Gdr Robotique

Framework for Human Robot Social Interactions Application to robocup@home Competitions

Jacques Saraydaryan[1,2], Raphael Leber [1], Fabrice Jumel[1,2]
[1]CPE Lyon, France
[2]CITI Lab., INRIA Chroma

Presented by Fabrice Jumel

RoboCup

WORLDWIDE 2021

"By the middle of the 21st century, a team of fully autonomous humanoid robot soccer players shall win a soccer game, complying with the official rules of FIFA, against the winner of the most recent World Cup.

4 Major Leagues

RoboCupSoccer

RoboCupRescue

RoboCup@Home

RoboCupIndustrial

RoboCupJunior

Robocup Soccer

2050 ?

RoboCup@Home

"Develop service and assistive robot technology with high relevance for future personal domestic applications.
It is the largest international annual competition for autonomous service robots and is part of the RoboCup initiative.

Social Standard Platform League (SSPL)

Open Platform League (OPL)

Domestic Standard Platform League (DSPL)

robocup@home

Dynamic Navigation

Decision in dynamic environnement

Interaction in natural language

Visual scene analysis

Environnement analasys

Gesture Recognition

Objects Manipulation

Objects Recognition

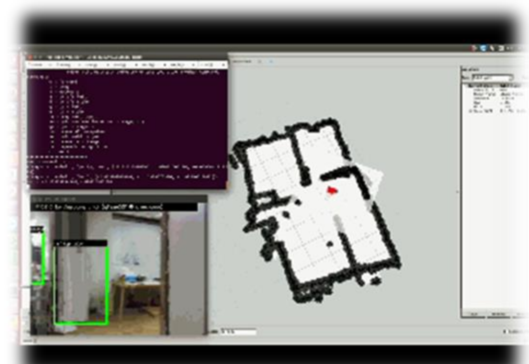
Human Robot Interaction

Human identification and re-identification

People following

...

Guide Robot
Companion Robot
Pesonal assistance Robot
Waiter Robot
Butler Robot



CHROMA

Context (1/2)

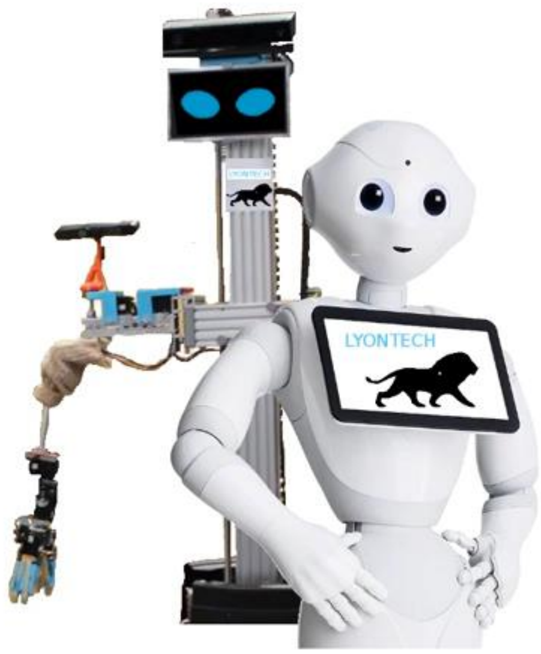
- **RoboCup@Home**
  - **Evaluation of** current **domestic robots** through **real life scenarios**.
  - **A set of benchmark tasks** is used to evaluate the robots' abilities and performance in realistic home environment settings.
  - Focus lies on **human-robot-interaction**, **navigation** and **mapping, object manipulation** in dynamic environments.

For example, during the "Party Host scenario" trials, robots provide general assistance to guests during a party (welcome, introduce a new guest to others, describe guests to the bartender, escort an exiting guest to a cab ...).

→ **(Focus) Needs of people management abilities :**

- **high level info : pose estimation, body description, clothing description**
- **Comprehensive People description**
- **Recognition / Finding of a specific people**
- **People tracking**

→ *In the case of a domestic robot, we need a framework able to provide all these features with **only onboard sensors as 2D camera***


LYONTECH

LyonTech

RoboCup@Home Team

2nd place SSPL World Champion 2021

and even more…

Raphael Leber[1], Sébastien Altounian[1], Simon Ernst[1,5], Florian Dupuis[1], Jeanne Fort[1], Fabrice Jumel[1,3,4], Cedric Mathou[1], Benoit Renault[2,3,4], Jacques Saraydaryan[1,3,4], Olivier Simonin[2,3,4]

[1]CPE Lyon, [2]INSA Lyon, [3]INRIA Chroma team, [4]CITI Lab., [5]Palo IT

Context (1/2)

- **RoboCup@Home**
  - **Evaluation of** current **domestic robots** through **real life scenarios**.
  - **A set of benchmark tasks** is used to evaluate the robots' abilities and performance in realistic home environment settings.
  - Focus lies on **human-robot-interaction**, **navigation** and **mapping, object manipulation** in dynamic environments.

For example, during the “Party Host scenario” trials, robots provide general assistance to guests during a party (welcome, introduce a new guest to others, describe guests to the bartender, escort an exiting guest to a cab ...).

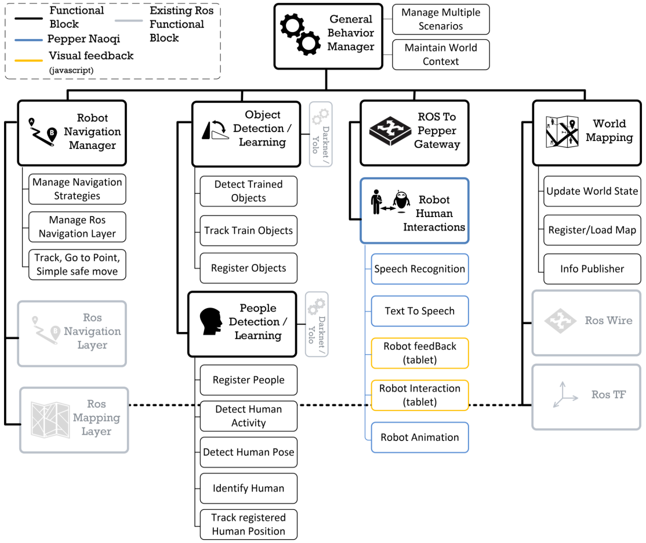
→ **(Focus) Needs of people management abilities :**
- **high level info : pose estimation, body description, clothing description**
- **Comprehensive People description**
- **Recognition / Finding of a specific people**
- **People tracking**

→ *In the case of a domestic robot, we need a framework able to provide all these features with **only onboard sensors as 2D camera***

Orchestration of high-level abilities

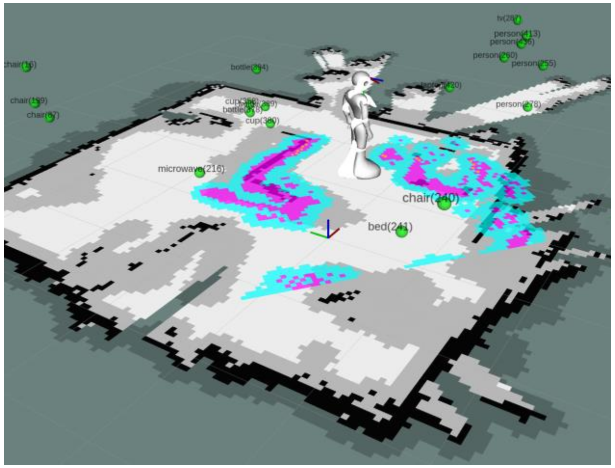
LyonTech Architecture

→ Object Detection/Learning
→ Human-Robot interaction
→ World Mapping
→ Robot Navigation



Navigation Selection Strategy

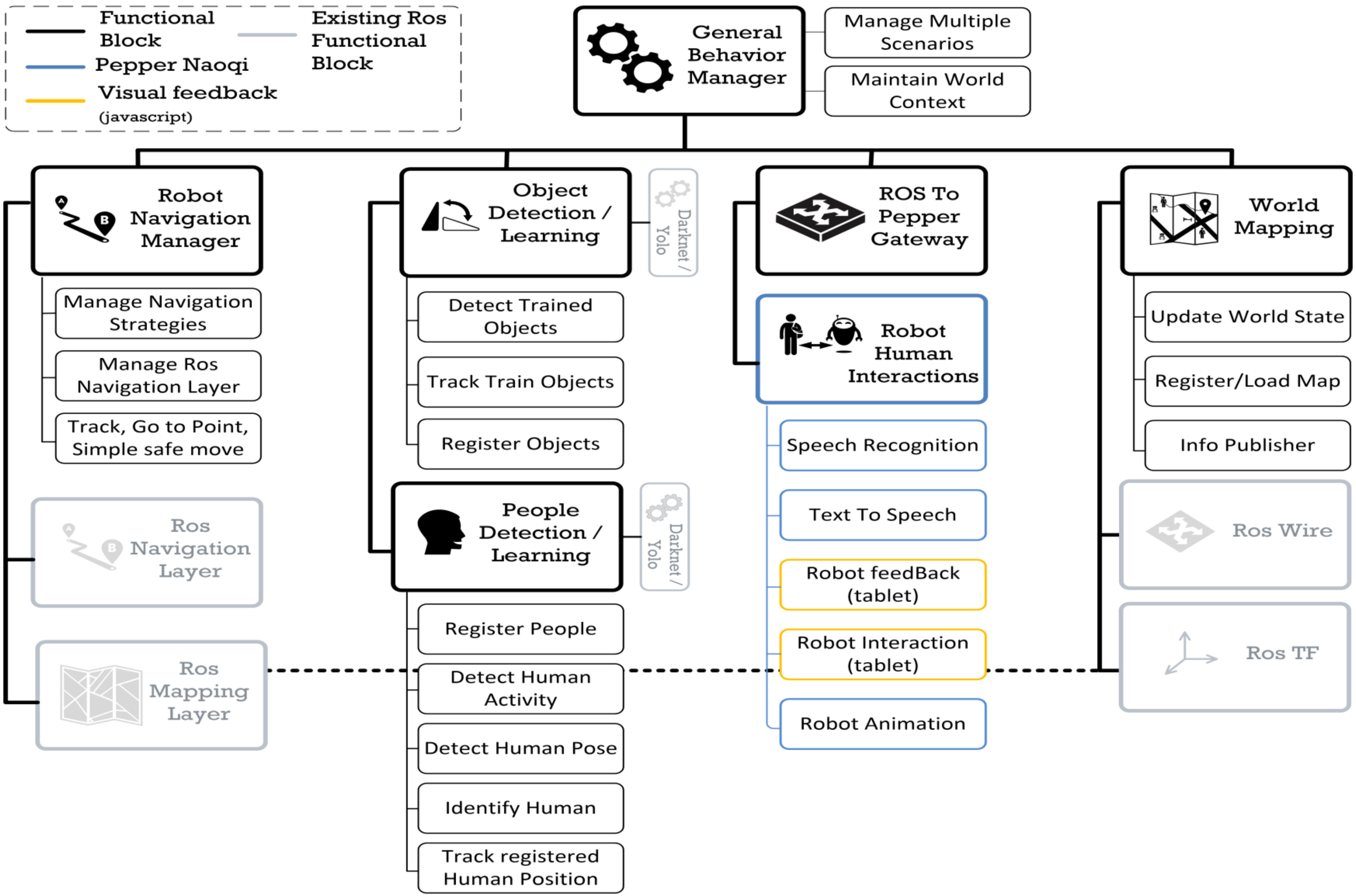
→ based on robot's environment context



High functions based on

→ DeepLaerning
→ Social Navigation

Framework for Human Robot Social Interactions



Focus on People Management

- Some characteristics (typically a person's positions) vary over time, use of tracking approach, as **Multiple Object Tracking (MOT)** [1,2], is needed.

- A modern approach would be to define all the characteristics needed and **train a neural network**. Unfortunately, the creation and labeling of such a large and complex dataset **is not possible.**

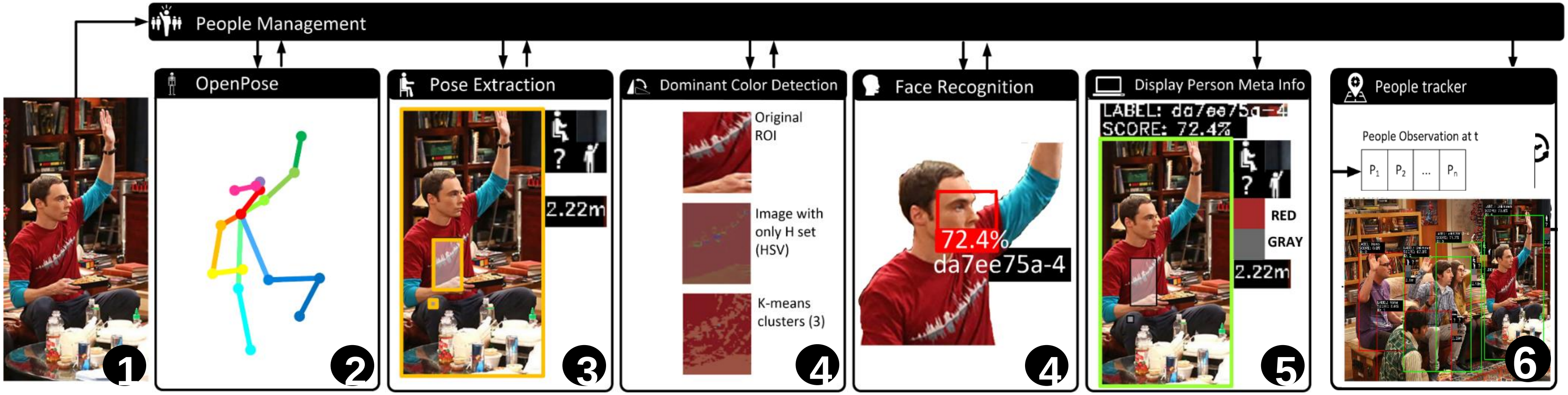
- **Practical approach is needed to aggregate different features** (mostly based on deeplearning based tools) and merge them:
  - Relevant works have been made on MOT applied to people tracking [1,2], but few of them are from a human (or robot) eye's perspective (e.g MOT16).
  - When people disappear and reappear, trackers need to re-identify people and associate them with a previous identity. This process, called **Person Re-Identification (PReID)** [3], uses different collected persons characteristics.
  - A RoboCup@Home team developed a general **tracking tool for MOT called "wire"** [4].
  - Another team defined a **specific framework for "Person-Following"** tasks [5] based on OpenPose tools [6] and color features extraction.

→ Need to get **modular goal oriented people features**
→ Need framework to **aggregate** people features and **track / re-identify** people over the time

[1] L. Wenhan et al . Multiple object tracking: A review. CoRR, abs/1409.7618, 2014, last 2017.
[2] A. Bewley, et al . Simple online and realtime tracking. In 2016 IEEE International Conference on Image Processing (ICIP), pages 3464–3468, 2016
[3] B. Lavi, et al. Survey on deep learning techniques for person re-identification task. CoRR , abs/1807.05284, 2018
[4] J. Elfring, et al. , Semantic world modeling using probabilistic multiple hypothesis anchoring. ,Robot. Auton. Syst. , 61(2):95–105, February 2013
[5] K. Minkyu et al. , An architecture for person-following using active target search. CoRR, abs/1809.08793, 2018.
[6] Z. Cao, et al., Open-pose: Realtime multi-person 2d pose estimation using part affinity fields. CoRR, abs/1812.08008, 2018

Focus on the People Management Architecture

We propose an architecture that provide people pose and posture, clothing colors,face recognition and offer tracking and re-identification abilities.



1. An Image is received
2. Joints are extracted through OpenPose
3. People joints are then processed to determine person Pose (Standing,Sitting,Lying,..), Region Of Interest (ROI) and person estimated distance
4. Extracted ROI are used to determine dominant colors of Tshirt and Trouser, and make face recognition.
5. All people information is gathered and displays
6. Tracker and re-identification can be provide

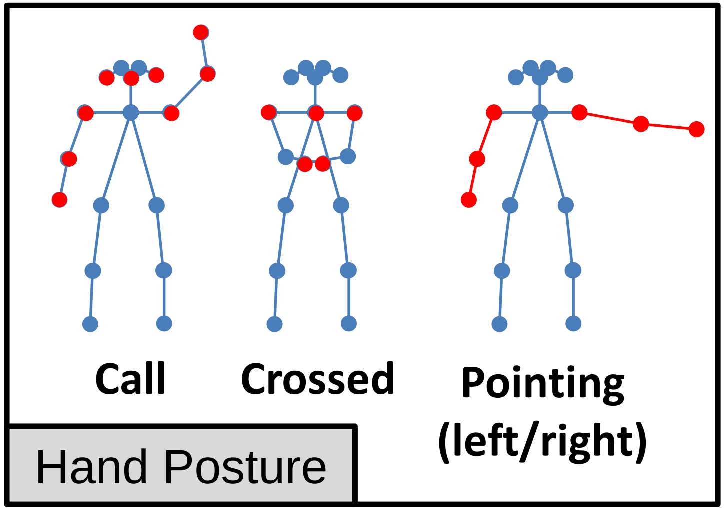
People Posture:

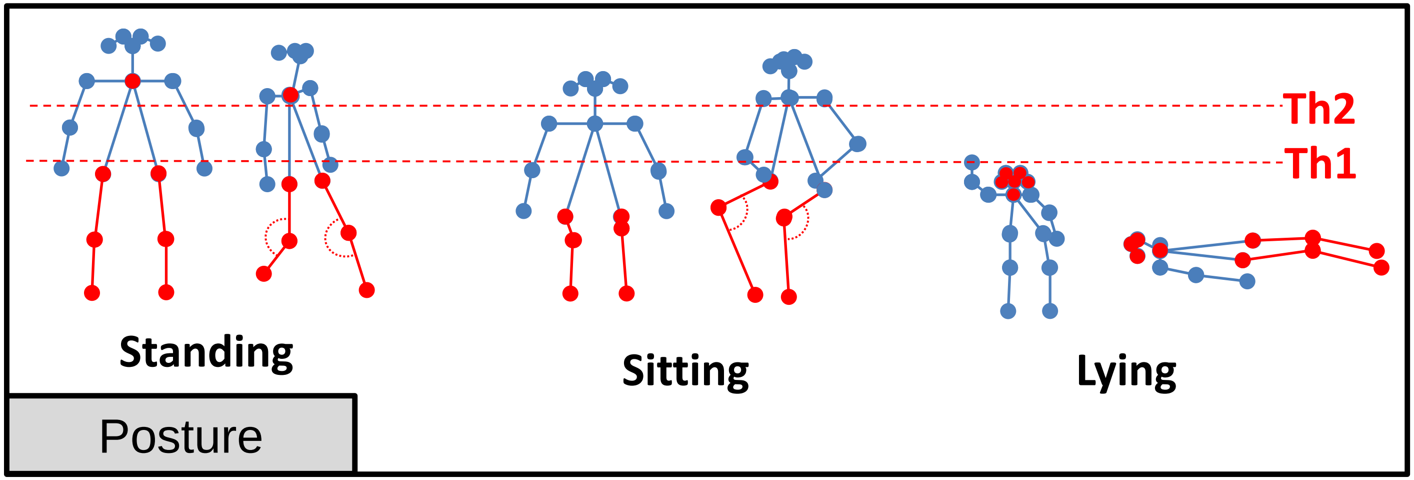
Goal : Compute people posture (hand and body)

Data: 2D body key points out of OpenPose

Process: Scoring system with one or several criteria on each posture (interest limbs/joints displayed in red)

Hypothesis (H1): Camera horizontal field of view is parallel to the ground (flat ground horizon doesn't go upper $^1/_2$ of image height)



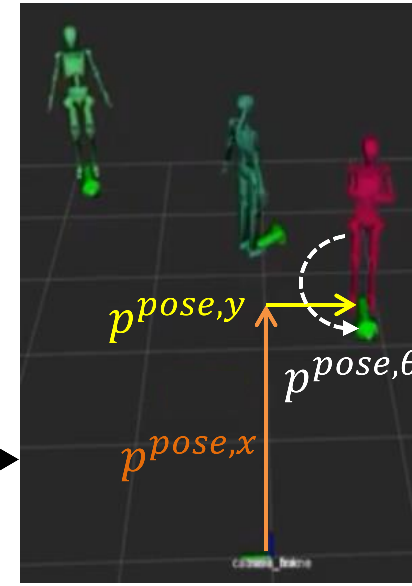


Based on **H1** Thresholds **Th1** and **Th2** are used in one of the Standing/Lying criteria

People Pose:

Goal : Compute people pose with a 2D camera

p^{pose} (x,y,θ) is the estimated pose (position and orientation), expressed in a “top-view” map with the robot as the origin



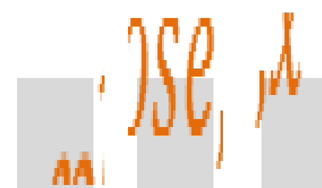
$p^{pose,\theta}$ estimation method

$$\psi = \frac{\sum bodypart\_confidence_{right} - \sum bodypart\_confidence_{left}}{\sum bodypart\_confidence_{right} + \sum bodypart\_confidence_{left}} \quad (1)$$

$$\boxed{p^{pose,\theta} \sim \alpha * \psi + \beta}$$

Equation (1) estimates people orientation ratio based on right and left confidence of people body parts (face and shoulder only). People front or back side are defined by shoulder sides and/or nose presence. Depending on front/back side, we compute α (−π/2 or π/2) and β (0 or π) in order to get an orientation angle $p^{pose,\theta}$

Examples:

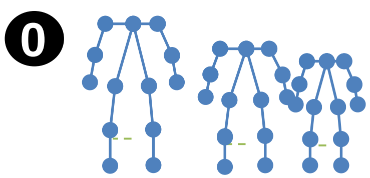


ψ	0.12	0.6	0.01	
α	π/2	-π/2	π/2	rad
β	π	0	π	rad
$p^{pose,\theta}$	3.33	-0.94	3.16	rad

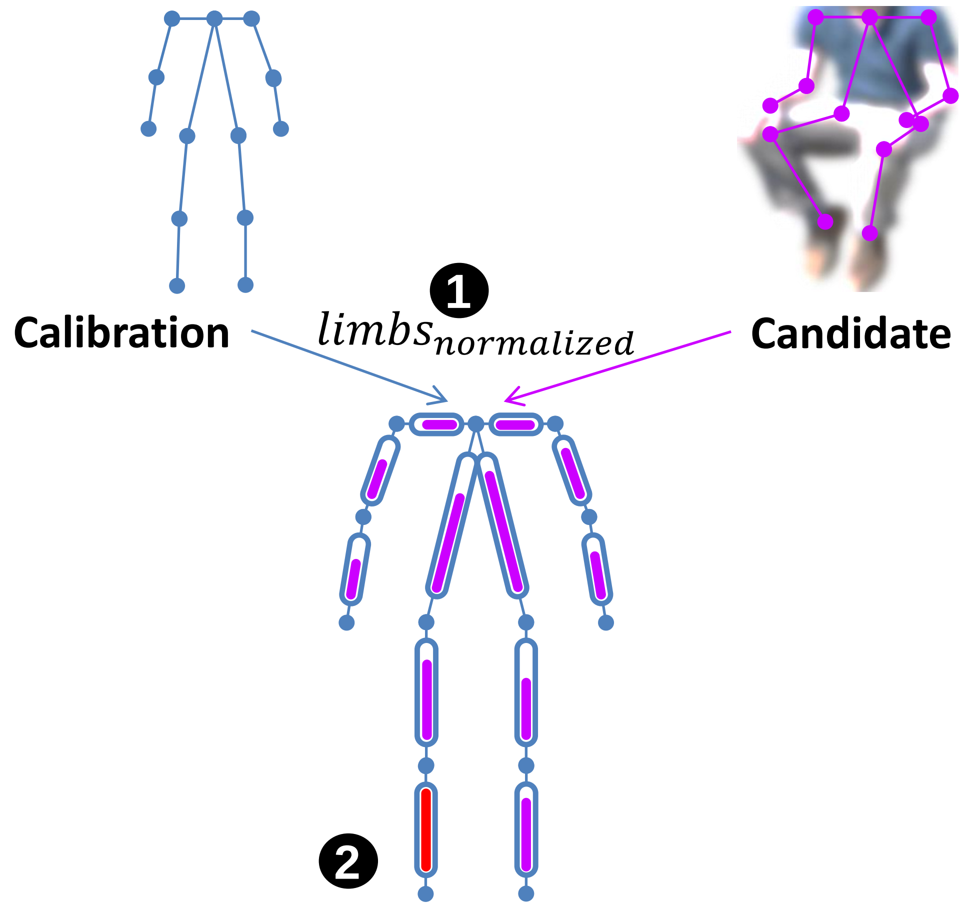
People Pose:

$p^{pose,x}$ $p^{pose,y}$ estimation method

Calibration data : Record at every meter of an average size person, straight on his legs and facing the camera, in order to maximize the limbs components on the 2D camera plane



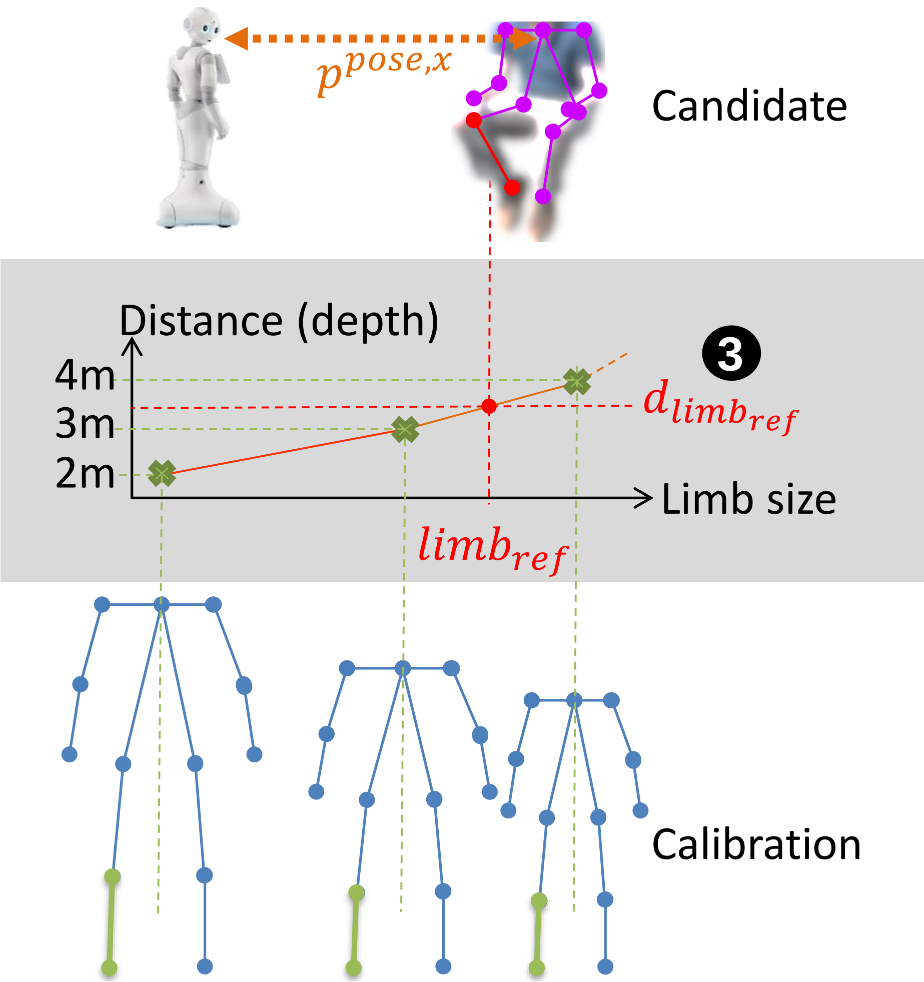
Hypothesis:
At least one limb is seen with most of its components on the image plane



$limb_{ref} = \max(limbs_{normalized})$

People Pose:

$p^{pose,x}$ $p^{pose,y}$ estimation method

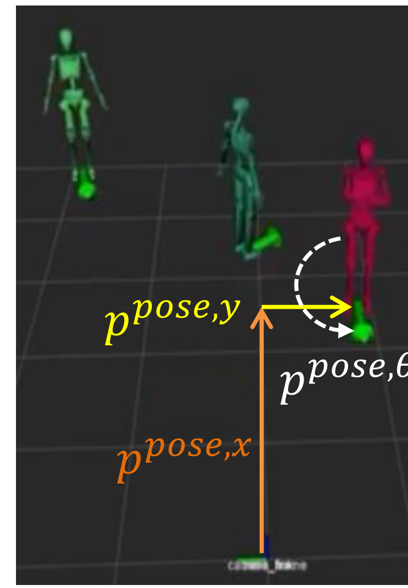


4

$$p^{pose,x} \sim d_{limb_{ref}}$$

5

$$p^{pose,y} \sim p^{pose,x} * \sin \frac{H_{FOV} * (p^{neck,x} - (\frac{i_w}{2}))}{i_w}$$



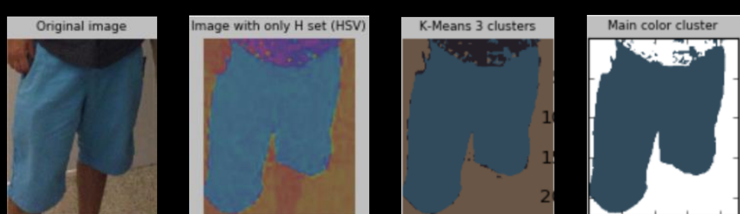
with H_{FOV} = **Horizontal Field Of View**
$p^{neck,x}$ = *image horiz. neck coordinate*
i_w = *Image width*

Color Detection

Workflow:

1. Convert RGB matrix of a given ROI to HSV Matrix
2. Compute Kmean cluster on the Hue value of the HSV color
3. Select main cluster
4. Set Saturation (S) and value (V) and associate the closest X11 [7] color name
5. Add information of darkness and grey white and black value through thresholds

1 2 3

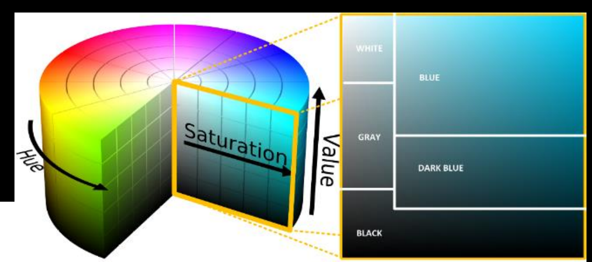


4 X11 color Name

BLUE

5 Adjusted Name

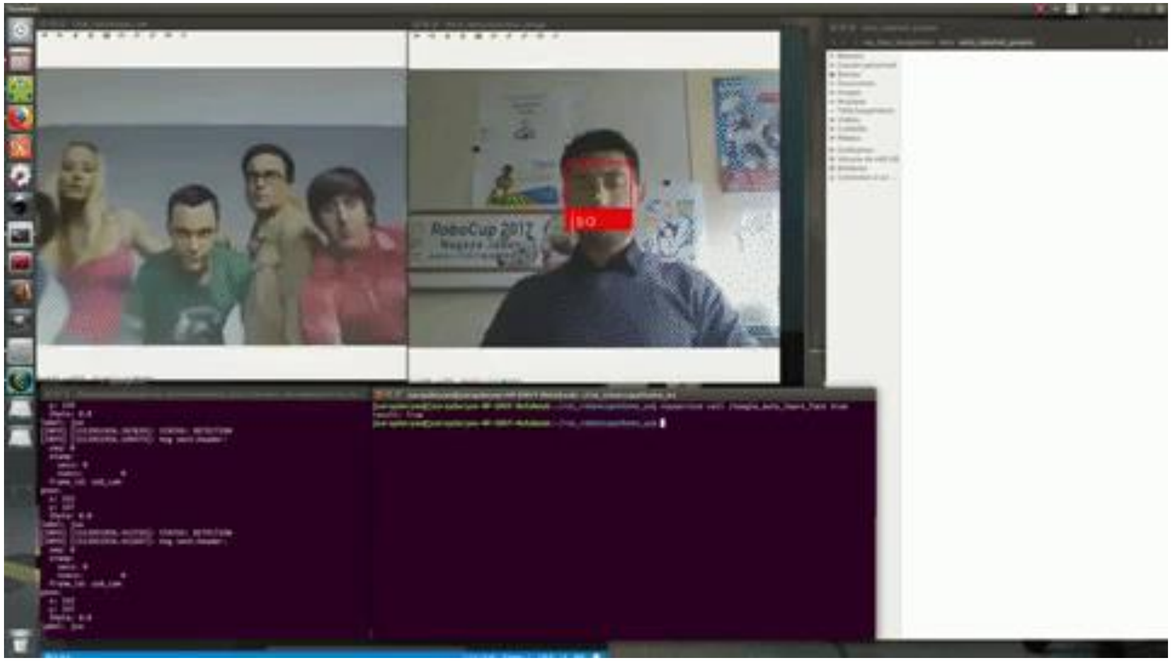
DARK BLUE



[7] B. Pettit et al., Css color module level 3. W3c recommendation, W3C, 2018

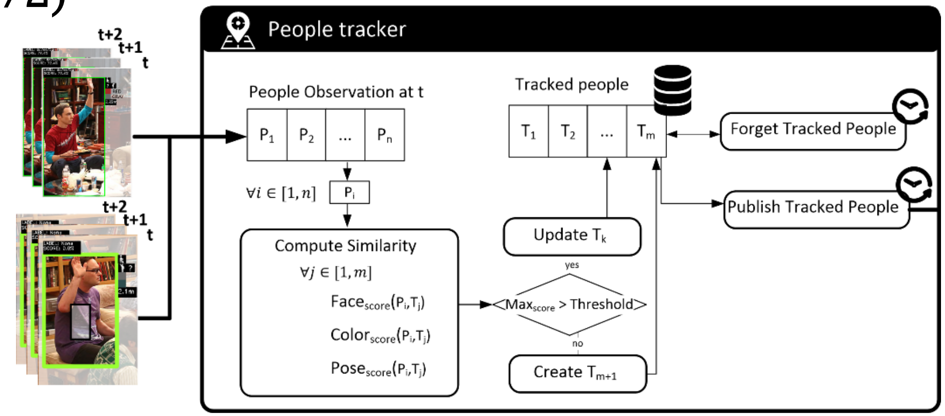
Face Recognition

- Based on the Adam Geitgey’s (based on the ResNet-34 [8]) library
- Complete the Face detection HOG [9] with Haar Cascades [10] and bounding box from OpenPose
- Add automatic face learning if unknown



[8] H. Kaiming H. et al. , Deep residual learning for image recognition. arXiv preprint arXiv:1512.03385 , 2015
[9] N. Dalal et al., Histograms of oriented gradients for human detection. Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 2005
[10] P. Viola et al. , Rapid object detection using a boosted cascade of simple features. In Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001 , volume 1, pages I–I, 2001

People Tracking (1/2)



- **Similarity Score**
  - Compute a similarity score between enriched person observation *p*, and tracked person T_i

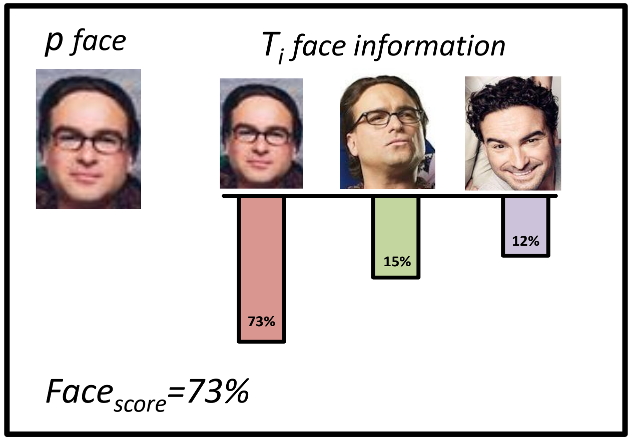
$$\mathbf{general_{score}(p, Tj)} = \sum \boldsymbol{w_i . Feature_{i_{score}}(p, T_j)}$$

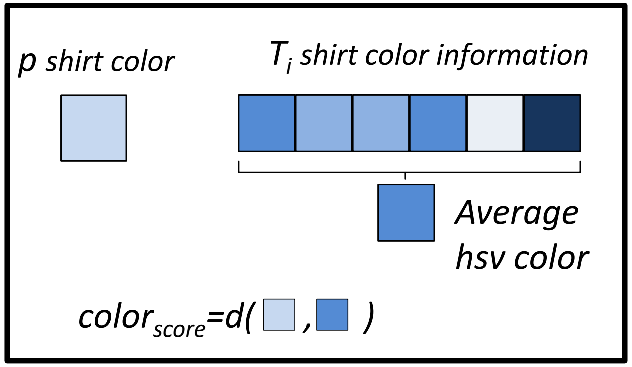
  - Where $Feature_{i_{score}}$ represents similarity of detected person features (e.g Face) and already track ones.
  - Score weights could be computed using a Boosting based algorithm [1](e.g AdaBoost) given a training labeled dataset.
- **Forget unused tracked person**
  - Periodically check if some tracked person has no been updated for a long time and remove it. This function is based on classical forgetting curve.

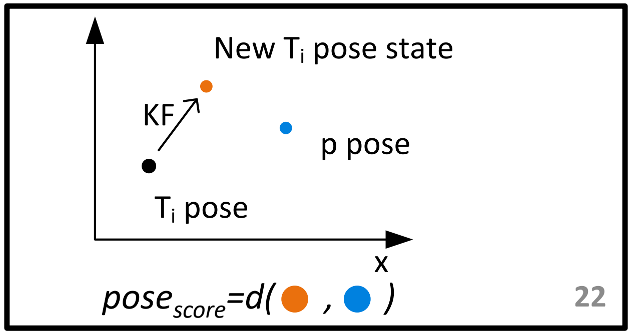
[1] L. Wenhan et al, Multiple object tracking: A review. CoRR, abs/1409.7618, 2014, last 2017.

People Tracking (2/2)

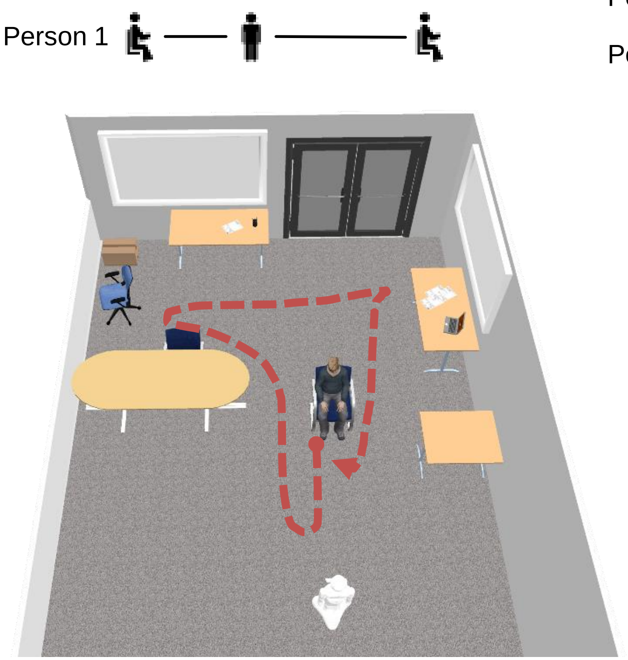
- **$face_{score}$**
  - *A same person can be associated to a set of faces. Each tracked person T_i maintains a set of face information.*
  - *$Face_{score}$ is egal to the percentage of the observed p face in T_i face information*
- **$color_{score}$**
  - *the color score is the distance d() (Hue or CIELAB ΔE*distances) between, for example, the observed p shirt color and the average hsv color of a Tracked person T_i*
- **$pose_{score}$**
  - *Kalman Filter is applied on the current tracked person T_i pose with the observed person pose p_{pose}*
  - *the pose score is the distance between observation and the new state of the system.*

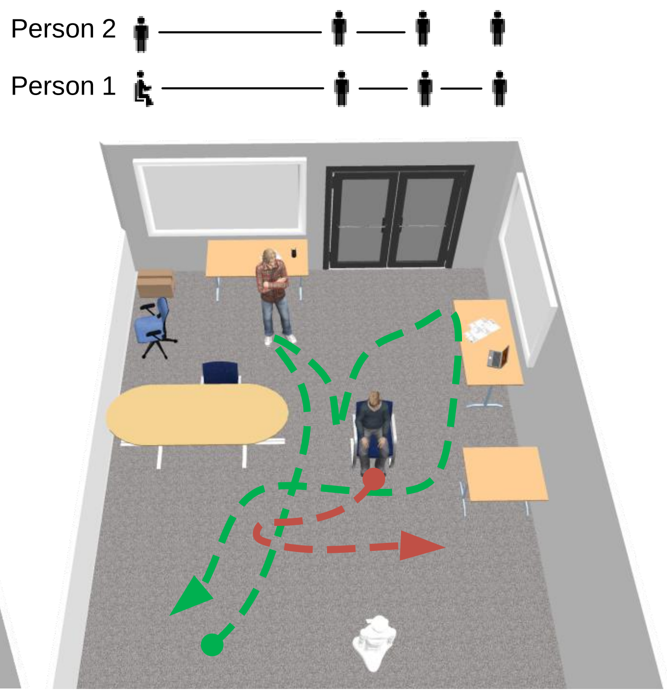


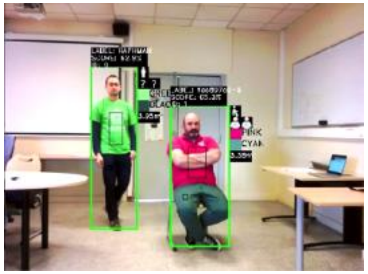


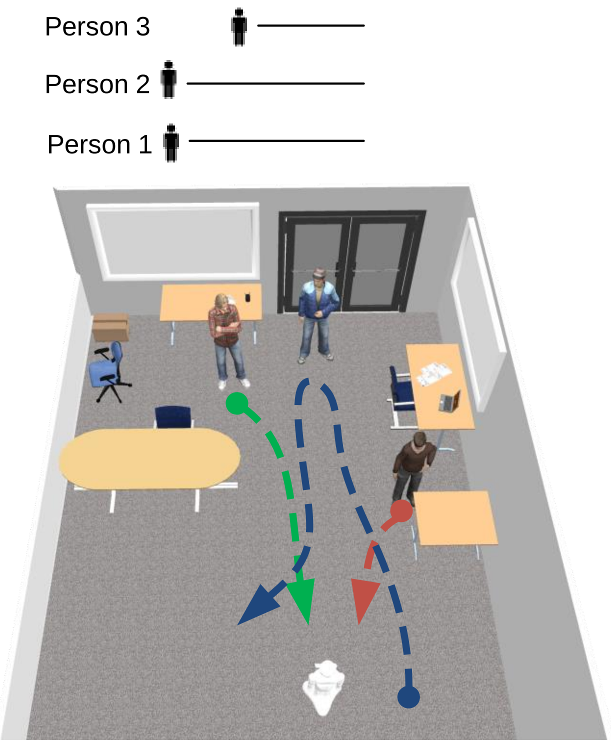


Scenario

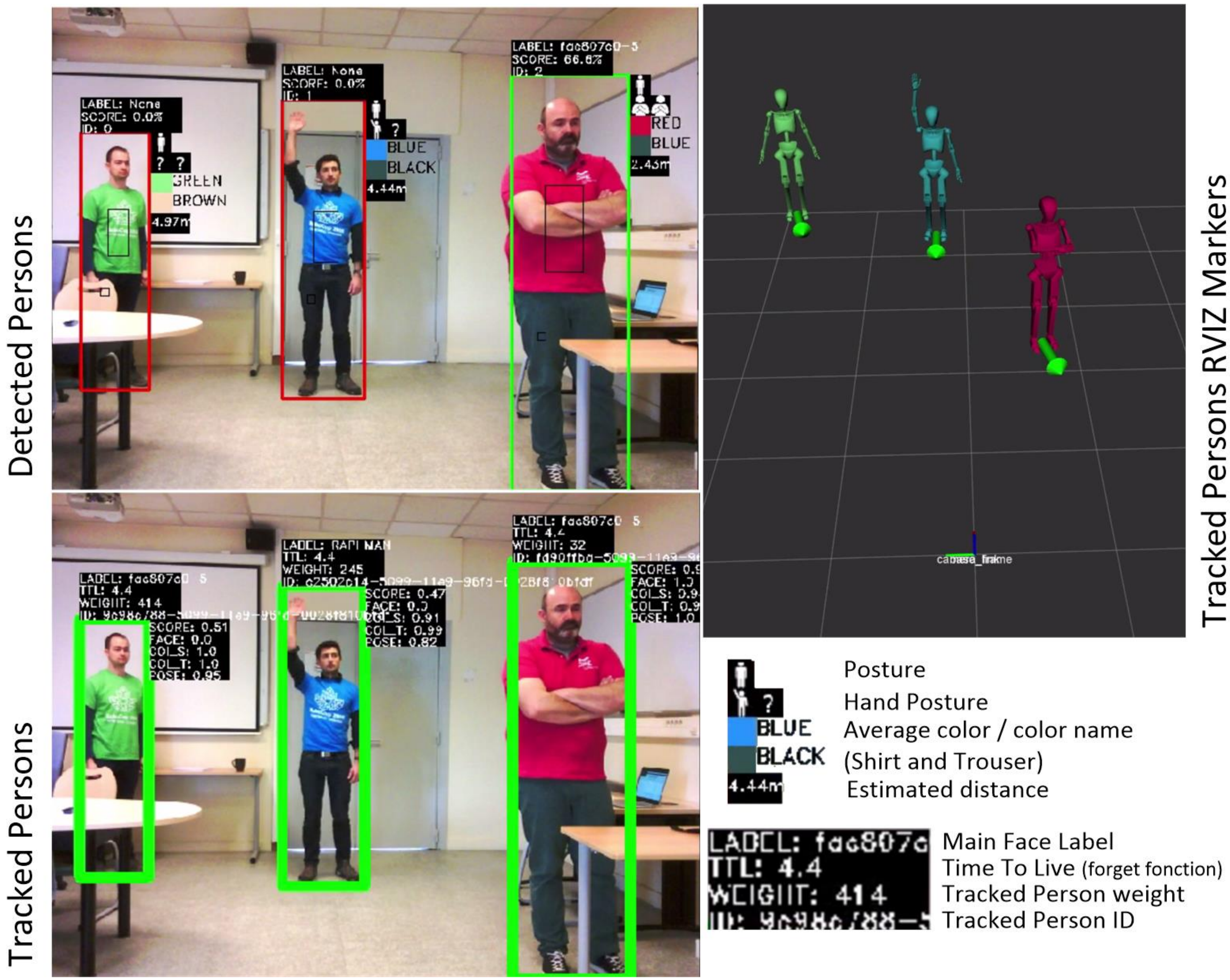




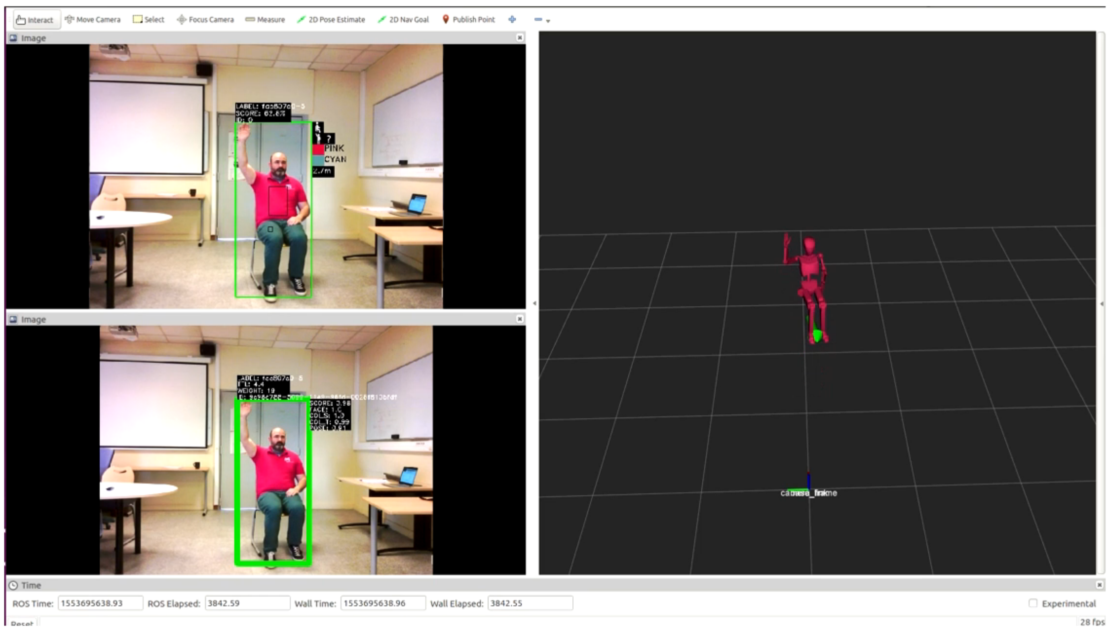


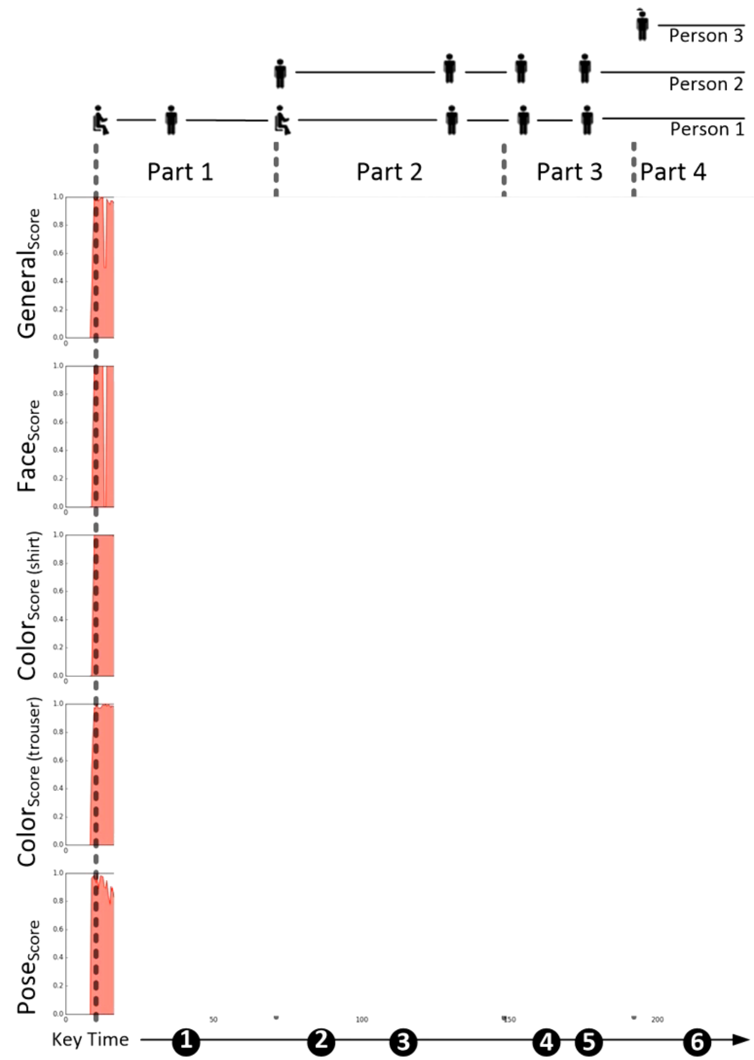


Robot FeedBack

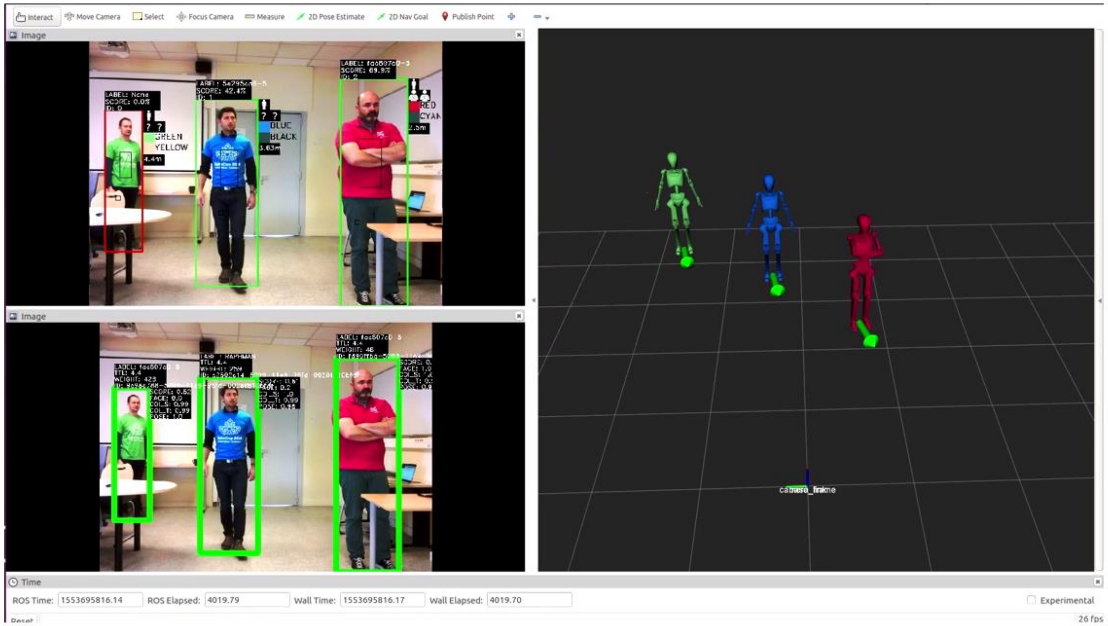


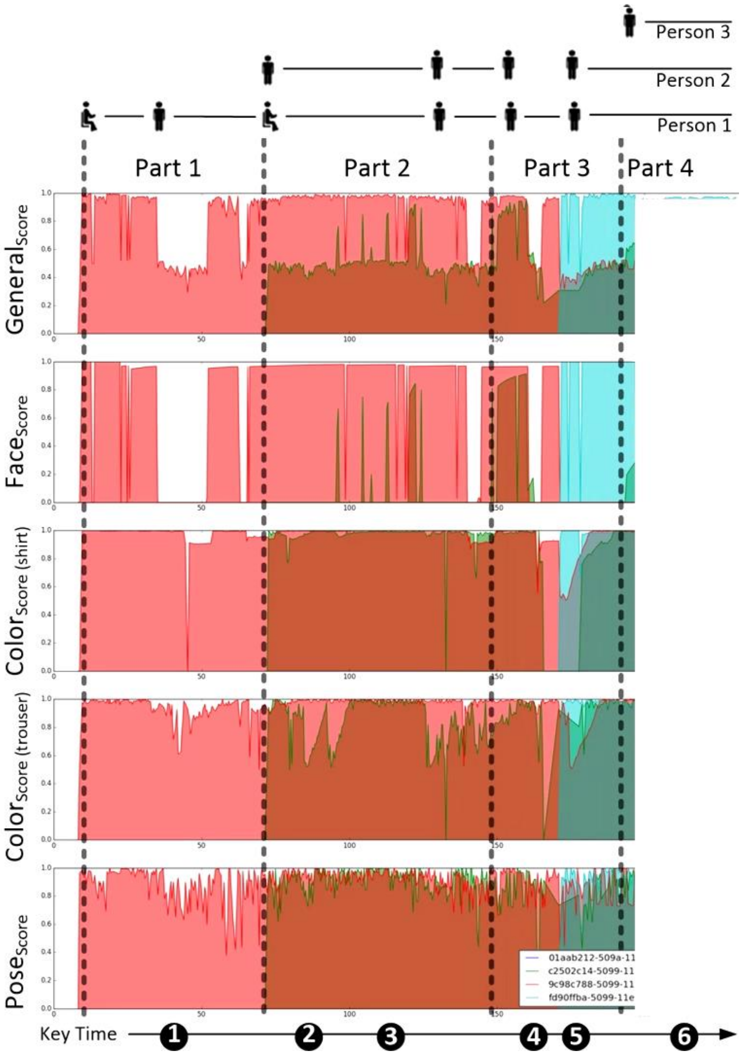
Results


Interact
Move Camera
Select
Focus Camera
Measure
2D Pose Estimate
2D Nav Goal
Publish Point
Image
Image
Time
ROS Time: 1553695638.93
ROS Elapsed: 3842.59
Wall Time: 1553695638.96
Wall Elapsed: 3842.55
Experimental
28 fps

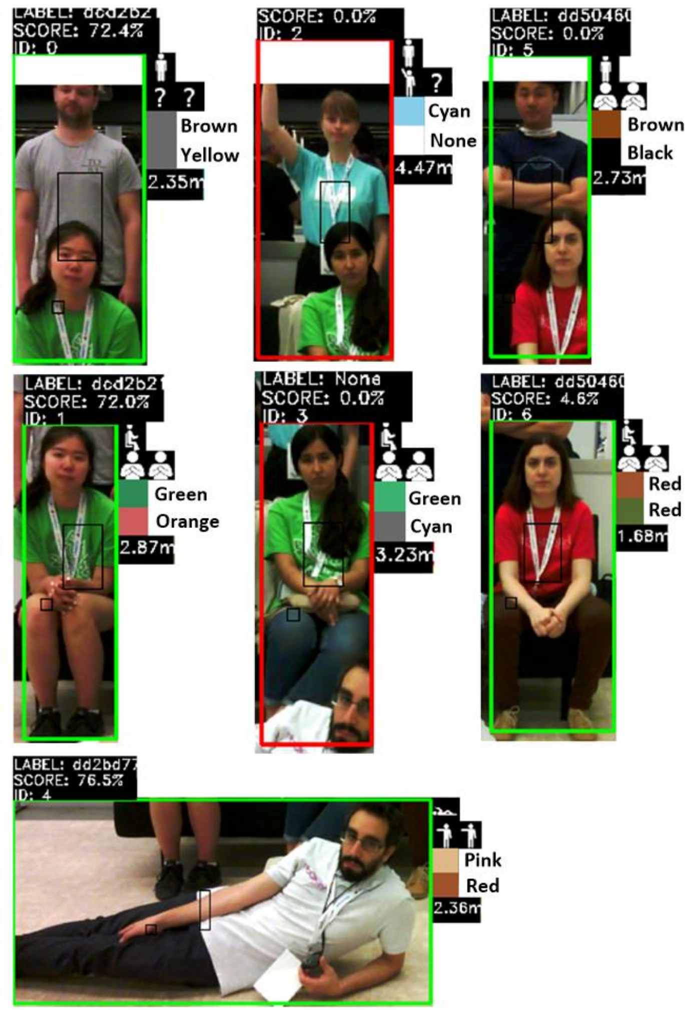

Person 3
Person 2
Person 1
Part 1
Part 2
Part 3
Part 4
General_score
Face_score
Color_score (shirt)
Color_score (trouser)
Pose_score
Key Time
1
2
3
4
5
6

Results




Person 3
Person 2
Person 1
Part 1
Part 2
Part 3
Part 4
General Score
Face Score
Color Score (shirt)
Color Score (trouser)
Pose Score
Key Time
1
2
3
4
5
6

RoboCup@Home 2018 results



Future Works

About People Management

- The dependence between people position and people height can be reduced through other features (e.g. age, gender,).
- A coming block, is the implementation of the work of our colleagues [11] to get 45 more features (e.g. Causal/Formal upper/lower clothes, carrying plastic bag, gender)
- Adjust weights of the scoring system of the tracker with reinforcement or adaptive learning technics
- Extend Kalman Filter approach with people speed estimation

More Generally

- Add link with Geographic Database
- Equivalent work through CNN (and LSTM Long / Short Term Memory)
- Add more Knowledge Representation and Reasoning to architectures

[11] Y. Chen, et al , Pedestrian attribute recognition with part-based CNN and combined feature representations. In VISAPP2018

Framework for Human Robot Social Interactions

Jacques Saraydaryan[1,2], Raphael Leber [1], Fabrice Jumel[1,2]
[1]CPE Lyon, France
[2]CITI Lab., INRIA Chroma

Presented by Fabrice Jumel

THANK YOU FOR YOUR ATTENTION

References

[1] L. Wenhan et al . Multiple object tracking: A review. CoRR, abs/1409.7618, 2014, last 2017.
[2] A. Bewley, et al . Simple online and realtime tracking. In 2016 IEEE International Conference on Image Processing (ICIP), pages 3464–3468, 2016
[3] B. Lavi, et al. Survey on deep learning techniques for person re-identification task. CoRR , abs/1807.05284, 2018
[4] J. Elfring, et al. , Semantic world modeling using probabilistic multiple hypothesis anchoring. ,Robot. Auton. Syst. , 61(2):95–105, February 2013
[5] K. Minkyu et al. , An architecture for person-following using active target search. CoRR, abs/1809.08793, 2018.
[6] Z. Cao, et al., Open-pose: Realtime multi-person 2d pose estimation using part affinity fields. CoRR, abs/1812.08008, 2018

[7] P. Pettit et al , Css color module level 3. W3c recommendation, W3C, 2018
[8] H. Kaiming et al. , Deep residual learning for image recognition. arXiv preprint arXiv:1512.03385 , 2015
[9] N. Dalal et al. , Histograms of oriented gradients for human detection. Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 2005
[10] P. Viola et al. , Rapid object detection using a boosted cascade of simple features. In Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001 , volume 1, pages I–I, 2001
[11] Y. Chen, et al , Pedestrian attribute recognition with part-based CNN and combined feature representations. In VISAPP2018