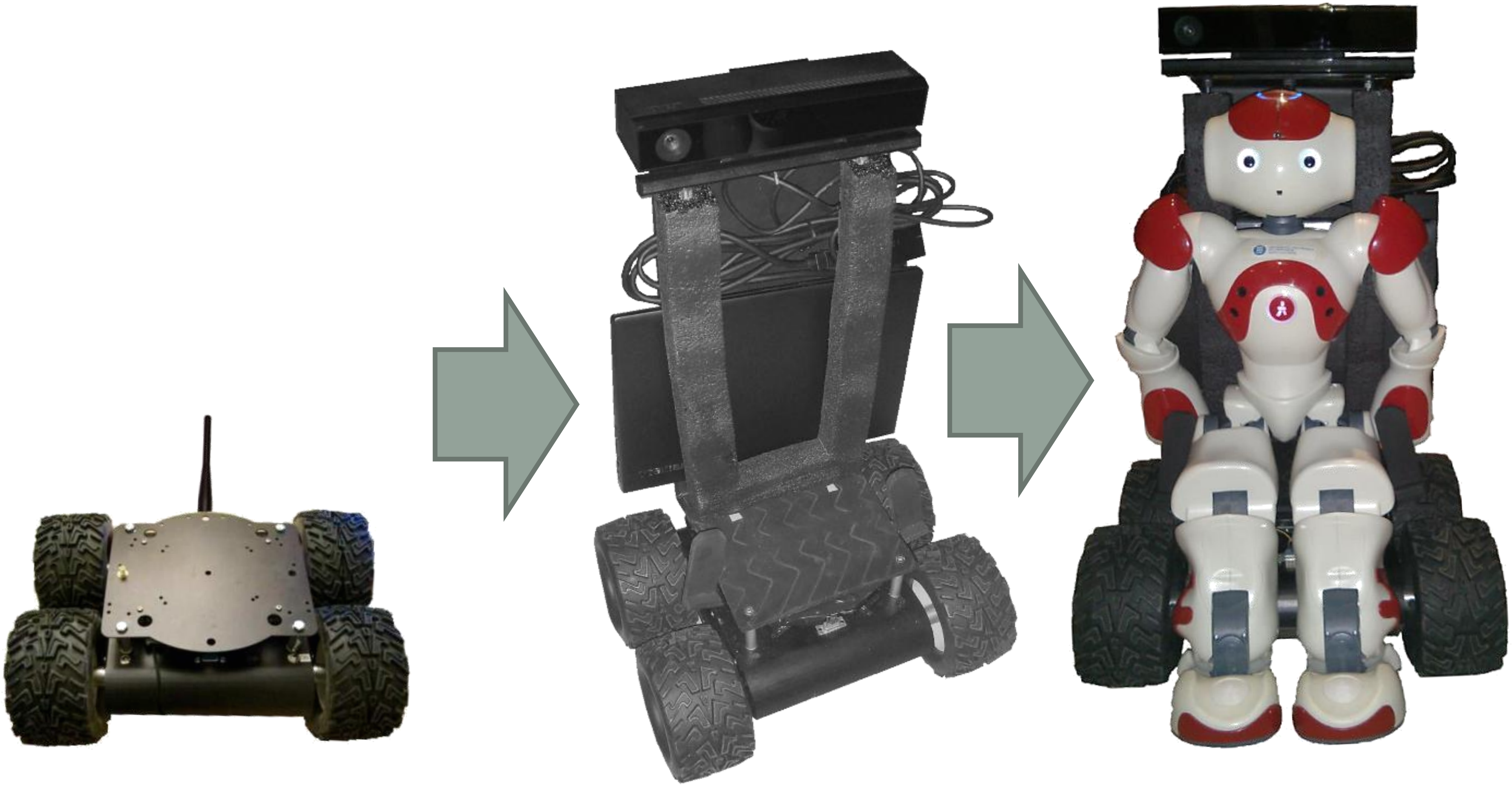
- Microsoft Kinect version 2.
  - Windows 8.1 driver and USB 3.0.

- NAO.
  - CPU Geode.
  - NoaQi OS.

- Two laptops:
  - Intel i5
  - Intel Core 2 duo

- Wifibot.
  - Intel Atom.
  - Ubuntu 12.04.

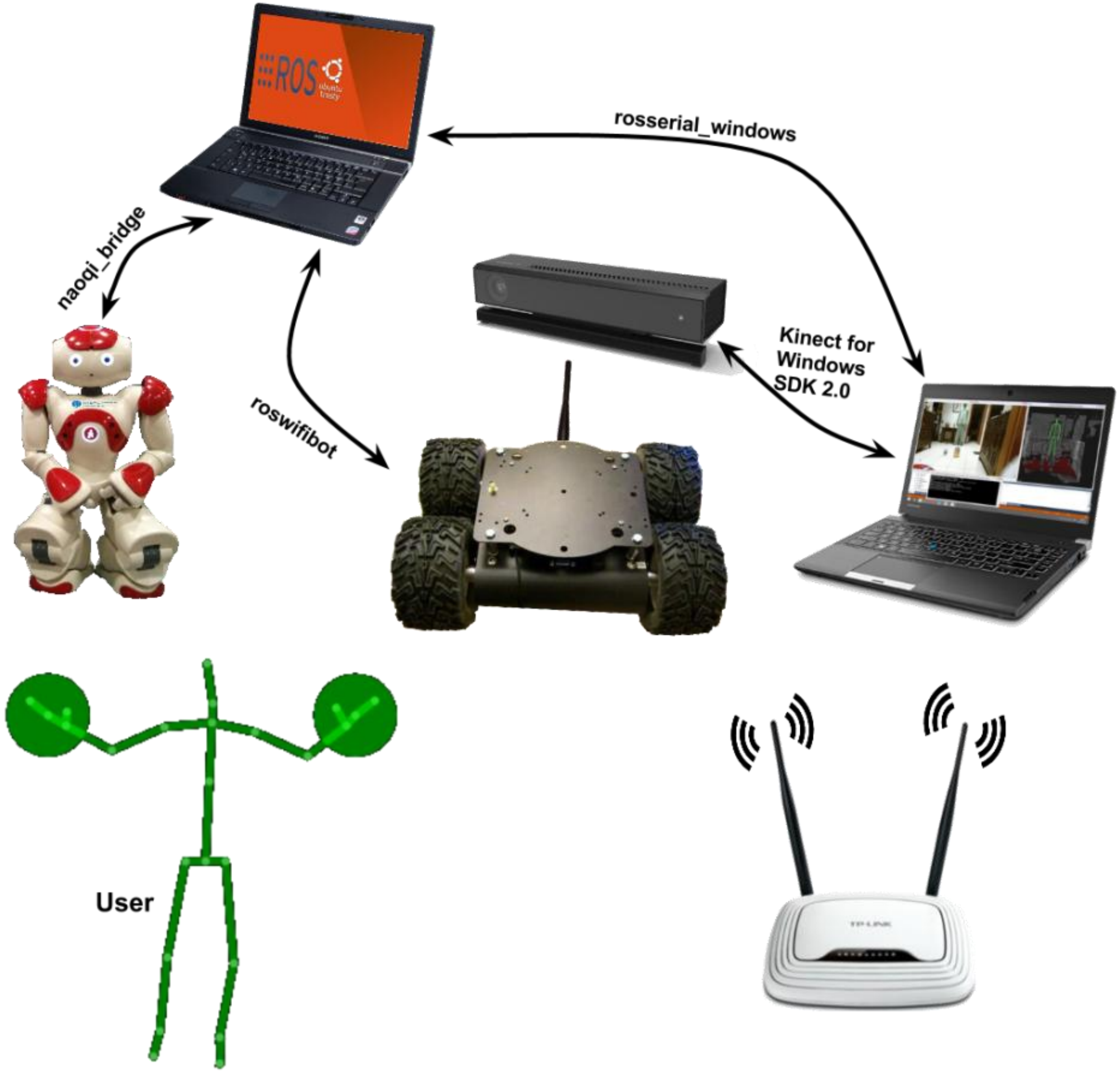
Hardware resources modifications



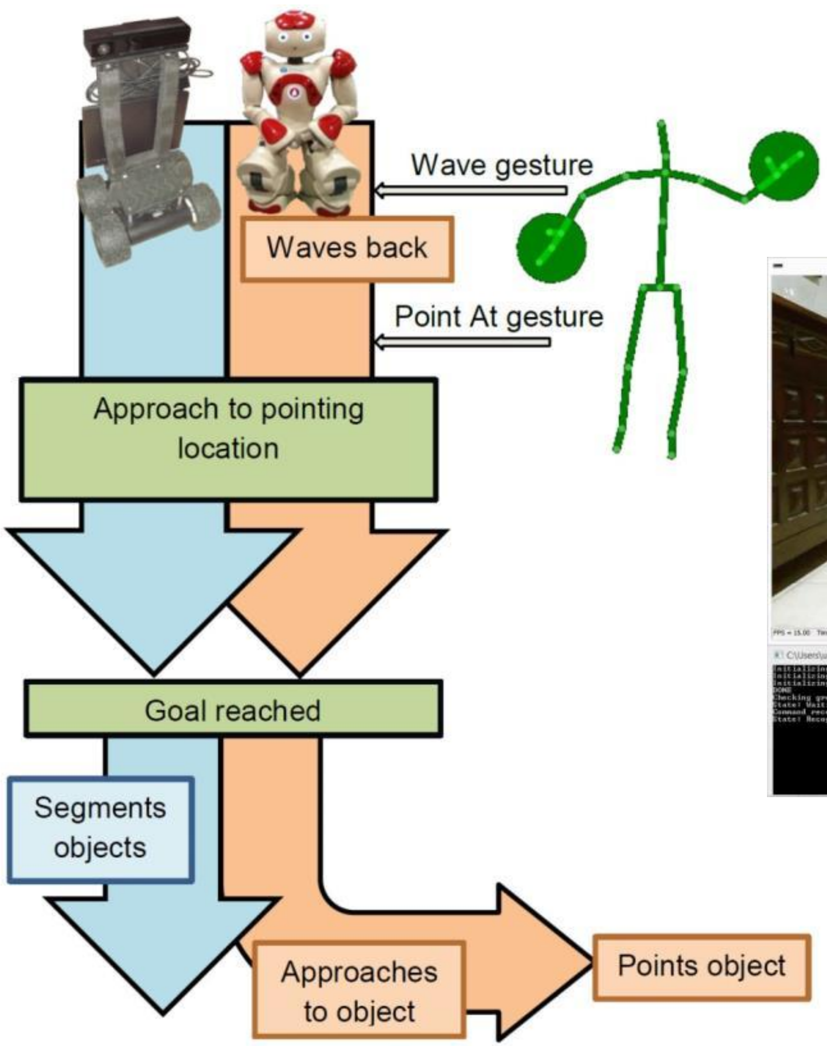
Software resources

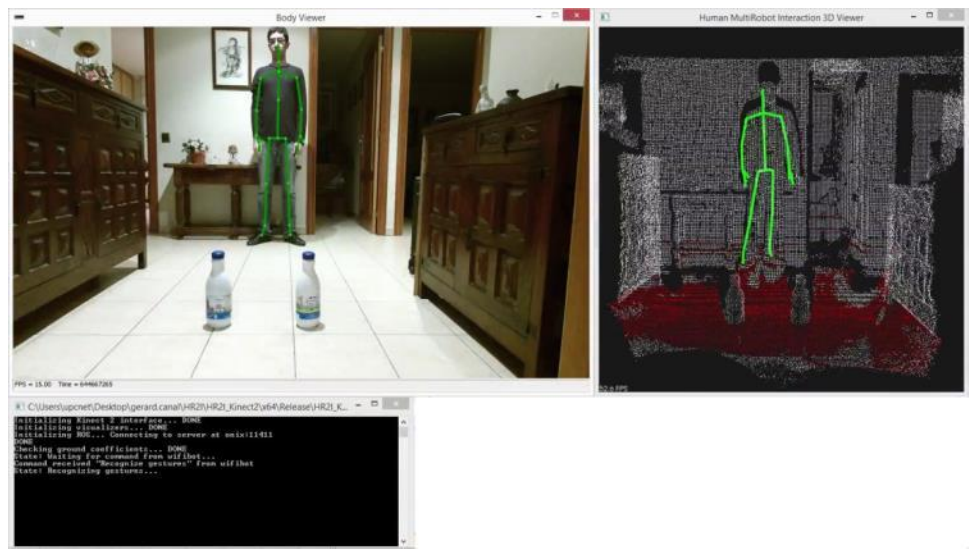
- ROS: Robot Operating System.
    - To program the robots.
  - SMACH to implement the Finite State Machines in Python.
  - Indigo Igloo version in Ubuntu 14.04.
- Kinect for Windows SDK 2.0.
  - C++ mode.
- PCL: Point Cloud Library.
  - Implemented in C++.

System overview



System overview





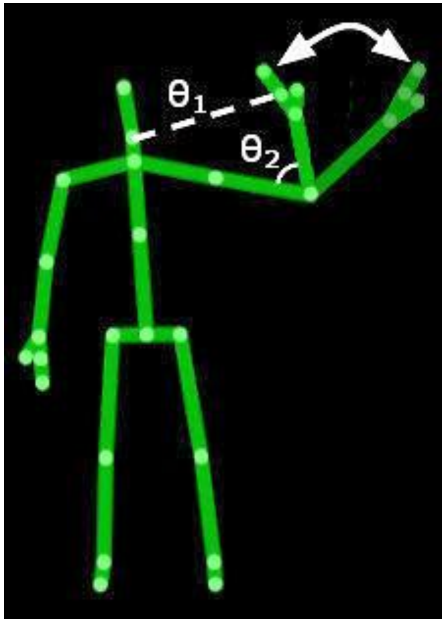
Gesture Recognition

- Two types of gestures:
  - Static
  - Dynamic
- One gesture of each type:
  - Wave
  - Point at
- Described by means of skeletal features [1].

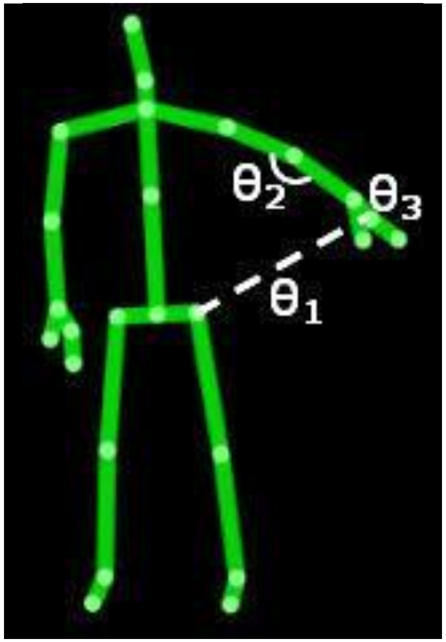
[1] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake. Real-time human pose recognition in parts from single depth images. In *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition, CVPR '11*, pages 1297– 1304, Washington, DC, USA, 2011. IEEE Computer Society.

Skeletal features

- Wave gesture:
  - θ_1: Neck – Hand distance
  - θ_2: Elbow angle

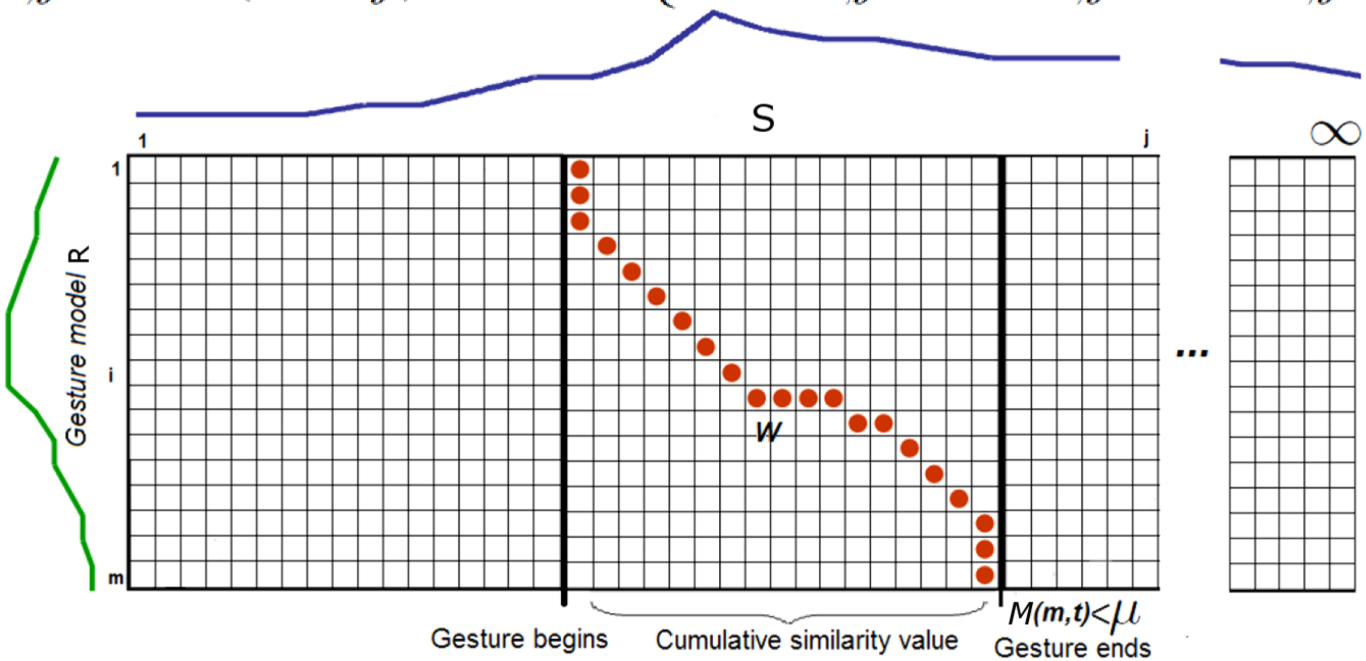


- Point at gesture:
  - θ_1: Hand – Hip distance
  - θ_2: Elbow angle
  - θ_3: Hand 3D position



Gesture recognition: Dynamic Time Warping

- Using a weighted L1 distance measure: $d_1(r,s) = \sum_{i=1}^{k} \alpha_i |r_i - s_i|$
- $M_{i,j} = d_1(r_i, s_j) + min\{M_{i-1,j}, M_{i-1,j-1}, M_{i,j-1}\}$



- A gesture is recognized when the input sequence is close enough to the model: $M_{m,k} < \mu, k \in [1, \ldots, \infty]$.

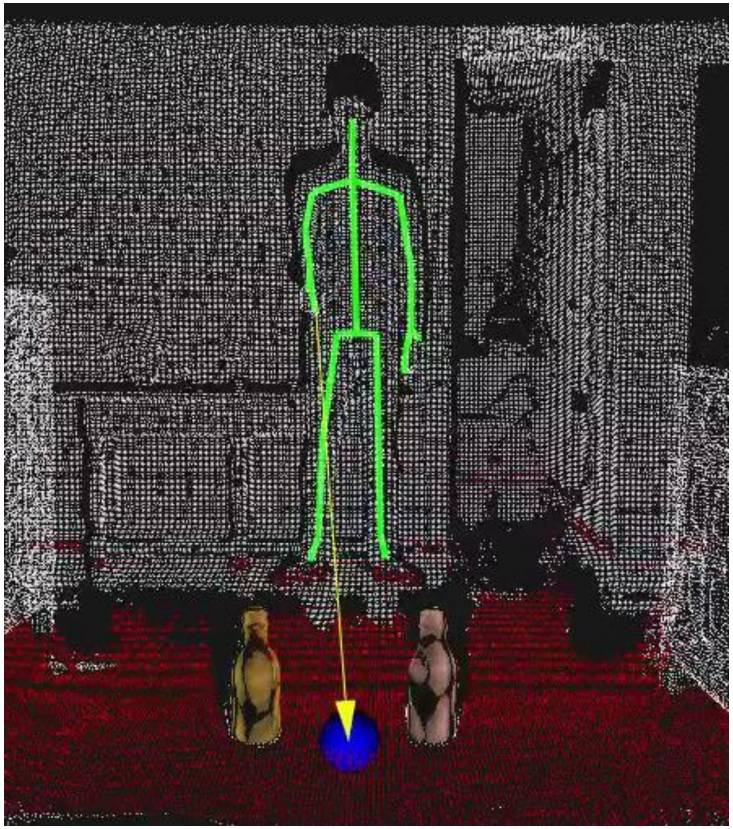
Static gesture recognition

- Check that features are within some thresholds and the involved limb is not moving during a certain number of frames.
  - $\theta_1 > T1, \theta_2 > T2$

- Dynamic and Static recognition performed in a multi-threaded joint way.

Gesture recognition: Pointing gesture related methods

- Ground plane detection by RANSAC model fitting [2].
- Pointed point extraction using skeletal joints information.
- Object segmentation by Euclidean Cluster Extraction [3].

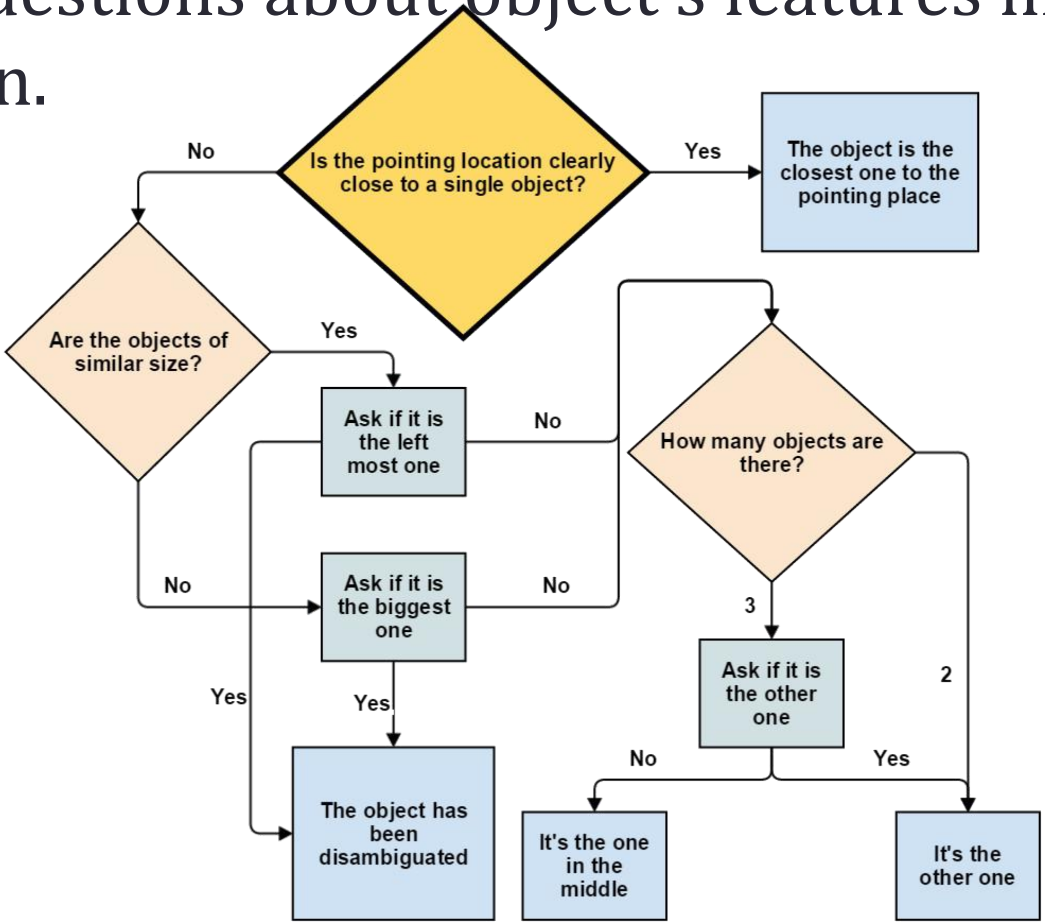


[2] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, June 1981.

[3] R. B. Rusu. Clustering and segmentation. In *Semantic 3D Object Maps for Everyday Robot Manipulation*, volume 85 of *Springer Tracts in Advanced Robotics*, chapter 6, pages 75–85. Springer Berlin Heidelberg, 2013.

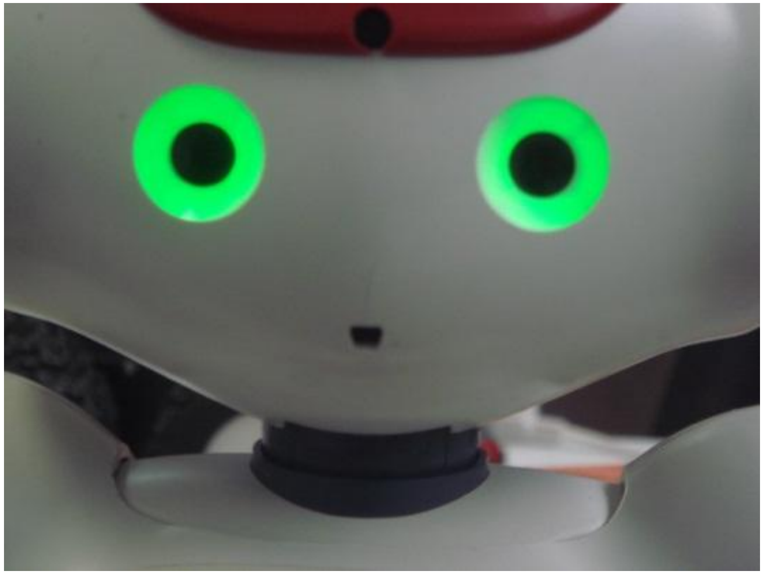
HRI methods: Object disambiguation

- Extra information may be needed in case of doubt.
- Solve it by means of a small spoken dialogue.
- Use of simple questions about object's features like size and position.



HRI methods: Interaction techniques

- The robot performs human-like gestures.
- Non-repetitive verbalization of its actions to enhance understanding.



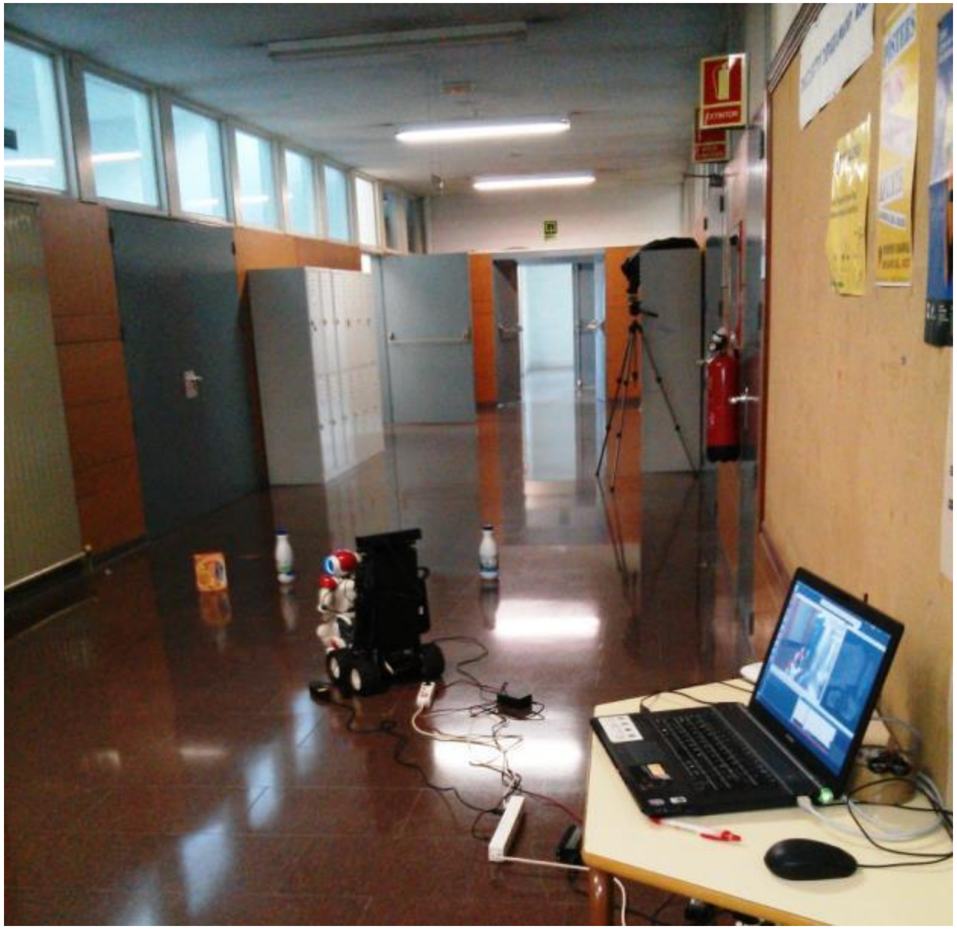
Results: Recognition performance. Jaccard index

- Performance measured on a labeled set:
  - 61 gesture samples, 27 static and 34 dynamic
  - 2082 gesture frames
- Overlap / Jaccard index as performance metric.

- LOOCV test mean Jaccard Index:
  - Static gestures: 0.46
  - Dynamic gestures: 0.49
  - Mean: 0.49

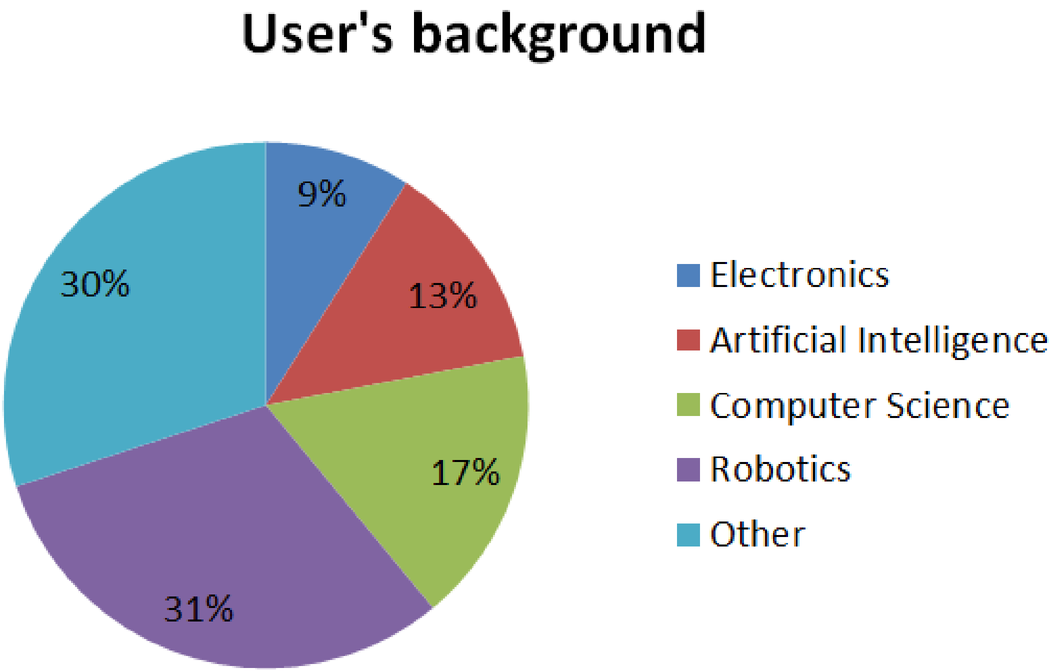
Results: User experience evaluation

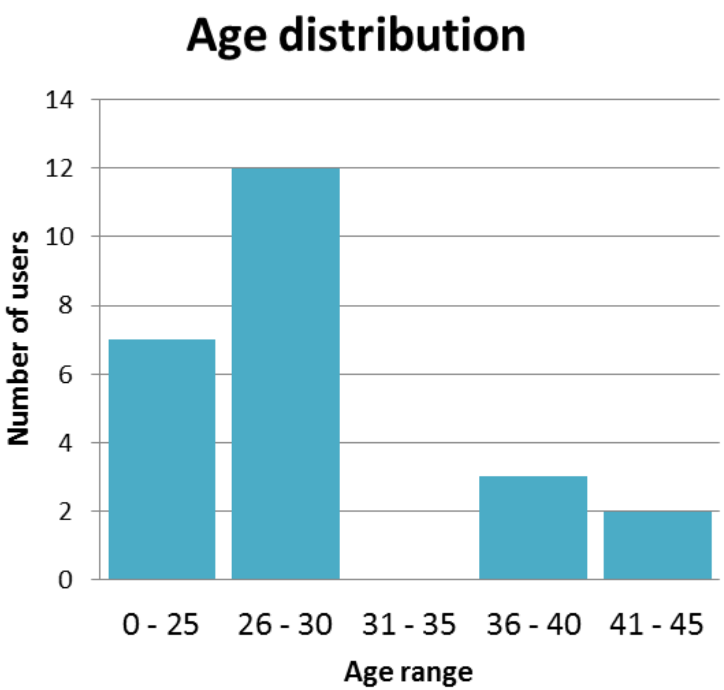
- Testing environment.



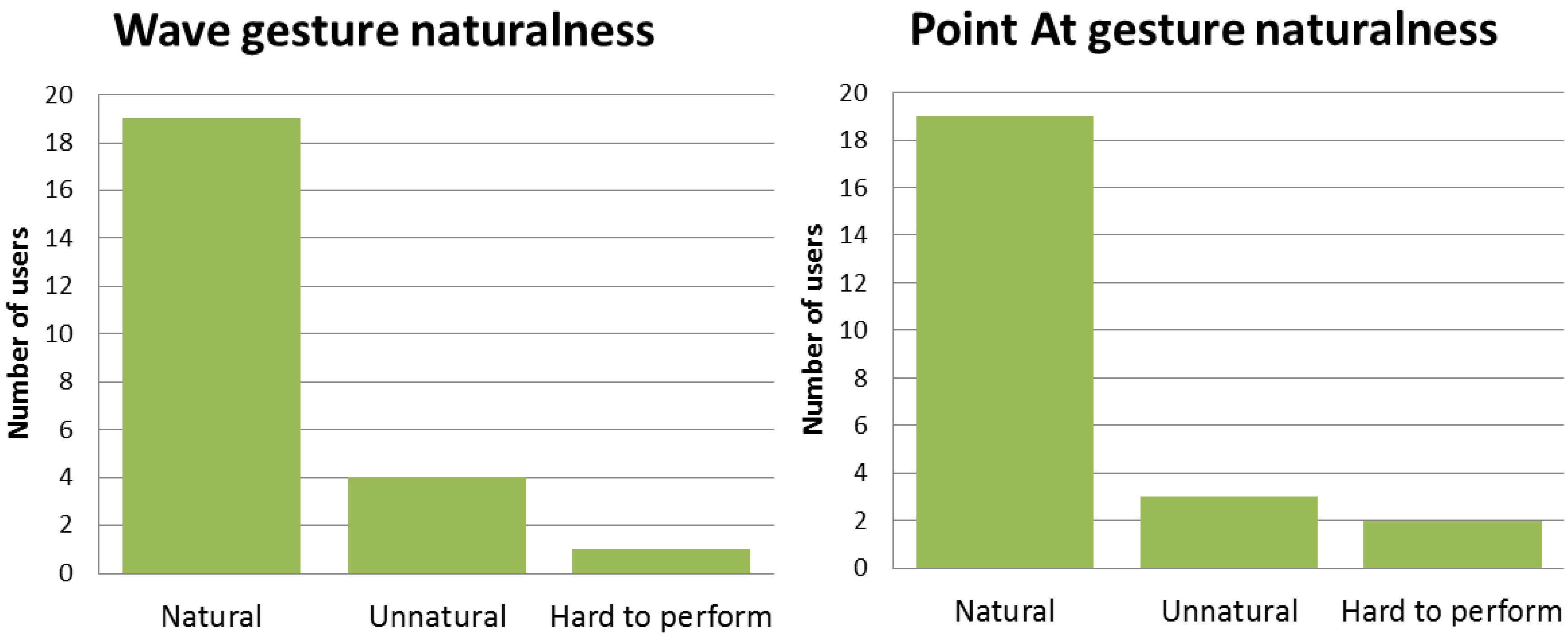
Results: User experience evaluation. Users survey

- 24 users tested the system





Results: User experience evaluation. Users survey



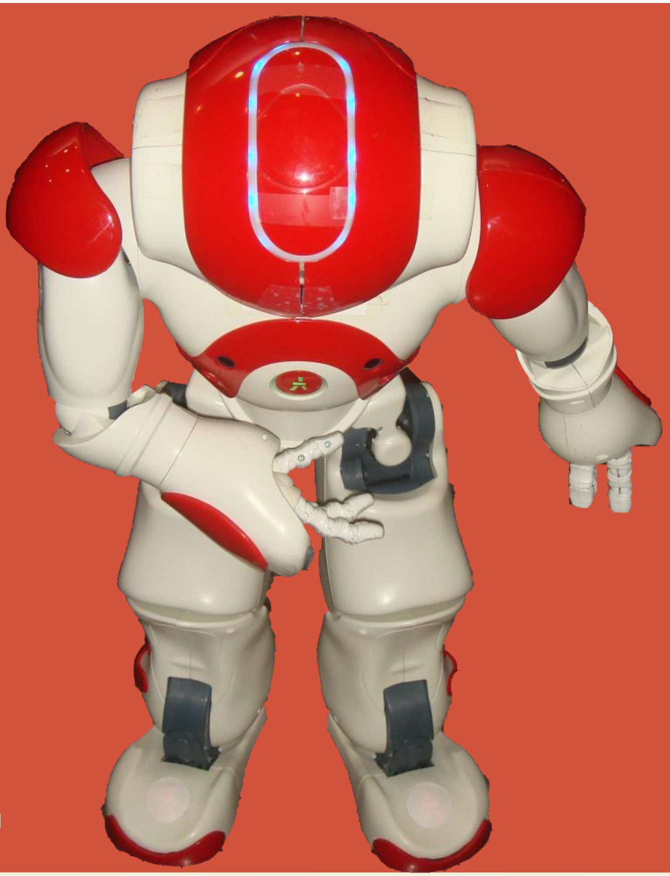
Demonstration

Conclusions

- Potential utility in household environments.
- Natural gestures as said by the test users.
- Easy to interact with the system and able to fulfill a task successfully in most of the cases.
- Working in near real time (~20 FPS), with correct response times.
- Generic and scalable framework.

Future improvements

- Enhancement of the pointing location estimation:
  - Solve user pointing imprecisions by learning from them.
  - Use of other cues such as gaze direction.
  - Hand pose estimation.
- More precise navigation (no free path assumption, scene understanding).
- Affective and cognitive interaction.



THANK YOU.

***No robot was harmed in the making of this paper.*