AI-Enhanced Computer Vision for Service Robots

Dr. Dimitrios Tzovaras

Director

Information Technologies Institute (ITI)

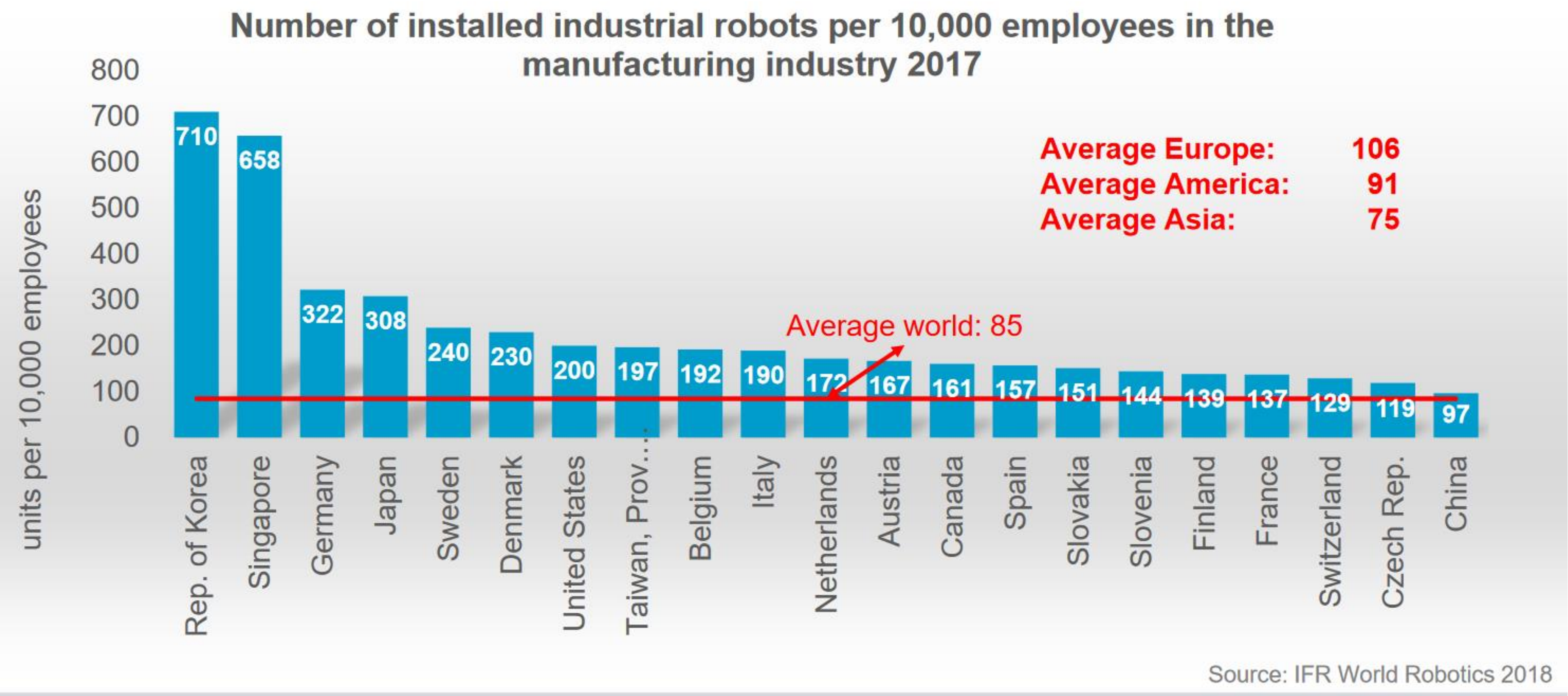
Centre for Research and Technology Hellas (CERTH)

Athens, August 2019

Robots now and in the future

the robot market according to the IFR[1]

Robot density rises Globally

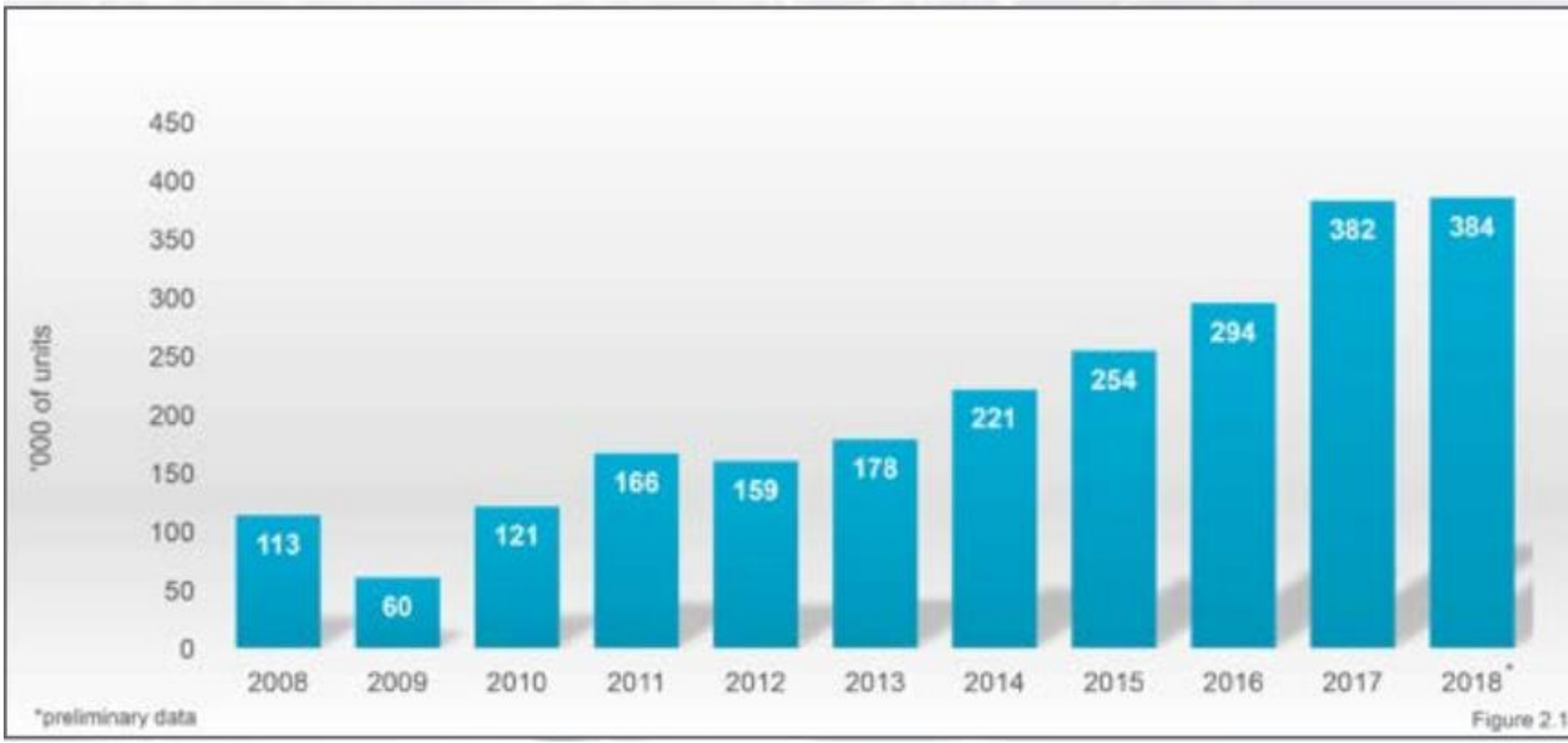


[1] International Federation of Robotics; https://ifr.org/

Robots now and in the future

the robot market according to the IFR[1]

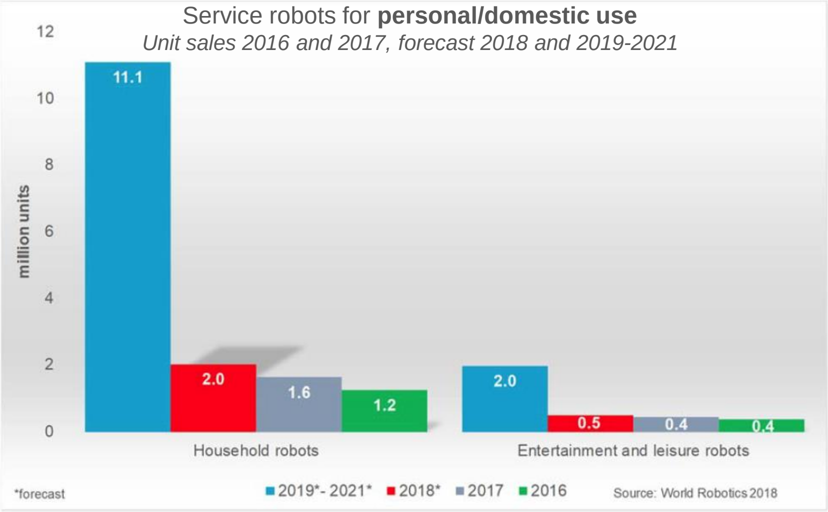
Estimated worldwide supply of industrial robots

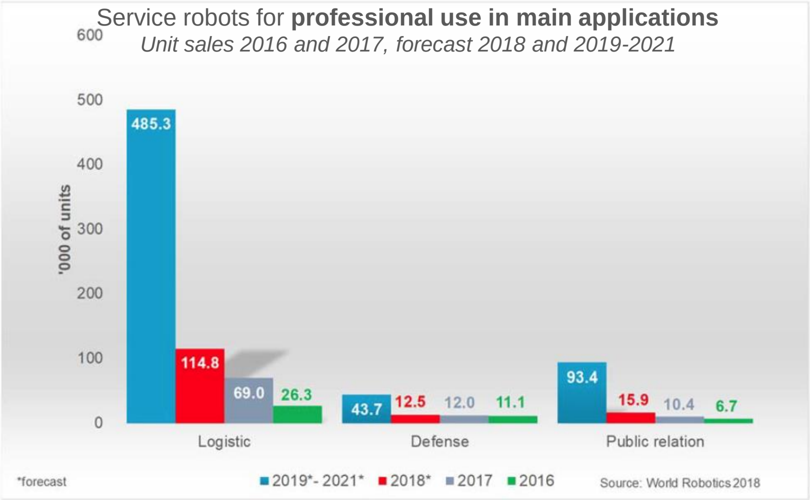


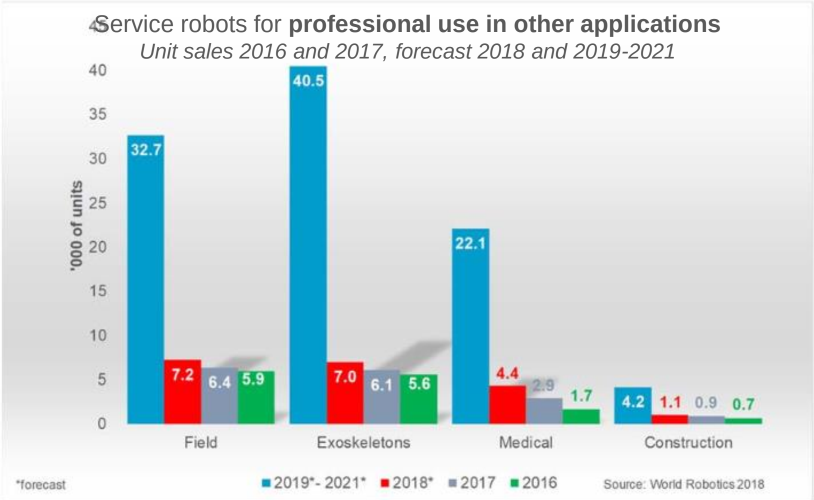
[1] International Federation of Robotics; https://ifr.org/

Robots now and in the future

the robot market according to the IFR[1]







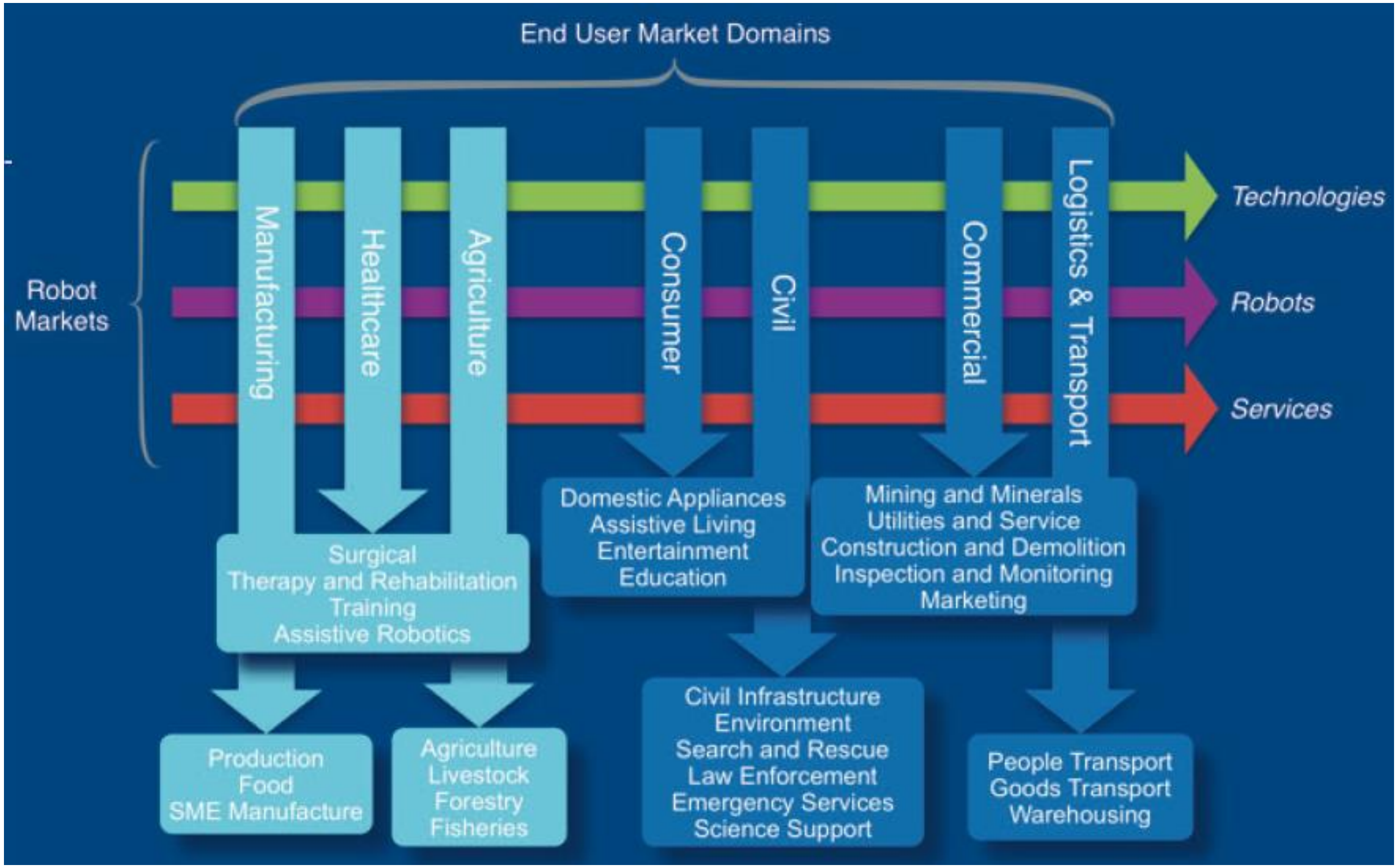
[1] International Federation of Robotics; https://ifr.org/

The robots market

…and the role of cross cutting technologies

Technologies that advance robot autonomy

are essential for most service robot applications



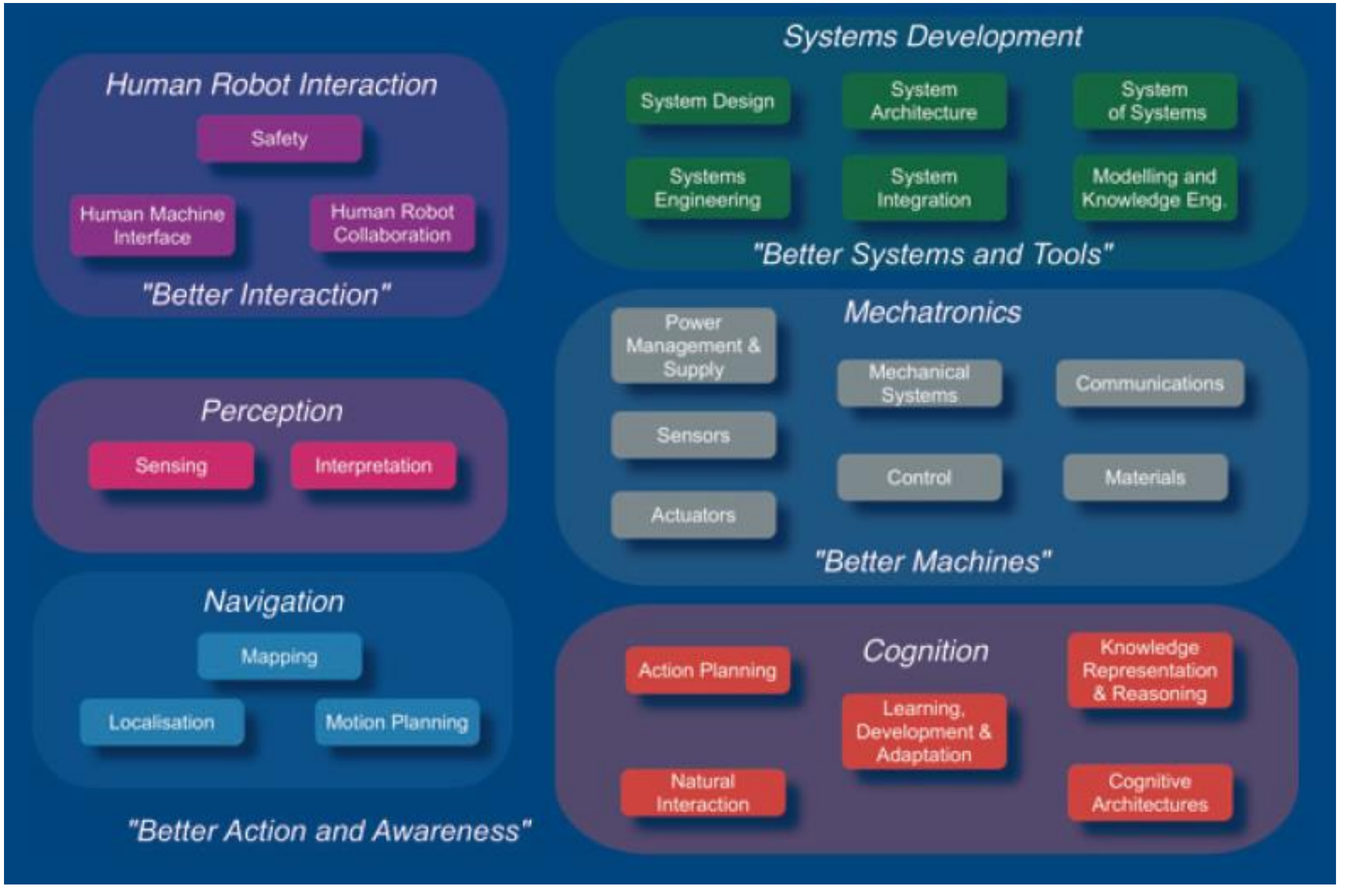
SPARC Robotics Multi Annual Roadmap (MAR)

The robots market

…and the role of cross cutting technologies

Better <u>Action</u> and <u>Awareness</u> is key to robot autonomy

***Perception**, **Cognition, Navigation** and **HRI** are essential abilities to this end*



SPARC Robotics Multi Annual Roadmap (MAR)

The robots market

…and the role of cross cutting technologies

Better Action and Awareness is key to robot autonomy

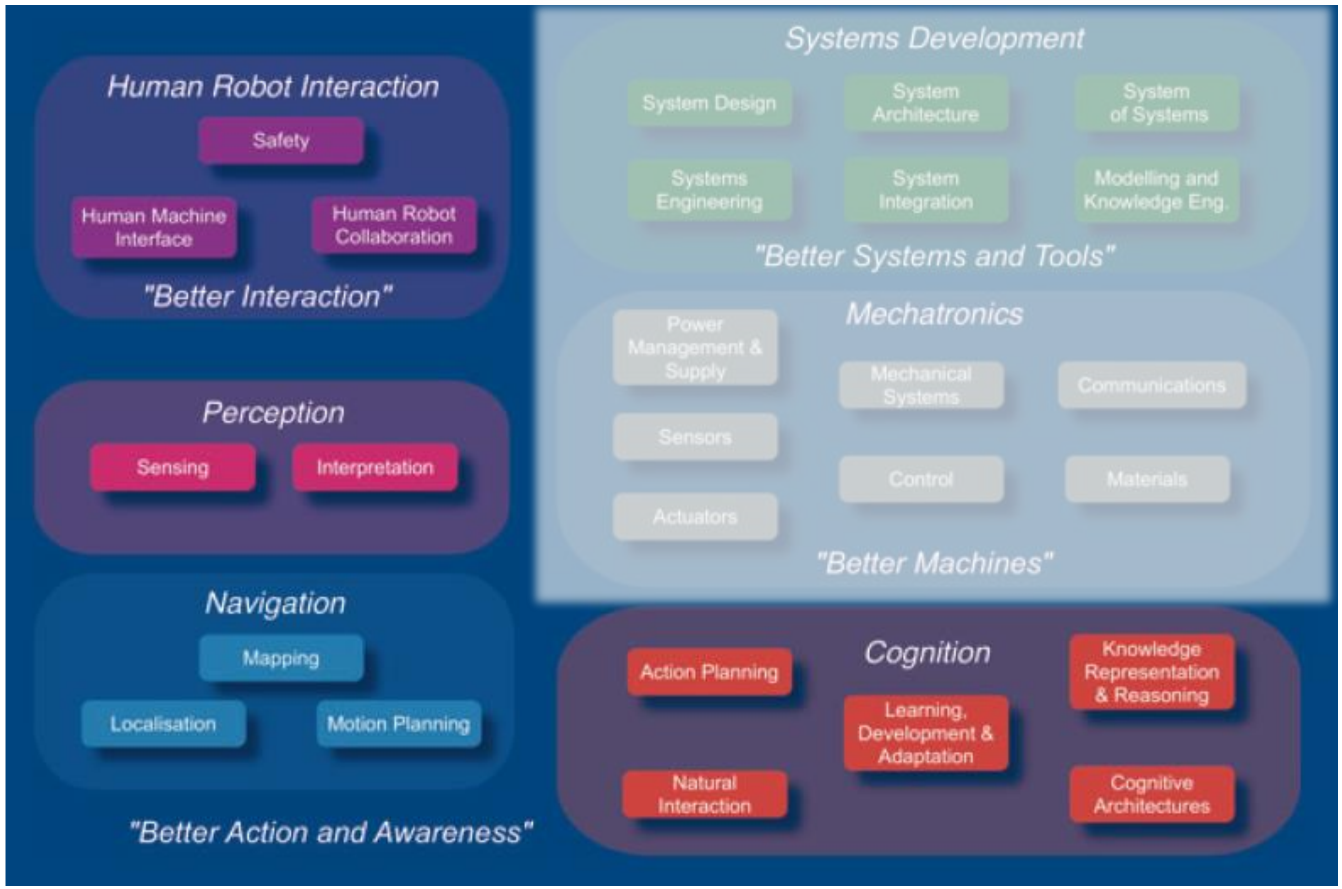
Perception, Cognition, Navigation *and* ***HRI*** *are essential abilities to this end*

We focus on AI-Enhanced Computer Vision for:

Sensing and interpretation of the environment

Mapping, localization and motion planning

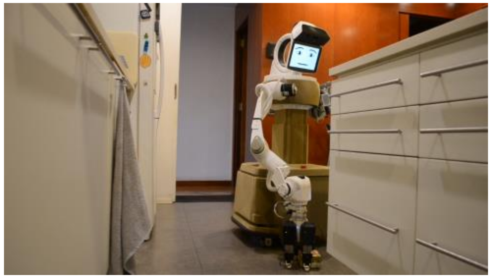
Knowledge representation and reasoning



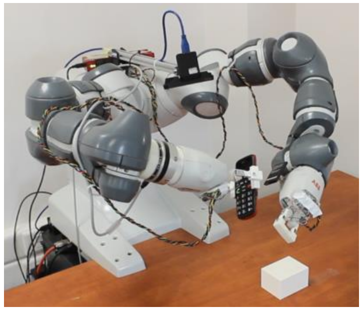
SPARC Robotics Multi Annual Roadmap (MAR)

Applications

- Personal service robots **@home**



- Professional service robots **@agile manufacturing**



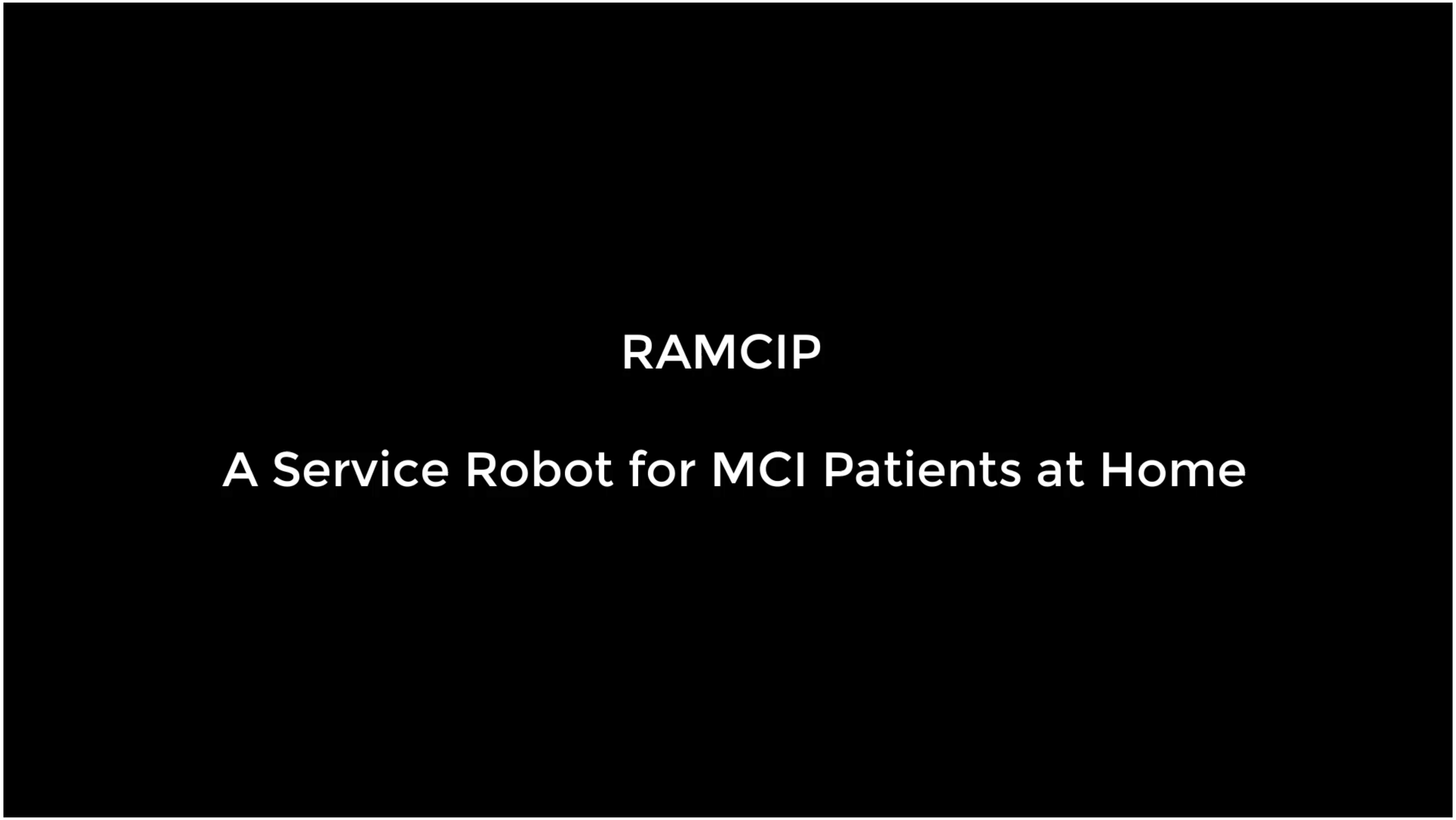
- Service robots **@field/construction sites**

Applications

- Personal service robots **@home**

Personal service robots @home

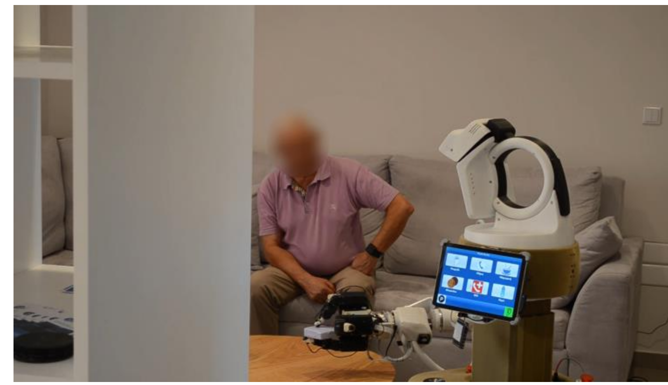
overview of the RAMCIP robot



Personal service robots @home
perceiving the home environment

- In terms of perception, an autonomous domestic assistive robot should
  - Know the home environment - **mapping**
  - Be capable to **recognize objects** and **estimate their pose**
    - Accurately enough to enable grasping
  - Be capable to **monitor human activity and understand behavior**
  - **Take decisions** on when and how to assist

*…so as to provide **autonomous, proactive assistance to the end user***



Mapping of indoor environments

Metric and semantic mapping

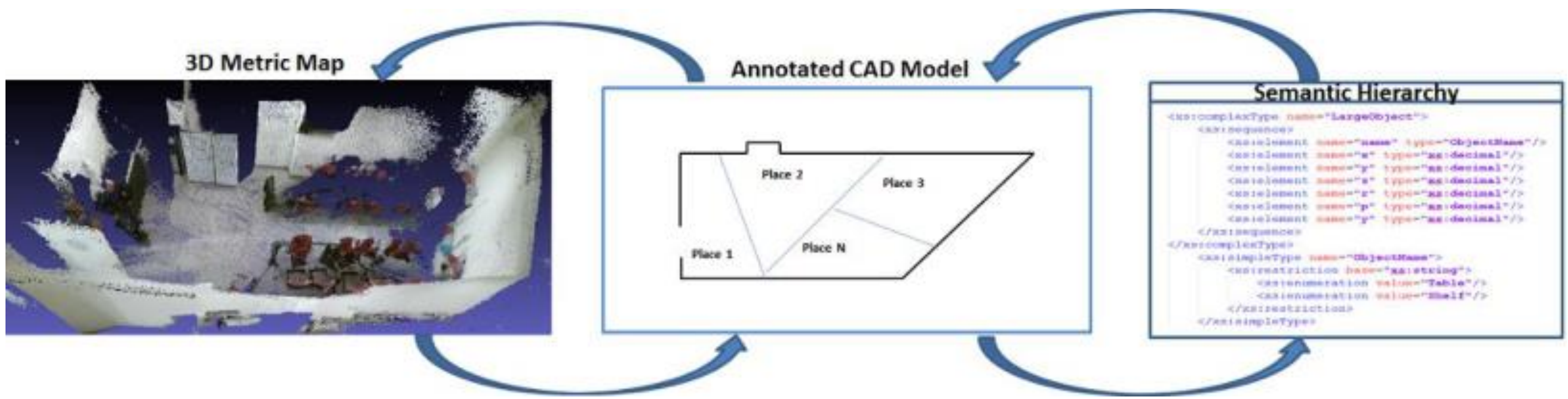
- Metric mapping
  - Employ Visual Odometry to construct a **topological map** [1]
    - Motion estimates by visual odometry and general graph optimization (g2o) for loop closure
    - New map node added according to geometric criterion (change in pose)
    - **Outcome: Dense map of the explored environment**

[1] Kostavelis, et al. "Learning spatially semantic representations for cognitive robot navigation." *Robotics and Autonomous Systems* 61.12 (2013): 1460-1475.

Mapping of indoor environments

Metric and semantic mapping

- Hierarchical modelling of the domestic space
  - Small (e.g. cup) and large (e.g. table) **objects semantics and relations**
  - Hierarchical map allowing updates (i.e. cup last found on **this table**)
  - Enabling the robot to search for needed object in the house

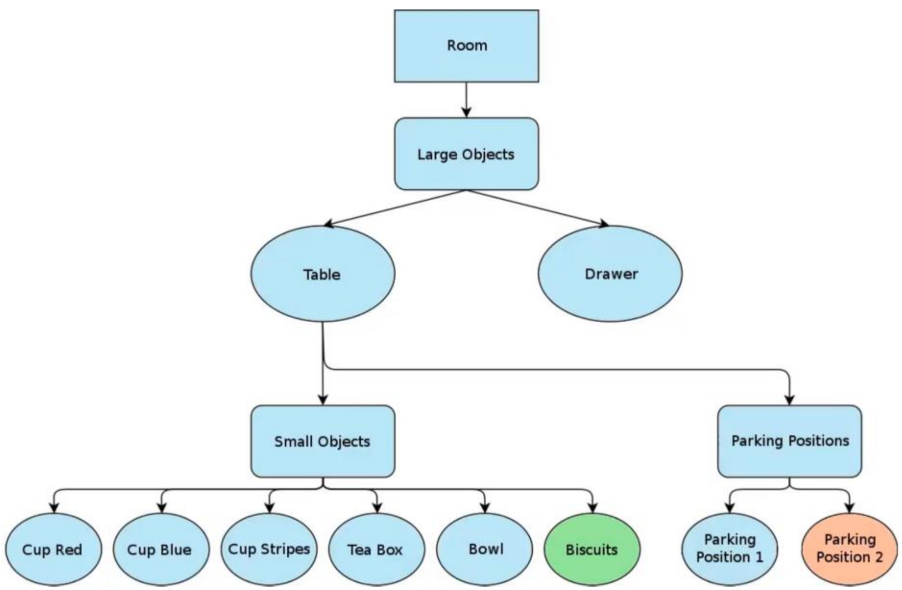


Information Technologies Institute

Mapping of indoor environments

Metric and semantic mapping

- Hierarchical modelling of the domestic space
  - Small (e.g. cup) and large (e.g. table) **objects semantics and relations**
  - Hierarchical map allowing updates (i.e. cup last found on **this table**)
  - Enabling the robot to search for needed object in the house




```xml
<?xml version="1.0" encoding="UTF-8" standalone="no" ?>
<pl:apartment
   xmlns:pl="http://tempuri.org/config/scene">
   <pl:room>
      <pl:id>0</pl:id>
      <pl:name>kitchen</pl:name>
      <pl:points>
         <pl:xcoord>100</pl:xcoord>
         <pl:ycoord>150</pl:ycoord>
      </pl:points>
      <pl:points>
         <pl:xcoord>200</pl:xcoord>
         <pl:ycoord>250</pl:ycoord>
      </pl:points>
      <pl:largeArticulated>
         <pl:id>600</pl:id>
         <pl:type>Fridge</pl:type>
         <pl:xcoord>125</pl:xcoord>
         <pl:ycoord>225</pl:ycoord>
         <pl:zcoord>1</pl:zcoord>
         <pl:roll>125</pl:roll>
         <pl:pitch>225</pl:pitch>
         <pl:yaw>1</pl:yaw>
         <pl:graspingPoints>1</pl:graspingPoints>
         <pl:small>
            <pl:id>1000</pl:id>
            <pl:type>medication</pl:type>
            <pl:xcoord>125</pl:xcoord>
            <pl:ycoord>225</pl:ycoord>
            <pl:zcoord>1</pl:zcoord>
            <pl:roll>125</pl:roll>
            <pl:pitch>225</pl:pitch>
            <pl:yaw>1</pl:yaw>
            <pl:properties>
               <pl:color>white</pl:color>
            </pl:properties>
            <pl:relationWithLarge>2</pl:relationWithLarge>
            <pl:parent>600</pl:parent>
            <pl:graspingPoints>1</pl:graspingPoints>
            <pl:defaultPosition>600</pl:defaultPosition>
         </pl:small>
         <pl:parkingpos>
            <pl:xcoord>2.29</pl:xcoord>
            <pl:ycoord>-2.55</pl:ycoord>
            <pl:zcoord>0</pl:zcoord>
            <pl:roll>0</pl:roll>
            <pl:pitch>0</pl:pitch>
            <pl:yaw>0.435</pl:yaw>
         </pl:parkingpos>
      </pl:largeArticulated>
   </pl:room>
</pl:apartment>
```

Mapping of indoor environments

Metric and semantic mapping

- Hierarchical modelling of the domestic space

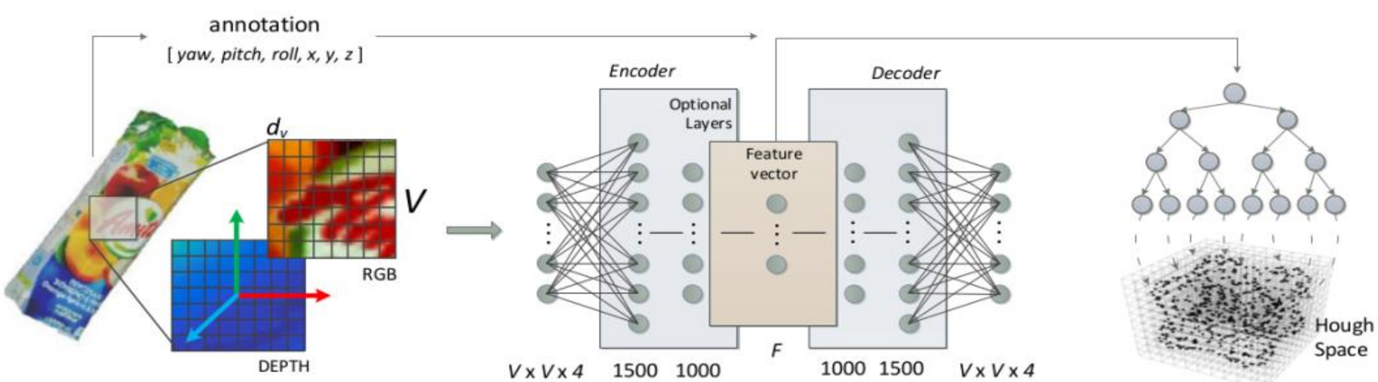
Request the "Biscuits" position from the Hierarchical Map

Object recognition

Object recognition and 6DoF pose estimation

- Steps of proposed method [1]

  1. **Scene Segmentation**
     - Planar supporting surface segmentation (Random Sample Consensus - RANSAC)
  2. **Training & Hypotheses Generation**
     - 2.5D Patch extraction
     - Feature Learning using **sparse autoencoders**
     - Hough Forests classifier – Hough Voting

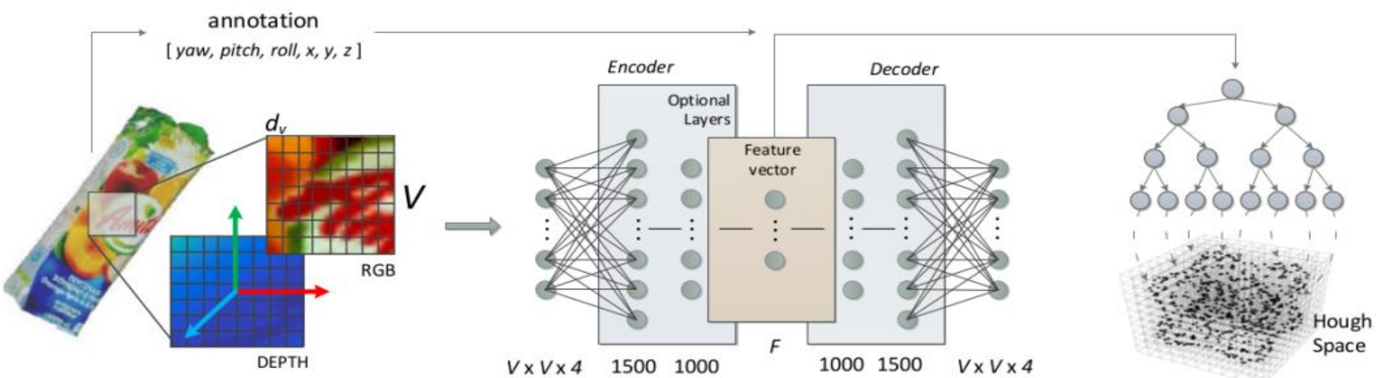


[1] A. Doumanoglou, R. Kouskouridas, S. Malassiotis and T. K. Kim, 6D Object Detection and Next-Best-View Prediction in the Crowd, IEEE Proceeding of Computer Vision and Pattern Recognition **(CVPR), 2016**.

Object recognition

Object recognition and 6DoF pose estimation

- Steps of proposed method [1]
  1. Scene Segmentation
  2. Training & Hypotheses Generation
  3. **Hypotheses joint optimization - refinement**
     - Depth and RGB similarity between original scene and objects rendered in the scene
     - Pose correction based on planar model coefficients



[1] A. Doumanoglou, R. Kouskouridas, S. Malassiotis and T. K. Kim, 6D Object Detection and Next-Best-View Prediction in the Crowd, IEEE Proceeding of Computer Vision and Pattern Recognition **(CVPR), 2016**.

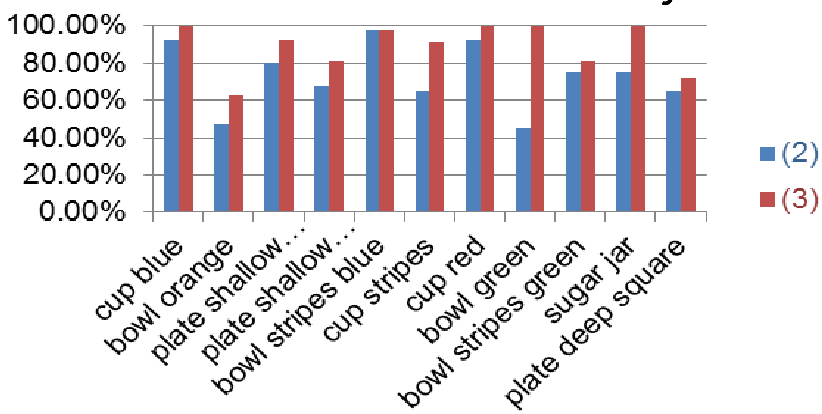
Information Technologies Institute

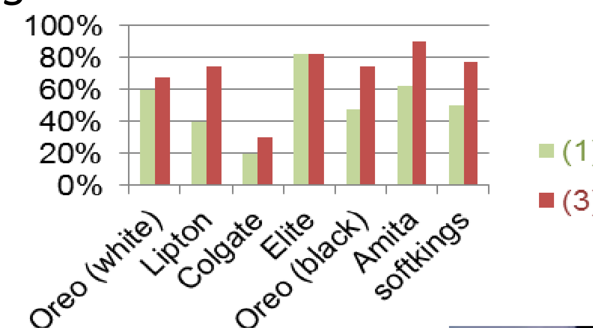
Object recognition

Object recognition and 6DoF pose estimation

Experimental Results (Comparison with State of Art algorithms)

Object recognition

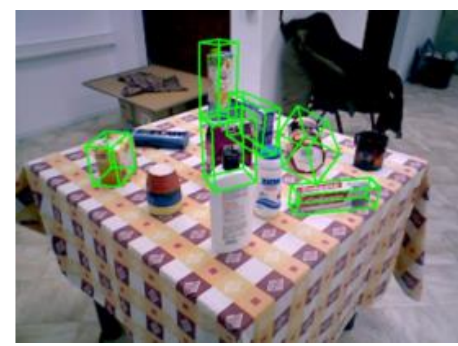




Pose Estimation Accuracy (towards grasping)

(1)	0.45 – 1.92 cm
(2)	0.59 – 2.1 cm
(3)	0.27 – 0.98 cm

Range of surface-to-surface distance between detected and ground truth object

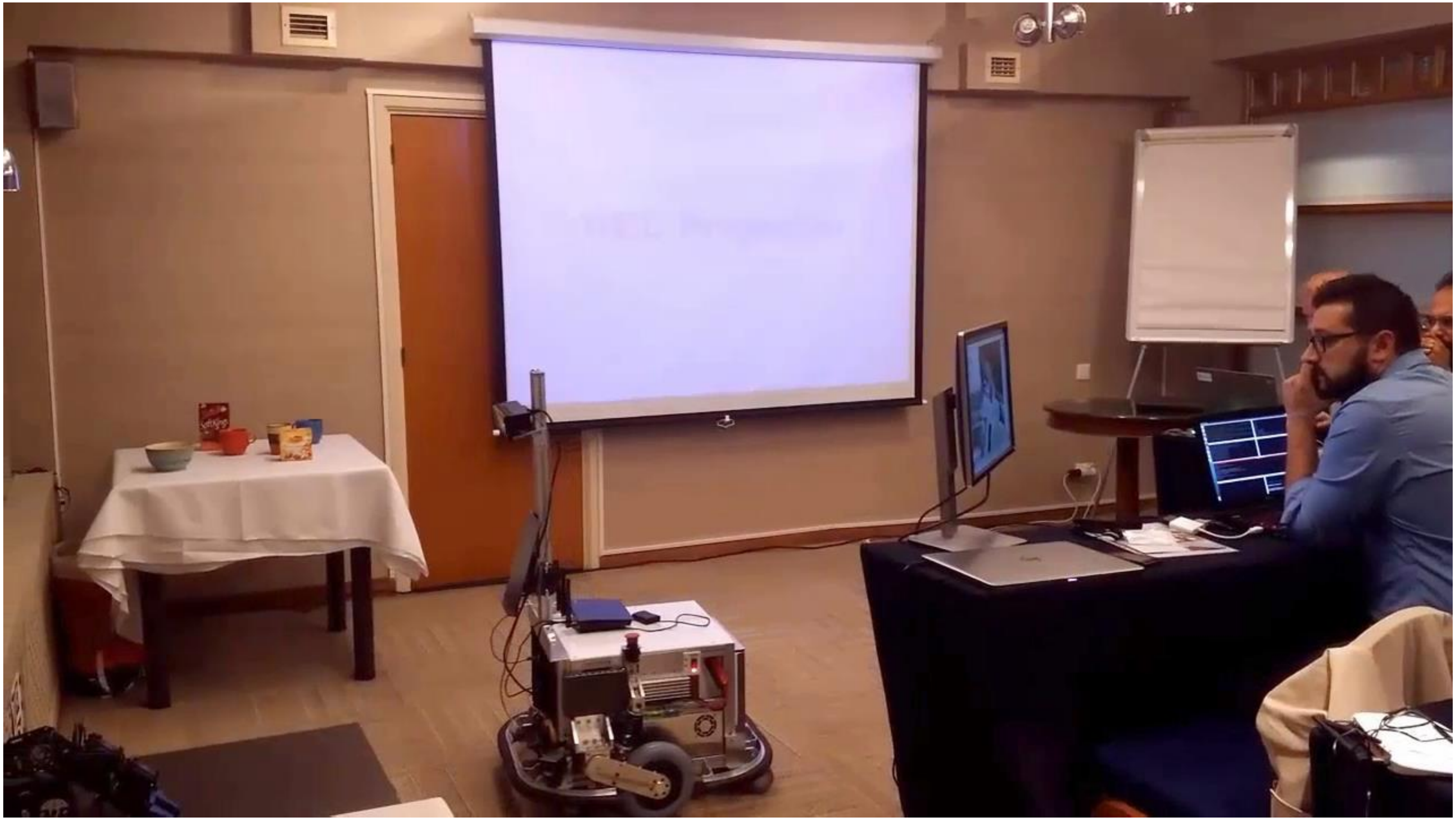


1) **Berkley Textured Object Recognition Algorithm (textured objects)**
2) **LINEMOD pipeline (non-textured objects)**
3) **6D Object Detection and Next-Best-View Prediction in the Crowd (textured + non-textured objects)**

[1] A. Doumanoglou, R. Kouskouridas, S. Malassiotis and T. K. Kim, 6D Object Detection and Next-Best-View Prediction in the Crowd, IEEE Proceeding of Computer Vision and Pattern Recognition **(CVPR), 2016**.

Object recognition

early experimentation with mobile robot platform



Object recognition

Object recognition and 6DoF pose estimation coupled with grasping in real homes

Large domestic objects recognition

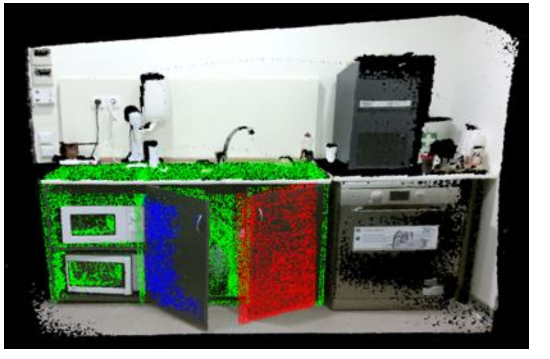
large articulated objects pose estimation

- Rough alignment of object model
  - Based on **robot location estimate** and **known environment map**
- Approach based on Articulated ICP (AICP) [1] applied to the model
- Object registration in the robot's perceived scene
  - **Robot localization** refinement
  - **Object position estimate** refinement
- Object state identification (closed, open - degrees)

First iteration of AICP

Cabinet base - aligned

Cabinet doors - aligned



[1] S. Pellegrini et. al. "A Generalisation of the ICP Algorithm for Articulated Bodies." In *BMVC*, vol. 3, p. 4. 2008.

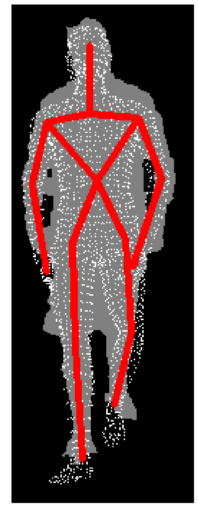
Information Technologies Institute

Large domestic objects recognition
coupled with grasping in real homes

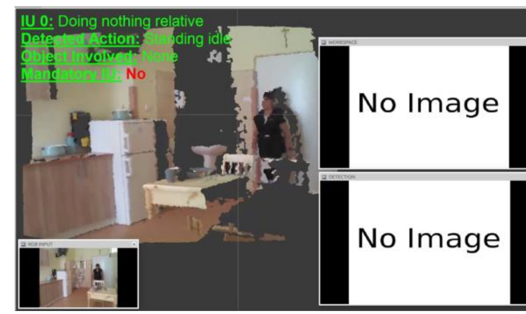
Human activity monitoring

for domestic service robot applications

- Human Tracking



- Human Activity Recognition and Behavior analysis

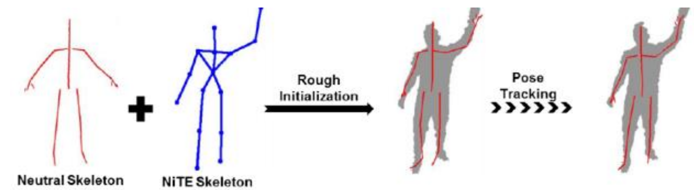


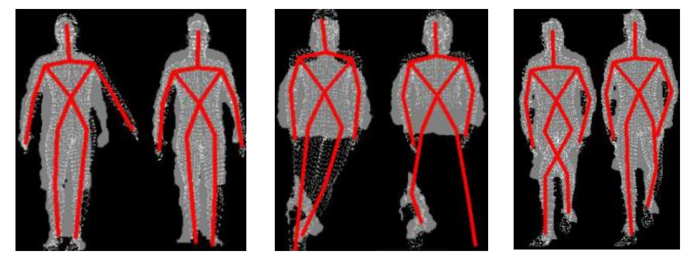
- Emotion Recognition

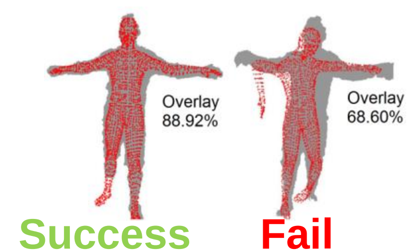
Human tracking

approach overview

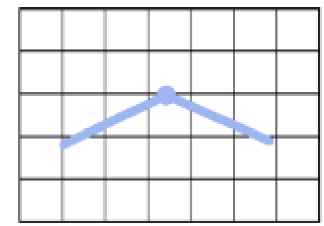
- **Proposed Human Pose Tracking method [1]**
  - Full body **model –based** pose tracking
    - **Initialization** from SoA single-shot discriminative method [2]
    - **Model-based tracking** building upon the “Dense Articulated Real Time Tracking” (DART) approach [3]
      - Optimization to minimize sum of distances between 3D points of the observation and the template
    - **Extensions** to improve tracking optimization
      - **Free space violation**
      - **Body part visibility**
      - **Leg intersection**
      - **Object interaction**





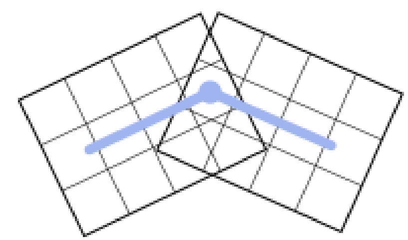


[1] M. Vasileiadis, S. Malassiotis, D. Giakoumis, C.S. Bouganis, D. Tzovaras, "Robust Human Pose Tracking For Realistic Service Robot Applications", 5th Int’l Workshop on Assistive Computer Vision and Robotics - **ACVR ‘17 of IEEE ICCV 2017**.
[2] Shotton, Al.., 2013. Real-time human pose recognition in parts from single depth images. Communications of the ACM, 56(1), pp.116-124.
[3] Schmidt, T, etAl, 2014, July. DART: Dense Articulated Real-Time Tracking. In Robotics: Science and Systems (Vol. 2, No. 1).



Model Representation

- Each rigid body part i forms a geometry defined implicitly by its Signed Distance Function: $SDF^i(\boldsymbol{x}, \boldsymbol{\theta}): R^3 \to R$
- Global $SDF_{mod}(\boldsymbol{x}, \boldsymbol{\theta}) : R^3 \to R$ is approximated by the composition of pre-computed local $SDF^i(\boldsymbol{x}, \boldsymbol{\theta})$



Optimization

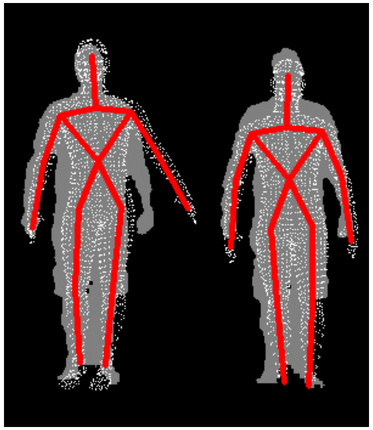
- Quasi-Newton optimization: Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm
- Can minimize any general real-valued function $f(x)$
- Approximates the Hessian matrix using simple rank-one updates specified by gradient evaluations

Human tracking

features introduced to improve performance

Free Space Violation

- Parts of the human template not corresponding to input data
- Deformed template projected on 2D-SDF depth image
- Free-space error $SDF_{fs}(\boldsymbol{\theta})$, defined as sum of values of the corresponding pixels on the 2D-SDF image



$$SDF_{overall}(\boldsymbol{\theta}) = SDF_{model}(\boldsymbol{\theta}) + \lambda SDF_{fs}(\boldsymbol{\theta}), \qquad \lambda \leq 1$$

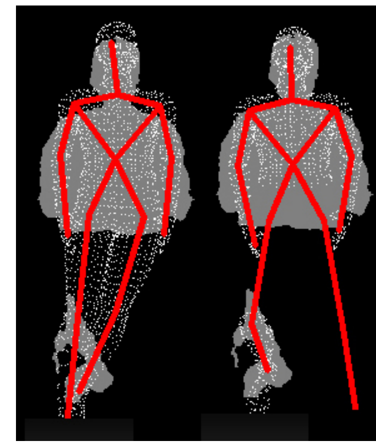
- Faster convergence, fewer iterations

Human tracking

features introduced to improve performance

Body Part Visibility

- Parts of the human template outside of the camera's FoV / occluded
- Template projected on image plane
- Part visibility determined by validity of data around limb midpoint / endpoint
- Non-visible body parts *i* are not taken into consideration during the optimization

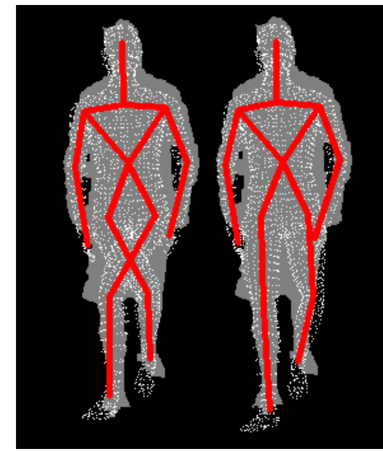


$$SDF^{i}(\boldsymbol{x}, \boldsymbol{\theta}) = \mathbf{0} \quad \boldsymbol{x} \in \Omega_i, \qquad \boldsymbol{\Delta q}_i = 0$$

- Pre-optimization

Leg Intersection

- Mix-up of the lower limbs, noisy observations between the legs, quick turn-arounds “trap” optimizer
- Each lower body part is approximated by 7 spheres s ($\boldsymbol{c}_s$, r_s)
- Leg intersection error based on sphere intersection



$$E_{intr}(\boldsymbol{\theta}) = \sum_{(s,t)\in P} \frac{1}{1 + e^{-(r_s + r_t - |\boldsymbol{c}_s(\theta) - \boldsymbol{c}_t(\theta)|)\gamma}}$$

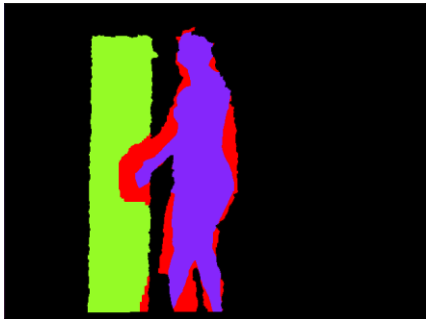
- Post-optimization. If error over threshold, recalculate for: a) R/L knees interchanged, b) R/L ankles interchanged

Human tracking

features introduced to improve performance

- **Human interaction with objects** is common in realistic settings
  - Can **severely affect** human tracking accuracy
  - *The optimizer tries to match the human template to both the human and non-human data*

- ***Implemented solution:*** Input preprocessing to remove such objects before optimization
  - Last tracked **human silhouette** along with a small **buffer zone** projected on new input
  - **Large non-overlapping areas** are considered as candidate objects
  - **Floodfill seeds** from center of each candidate area to remove smooth surfaces (doors, tables etc.)

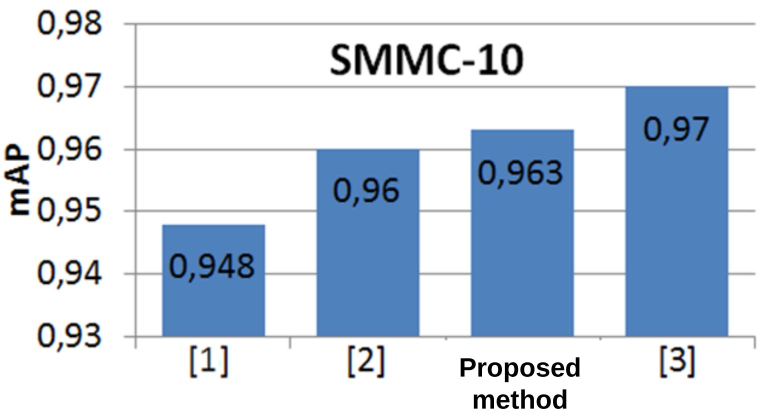


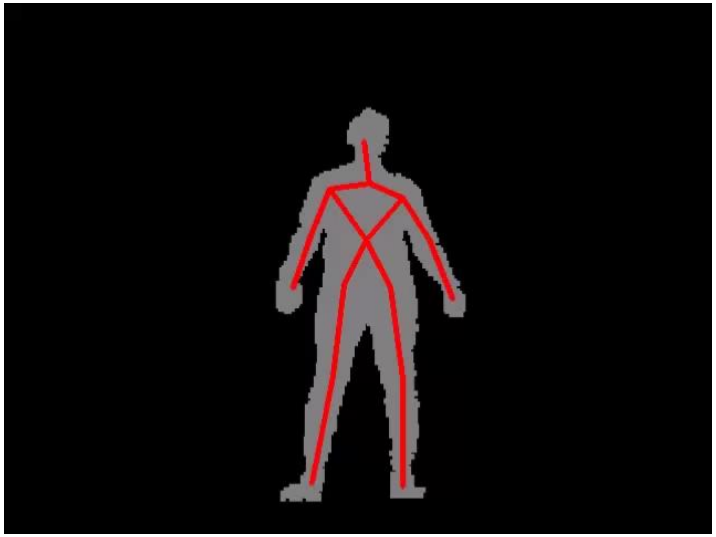
Human tracking

experimental results – public datasets

SMMC-10 Dataset

- **Sensor:** Mesa SwissRanger time-of-fight sensor
- **1 subject**, 28 sequences, front facing actions
- **Actions:** Waving, clapping, pointing, boxing, throwing, sitting, leg raising, kicking
- **Ground truth** from Vicon motion capture system
- **14 joints:** head, torso, R/L shoulder, R/L elbow, R/L hand, R/L hip, R/L knee, R/L foot.
- 8K point clouds, each one containing 25K points





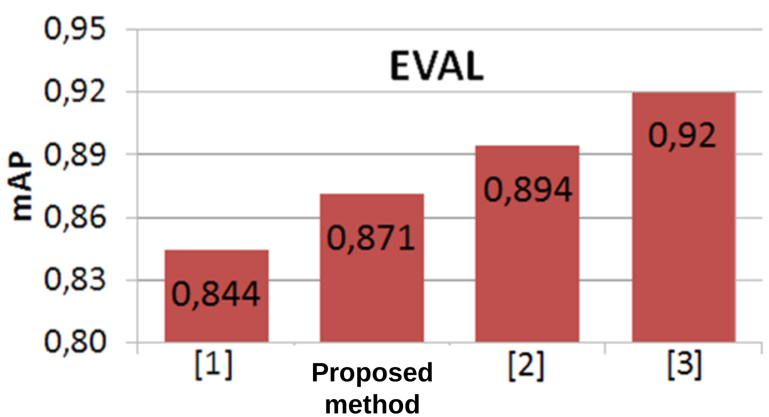
[1] Shotton et al. *Real-time human pose recognition in parts from single depth images*
[2] Ding & Fan. *Articulated gaussian kernel correlation for human pose estimation*
[3] Ye & Yang. *Real-time simultaneous pose and shape estimation for articulated objects*

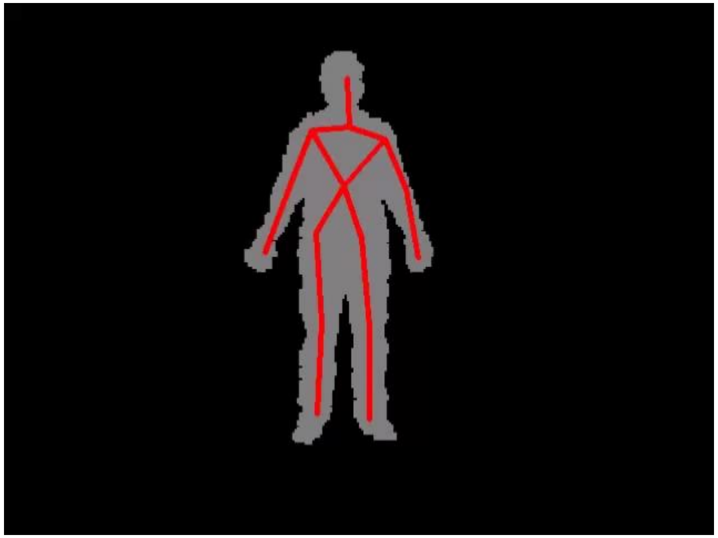
Human tracking

experimental results – public datasets

EVAL Dataset

- **Sensor:** Kinect v1 RGB-D camera
- **3 subjects**, 24 sequences, front facing actions
- **Actions:** Waving, clapping, boxing, bending, sitting, kicking, handstand, backflip
- **Ground truth** from Vicon motion capture system
- **12 joints:** head, neck, R/L shoulder, R/L elbow, R/L hand, R/L knee, R/L foot.
- 9K point clouds, each one containing 78K points



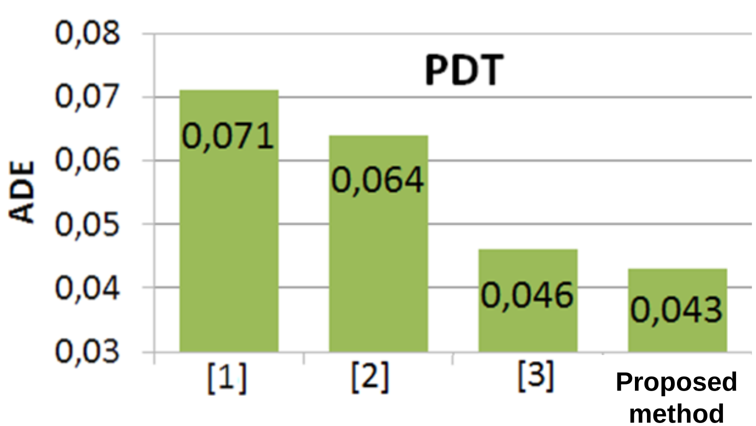


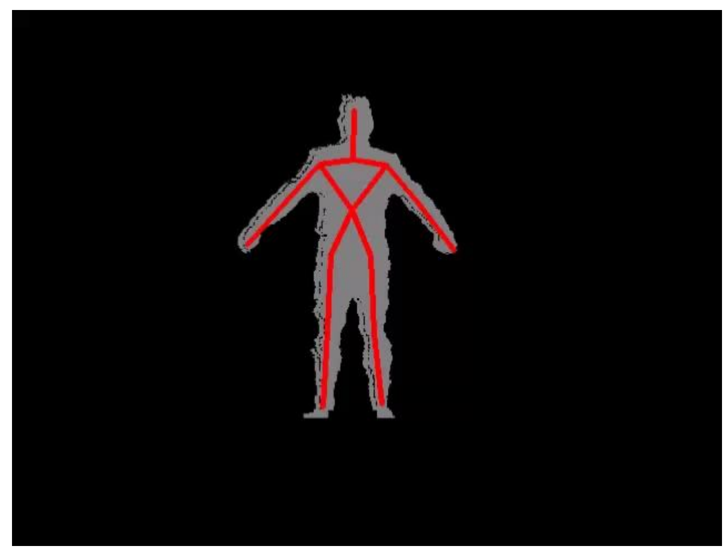
[1] Ganapathi, et al. *Real Time Human Pose Tracking from Range Data*
[2] Schmidt et al. *Dart: dense articulated real-time tracking with consumer depth cameras*
[3] Ye & Yang. *Real-time simultaneous pose and shape estimation for articulated objects*

Human tracking
experimental results – public datasets

PDT Dataset

- **Sensor:** Kinect v1 RGB-D camera
- **5 subjects**, 20 sequences, front facing actions
- **Actions:** Waving, boxing, bending, kicking, jumping, sitting on floor, moving around
- **Ground truth** from Phasespace motion capture system
- **15 joints:** head, neck, R/L shoulder, R/L elbow, R/L hand, Torso, R/L Hip, R/L knee, R/L foot
- 27K depth images, 640x480px





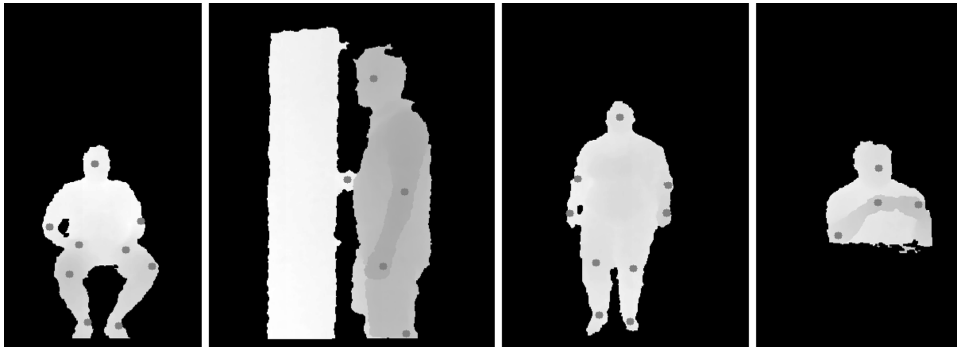
[1] Baak et al. *A data-driven approach for real-time full body pose reconstruction*
[2] Helten et al. *Personalization and evaluation of a real-time depth-based full body tracker*
[3] Ye & Yang. *Real-time simultaneous pose and shape estimation for articulated objects*

Human tracking

experimental results – realistic dataset

Custom Dataset Generation

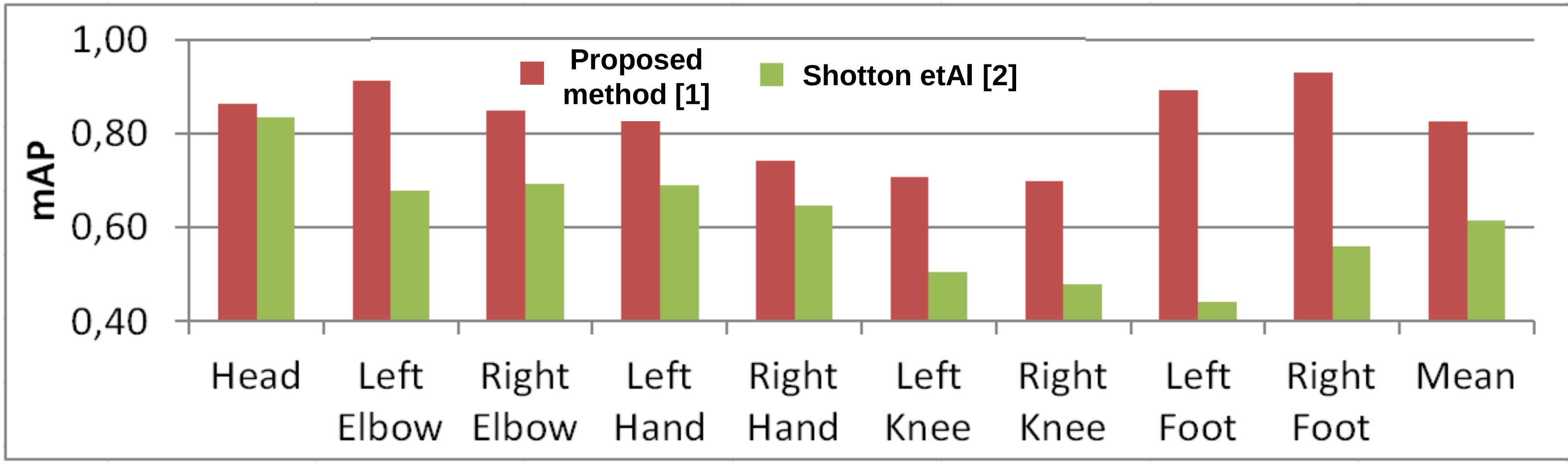
- Realistic human motion dataset for service robot AAL applications
- Kinect v1 depth camera, on-board of a service robot
- ~90s sequences, 11 subjects
- Actions of typical **activities of daily living relevant to AAL**
  - walking, eating, drinking, opening cupboard, taking pill, etc.
- Manual annotation of 9 skeleton joints

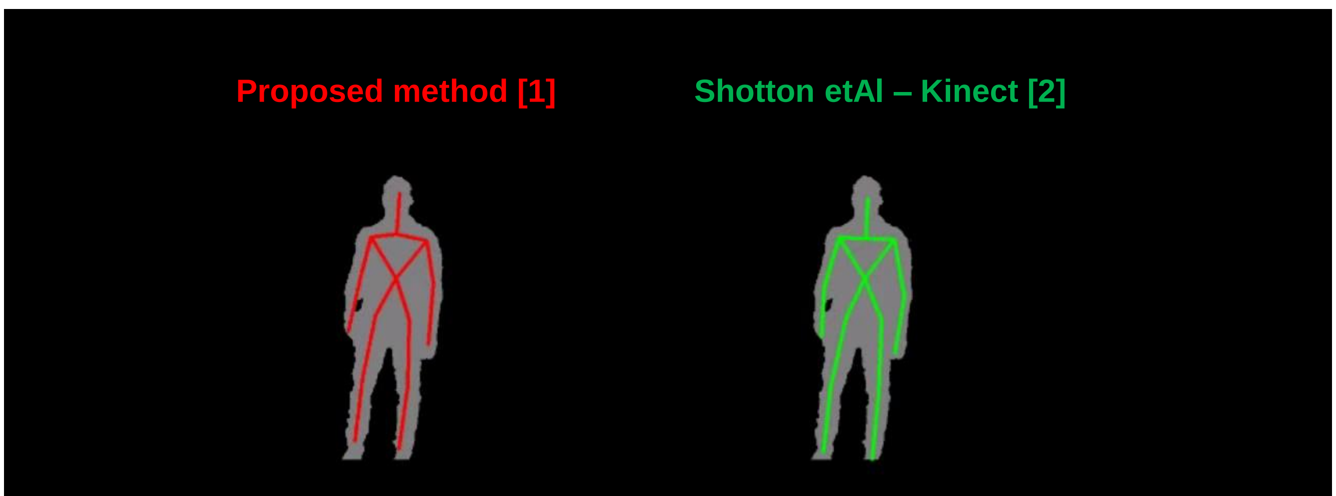


available online: http://ramcip-project.eu/ramcip-data-mng

Human tracking

experimental results – realistic dataset



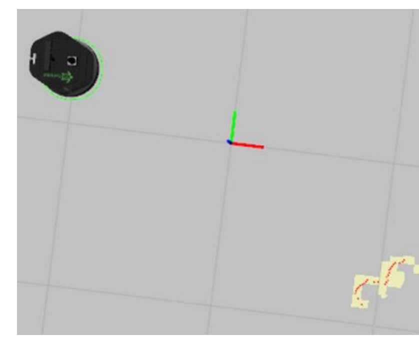


[1] M. Vasileiadis, S. Malassiotis, D. Giakoumis, C.S. Bouganis, D. Tzovaras, "Robust Human Pose Tracking For Realistic Service Robot Applications", 5th Int'l Workshop on Assistive Computer Vision and Robotics - **ACVR '17 of IEEE ICCV 2017**.

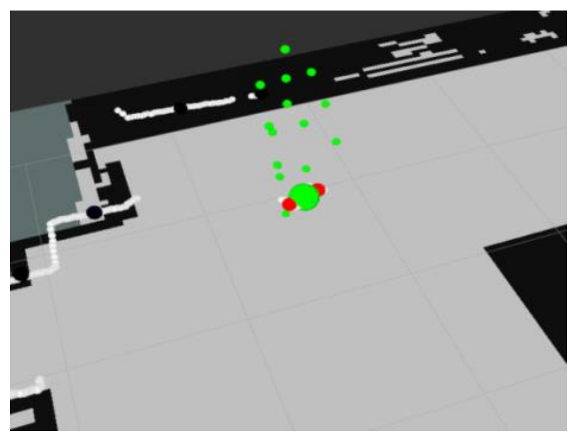
[2] Shotton et al. *Real-time human pose recognition in parts from single depth images*

Human tracking by RGBD and lasers fusion
enabling social-aware robot navigation

Overview of proposed method [1]



- **Modality 1**: Skeleton joints tracker - RGBD sensor
  - Robot tracks user while in RGBD sensor's FoV
- **Modality 2:** Leg-based human detector through LIDAR sensors of increased FoV
  - User position tracked even out of RGBD FoV
- **Fusion** b/w Modality 1 and Modality 2
- **Social-aware** adaptation of robot path planner…



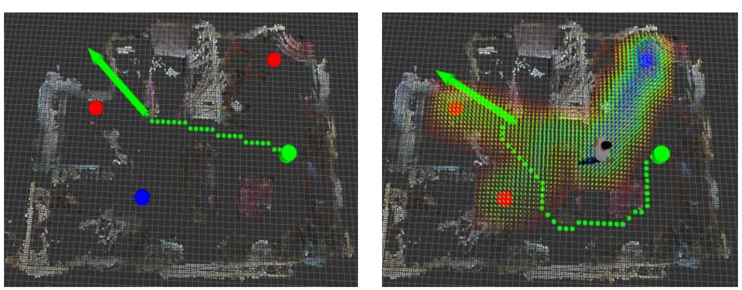
[1] Kostavelis, I., Kargakos, A., Giakoumis, D. and Tzovaras, D., 2017, July. Robot's Workspace Enhancement with Dynamic Human Presence for Socially-Aware Navigation. In International Conference on Computer Vision Systems (ICVS 2017), pp. 279-288. Springer, Cham. ***Best Conference Paper Award***

Human tracking by RGBD and lasers fusion

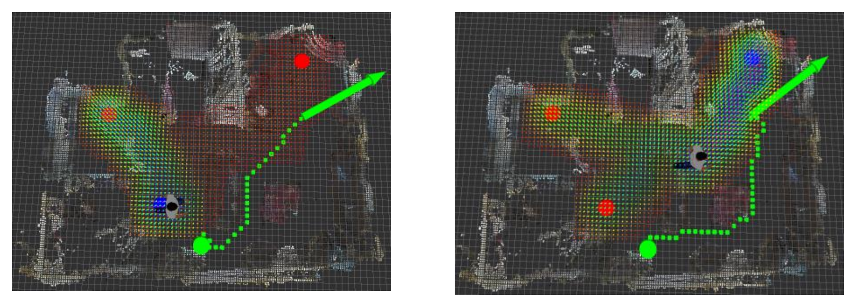
enabling social-aware robot navigation

- *The social aware robot navigation method* ***models****:*
  - **Human presence** with a sequence of Gaussian Kernels parameterized to the proxemics theory
  - The human's **short term motion intention** based on geometric criterions
    - Considering the current human pose with respect to the candidate human standing positions (frequently visited ones)

- *It* ***calculates*** *in real-time:*
  - The **robot global path** using a variation of D* Lite algorithm
  - The required robot **re-planned path on run-time** to avoid unintentional collisions and crossings with the human paths



Robot path planning: ***Left***, without considering human presence
Right, by considering human presence

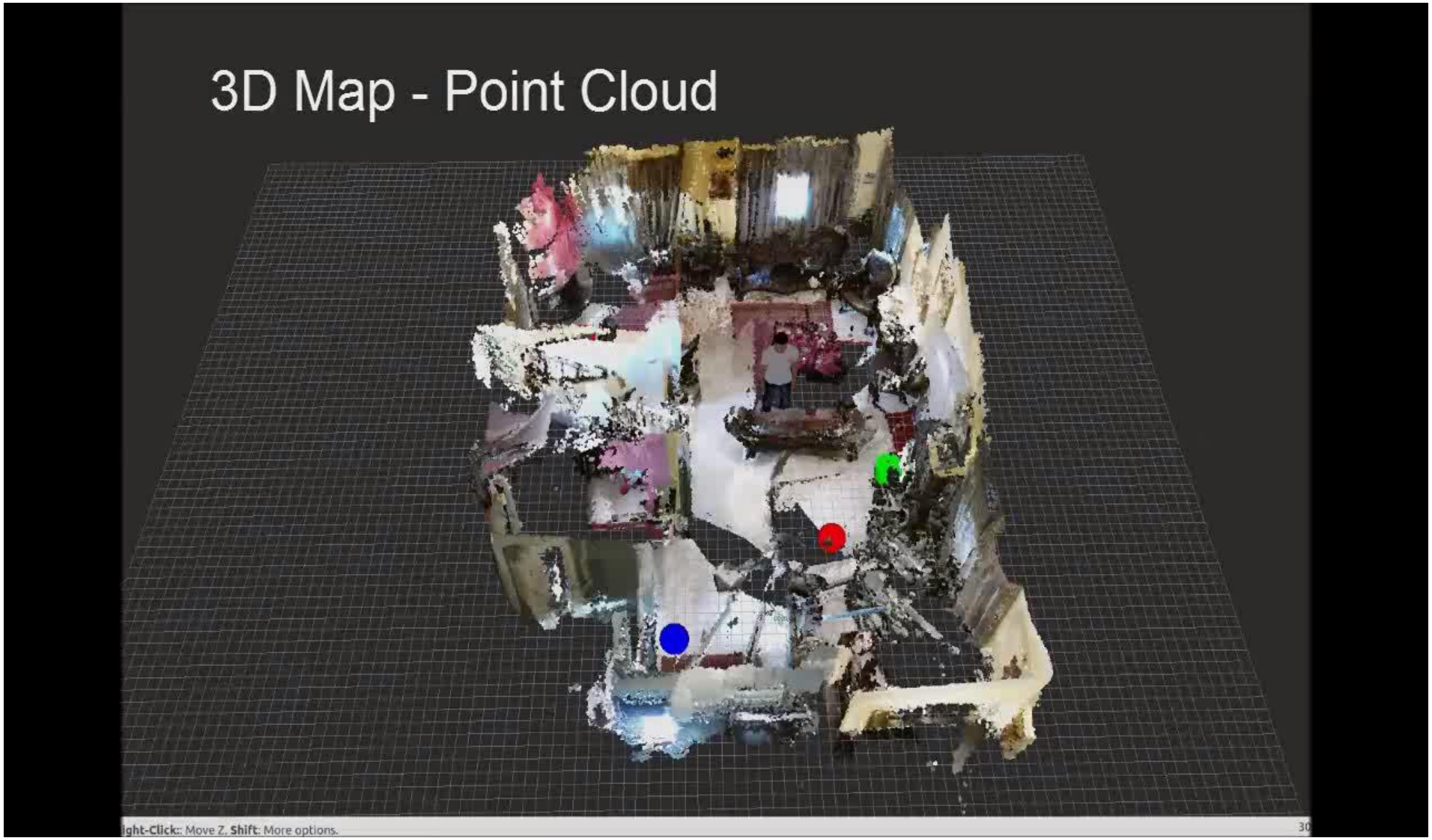


Robot path planning: ***Left***, without robot-human path intersection
Right, with robot-human path intersection

[1] Kostavelis, I., Kargakos, A., Giakoumis, D. and Tzovaras, D., 2017, July. Robot's Workspace Enhancement with Dynamic Human Presence for Socially-Aware Navigation. In International Conference on Computer Vision Systems (ICVS 2017) , pp. 279-288. Springer, Cham. ***Best Conference Paper Award***

Human tracking by RGBD and lasers fusion

short term human motion intention prediction and adaptive robot path planning



[1] Kostavelis, I., Kargakos, A., Giakoumis, D. and Tzovaras, D., 2017, July. Robot's Workspace Enhancement with Dynamic Human Presence for Socially-Aware Navigation. In International Conference on Computer Vision Systems (ICVS 2017) , pp. 279-288. Springer, Cham. ***Best Conference Paper Award***

Emotion recognition
based on computer vision and biosignals

Facial expressions recognition

- Face detection, facial landmark detection, face alignment and cropping
- Local Gabor Binary patterns extraction
  - Local Gabor Binary Pattern Histograms (LGBPH)
    - 18 Gabor channels (3 octaves, 6 orientations)
  - Three facial feature extractors
    - Operating on input image subsets (upper, lower, global)
- Dimensionality reduction (PCA)
- SVM classifier for each Action Unit (AU)
  - Trained on the extended Cohn Kanade database

Emotion recognition
based on computer vision and biosignals

Affect-related body activity analysis

- Depth-based **upper-body activity** tracking
  - Extraction of low-level postural features, high-level features and temporal dynamics
    - E.g. hands distance, body activity movement/power, body spatial expansion, symmetry, bending and statistical cues (mean, SD)
  - **Stress-oriented behavioural body activity** features
    - Activity level, sharp activities energy, activity symmetry, position and movement of head, body barycenters, specific gestures [1]
- **Biosignals** processing
  - E.g. Empatica E4 wristwatch, wirelessly connected to robot
  - Extraction of features from Inter-Beat-Interval (IBI) and Galvanic Skin Response (GSR) signals [2]

[1] Giakoumis, D., Drosou, A., Cipresso, P., Tzovaras, D., Hassapis, G., Gaggioli, A. and Riva, G., 2012. Using activity-related behavioural features towards more effective automatic stress detection. **PloS one**, 7(9), p.e43571.

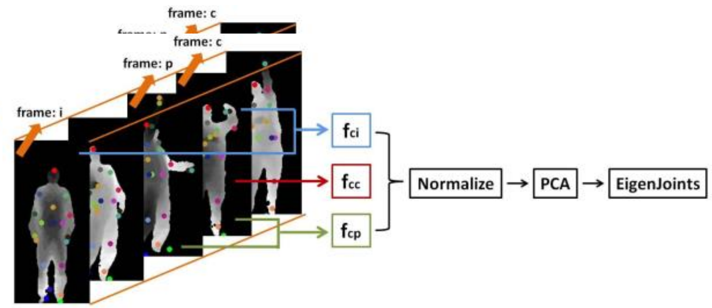
[2] Giakoumis, D., Tzovaras, D., Moustakas, K. and Hassapis, G., 2011. Automatic recognition of boredom in video games using novel biosignal moment-based features. **IEEE Transactions on Affective Computing**, *2*(3), pp.119-133.

Vision-based human activity recognition

approach overview (1/2)

- Aim: Detection of actions included in ADLs relevant to AAL
  - Based on human tracking through the robot's RGBD sensor

- **Overview of proposed approach [2]:**
  - Building upon the Eigenjoints-based method [1]
    - Extracts information about the **relative positions of the joints between frames** in video sequences



  - **Introducing extensions** towards robust action recognition in realistic domestic service robot applications

[1] X. Yang and Y. Tian. 2014. Effective 3d action recognition using eigenjoints. Journal of Visual Communication and Image Representation, 25(1), 2-11. (2014).

*[2] Stavropoulos, G., Giakoumis, D., Moustakas, K. and Tzovaras, D., 2017, June. Automatic action recognition for assistive robots to support mci patients at home. In Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments, **PETRA 2017**, (pp. 366-371). ACM.*

Vision-based human activity recognition

approach overview (2/2)

Proposed extensions to eigenJoints-based action recognition:

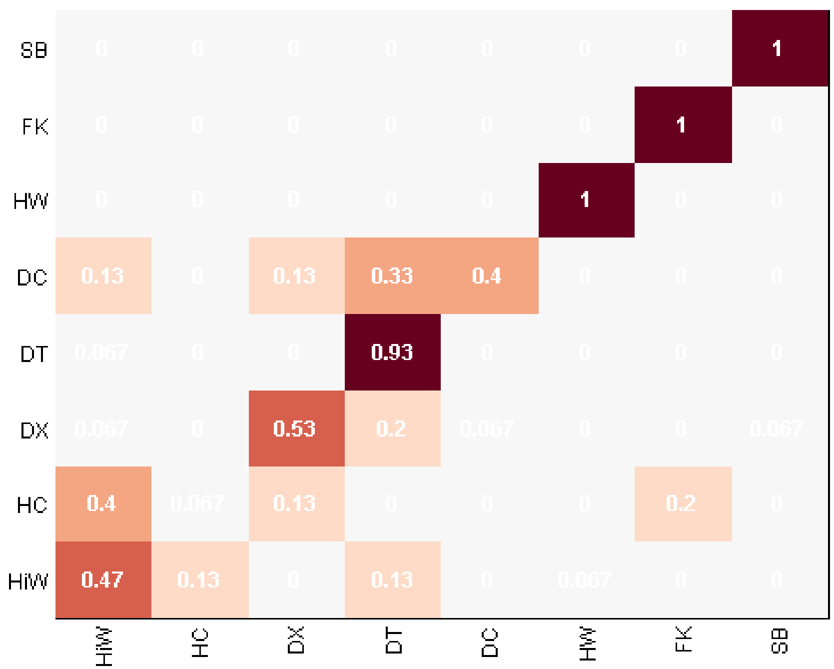
- **Motion trend** added as extra feature
  - By considering also the "next frame" in the video sequence, we add f_{cn} feature, analogous to f_{cp}, but extracted from the next frame
- **Accumulated travelled distance** of each joint over the video sequence added as extra feature
- Use only of the **corresponding joints** instead of all joints pairs in the f_{ci}, f_{cp} (and f_{cn} when used) features
  - Feature size and noise reduction of noise induced by action irrelevant joints (e.g. leg joints in a seated action)
- Detection of **objects manipulated** by the user
  - Information added to the action recognition method

[2] Stavropoulos, G., Giakoumis, D., Moustakas, K. and Tzovaras, D., 2017, June. Automatic action recognition for assistive robots to support mci patients at home. In Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments, ***PETRA 2017****, (pp. 366-371). ACM.*

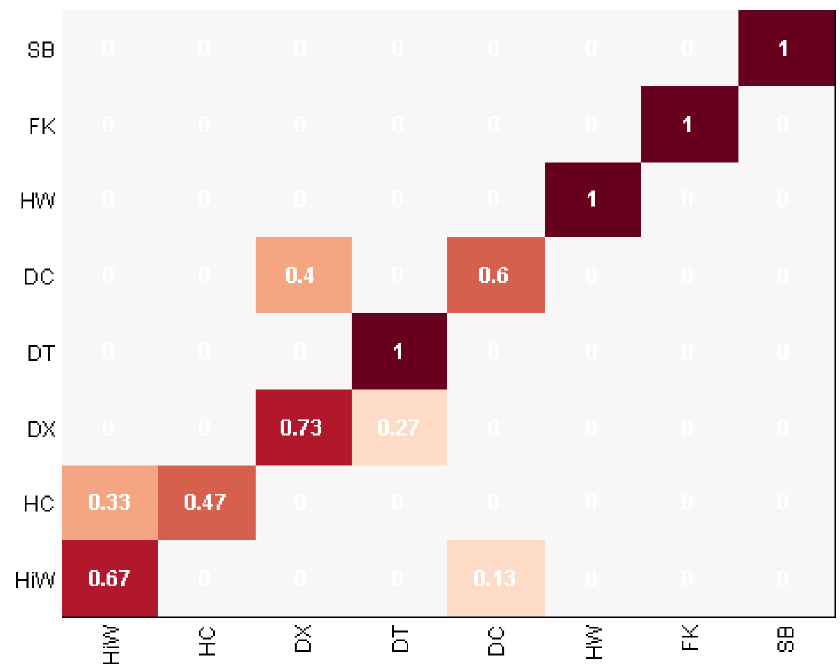
Vision-based human activity recognition
experimental results – public dataset

MSR Action Dataset

- **Confusion matrices, cross-subject** experiment, **Action Set 2**:



Original EigenJoints method



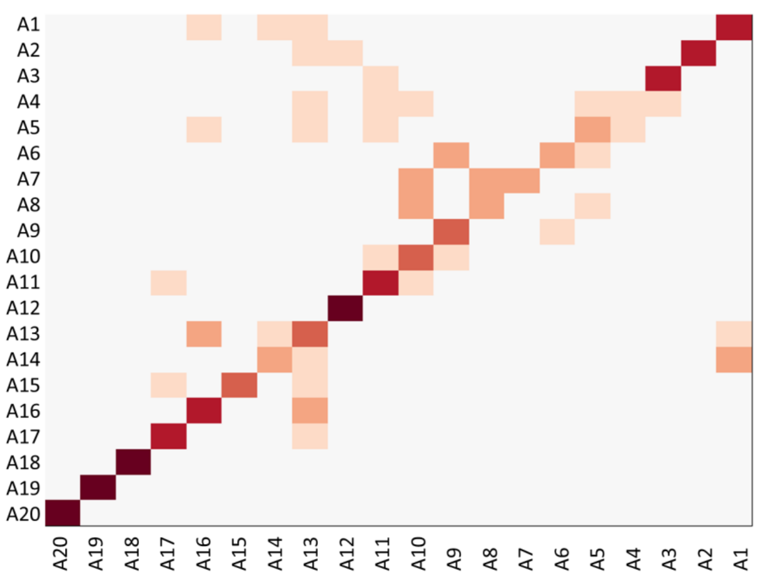
Proposed method

- Our method **improved performance**, especially **between actions with similar content**
    - e.g: in DX, DT and DC; all draw actions: Draw “X”, tick and cross respectively

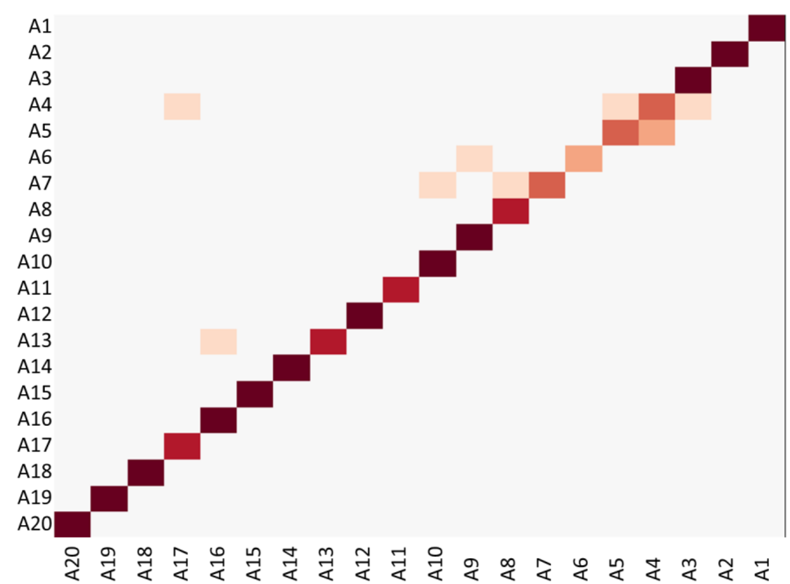
Vision-based human activity recognition

experimental results – realistic dataset

- **Confusion Matrices** on our custom, realistic daily activities dataset
- The proposed method **significantly improved action recognition performance:**
  - A9, A10 & A11: that are all “Open Cupboard” at different heights
  - A12, A13 & A14: “Eat”, “Alter” & “Drink” actions, where the manipulated object helps to distinguish between actions with very similar content



Original EigenJoints method

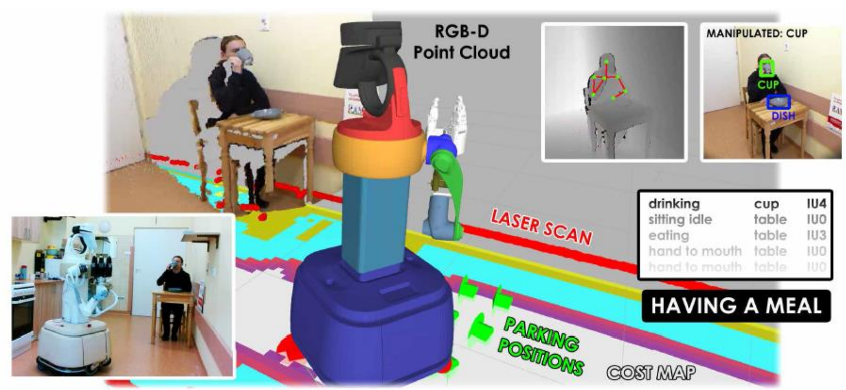


Proposed method, with all extensions, incl. manipulated object

Human behavior analysis

IU and DBN-based behavior monitoring on top of CV

- ***Aim:*** *Analysis of human behavior towards proactive assistive robot decisions*
- Development of novel user behavior analysis method [1], based on **Dynamic Bayesian Networks (DBNs)**
- Adoption of the **Interaction Unit (IU) Analysis**
    - Decomposition of complex activities into simple actions.
    - Systematic notation on how simple actions are associated with behavioral factors
    - Correlation of atomic actions with manipulated objects
- Application on common activities of daily living
    - Meal preparation cooking, medication intake and eating activities

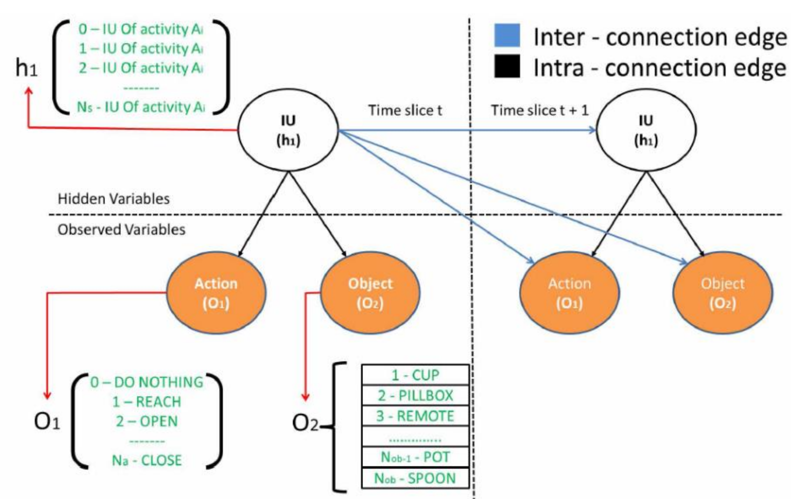


[1] Kostavelis, I., Vasileiadis, M., Skartados, E., Kargakos, A., Giakoumis, D., Bouganis, C. S., & Tzovaras, D. (2019). Understanding of human behavior with a robotic agent through daily activity analysis. International Journal of Social Robotics, 1-26.

Human behavior analysis

IU and DBN-based behavior monitoring on top of CV

- Modeling of the IU analysis with a Dynamic Bayesian Network for:
  - Activity recognition
  - Interpretation of the IU steps to extract insights about the way an activity is performed
- Modelling of normal and abnormal behavior
  - Statistics and post processing on the resulting Viterbi path of DBN network
  - Understanding of user’s normal and abnormal behaviors
    - ❑ to be used for robot decision making…



IU	Human Action	Manipulated Object	Environment State	Behavioral Factor Recognition/Recall/Action	Abnormal Response	Priority
1	Reach	pill box	pill box closed on table	The pill box is on the table	Forgets where the objects are, looking in the different area. Picks up different than desired object. Easily distracted	M
2	Alter	pill box	pill box in hand, opened	The pill box is closed	Opens the box and does not remember the reason	NM
3	Hand to mouth	cup	cup with water on table	Take the pill	Gets easily distracted	M
4	Alter	pill box	pill box closed	The pill box is opened	Gets easily distracted	M
5	Reach	pill box	pill box on table	Place the pill box on table	Gets easily distracted, misplaces the objects	NM
6	Reach	cup	cup with water on table	The cup is on the table	Forgets where the objects are, looking in a different area Takes the objects and later does not use them. Pick up different than the desired object	M
7	Hand to mouth	cup	cup with water on table	Drink water from the cup	Gets easily distracted	M
8	Reach	cup	cup empty on table	Place the cup on the table	Gets easily distracted	NM

[1] Kostavelis, I., Vasileiadis, M., Skartados, E., Kargakos, A., Giakoumis, D., Bouganis, C. S., & Tzovaras, D. (2019). Understanding of human behavior with a robotic agent through daily activity analysis. International Journal of Social Robotics, 1-26.

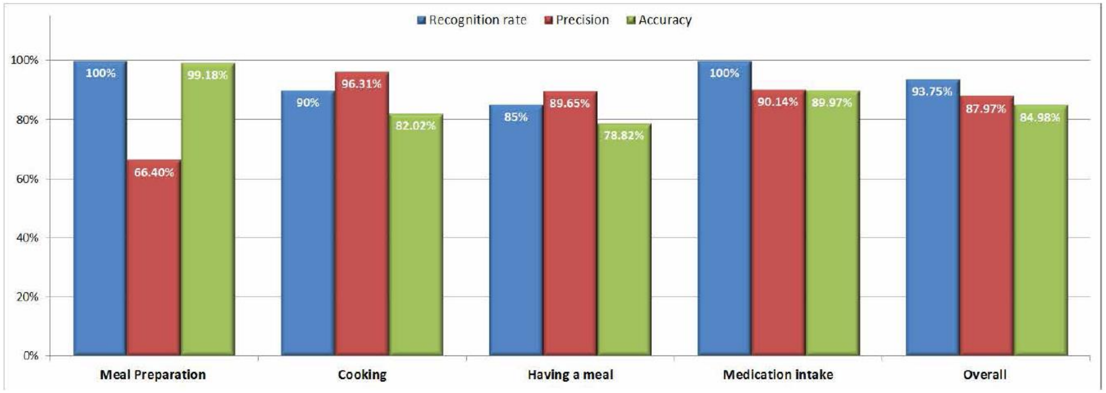
Object recognition, human tracking and activity monitoring

Integrated in the RAMCIP user behavior monitoring approach

IU and DBN-based behavior analysis

experimental evaluation

- Dataset collected at a simulated apartment
  - 18 subjects performed 4 activities, ~120 repetitions
  - 98% classification accuracy on activity recognition



  - Above 85% classification accuracy on IU analysis

Meal preparation

IU 1	IU 2	IU 3*	IU 4*	Overall
100%	89.74%	64.10%	79.41%	83.31%

Cooking

IU 1	IU 2*	IU 3*	IU 4*	Overall
85%	100%	85%	80%	87.50%

Having a meal

IU 1	IU 2	IU 3*	IU 4*	IU 5	Overall
93.75%	93.75%	100%	96.87%	93.75%	95.62%

Medication intake

IU 1*	IU 2*	IU 3*	IU 4	Overall
75%	81.25%	96.87%	75%	82.03%

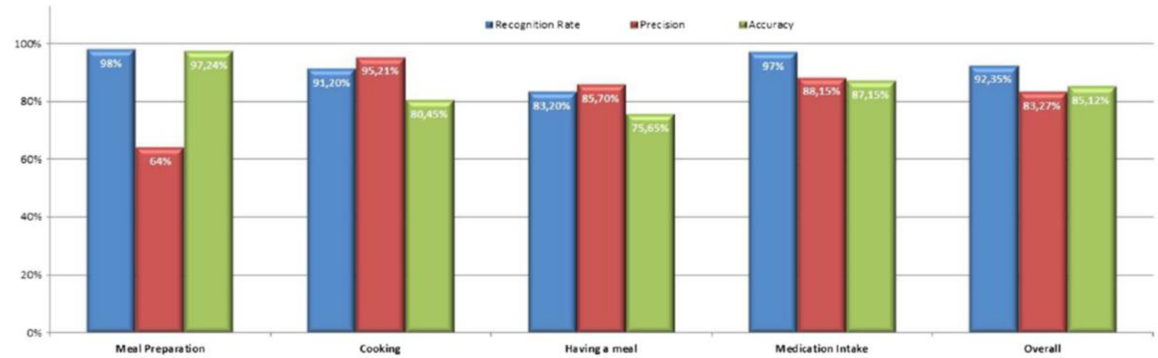
*mandatory IU step

[1] Kostavelis, I., Vasileiadis, M., Skartados, E., Kargakos, A., Giakoumis, D., Bouganis, C. S., & Tzovaras, D. (2019). Understanding of human behavior with a robotic agent through daily activity analysis. International Journal of Social Robotics, 1-26.

IU and DBN-based behavior analysis

experimental evaluation

- Dataset collected in real homes
  - 12 subjects performed 4 activities, for a whole week at their own homes, >300 repetitions
  - Approx. 92% classification accuracy on activity recognition



  - Approx. 85% classification accuracy on IU analysis

IU 1	IU 2	IU 3ª	IU 4ª	Overall
Meal preparation				
94.29%	87.14%	84.29%	87.14%	88.21%

IU 1	IU 2ª	IU 3ª	IU 4ª	Overall
Cooking				
88.57%	91.43%	84.29%	87.14%	87.86%

IU 1	IU 2	IU 3ª	IU 4ª	IU 5	Overall
Having a meal					
90.48%	88.10%	91.67%	72.62.87%	82.14%	85.00%

IU 1ª	IU 2ª	IU 3ª	IU 4	Overall
Medication intake				
86.90%	83.33%	86.90%	83.33%	85.12%

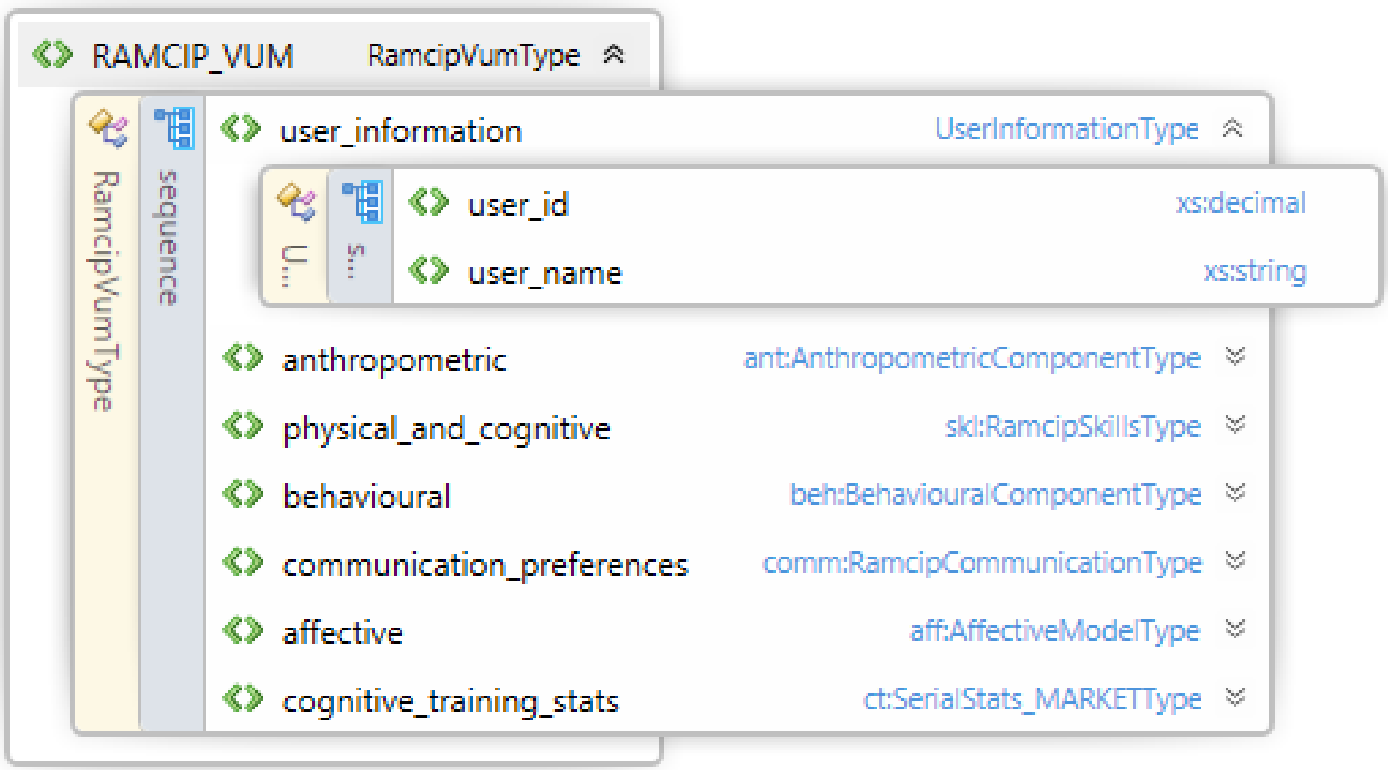
ªMandatory IU step

[1] Kostavelis, I., Vasileiadis, M., Skartados, E., Kargakos, A., Giakoumis, D., Bouganis, C. S., & Tzovaras, D. (2019). Understanding of human behavior with a robotic agent through daily activity analysis. International Journal of Social Robotics, 1-26.

User-centric Robot Cognition

VUM-based knowledge representation

- *Virtual User Model (VUM) –based representation of key end user aspects*
    - ***VUM, semantic map and real-time user behavior observations drive robot decisions for personalized context-aware operation***

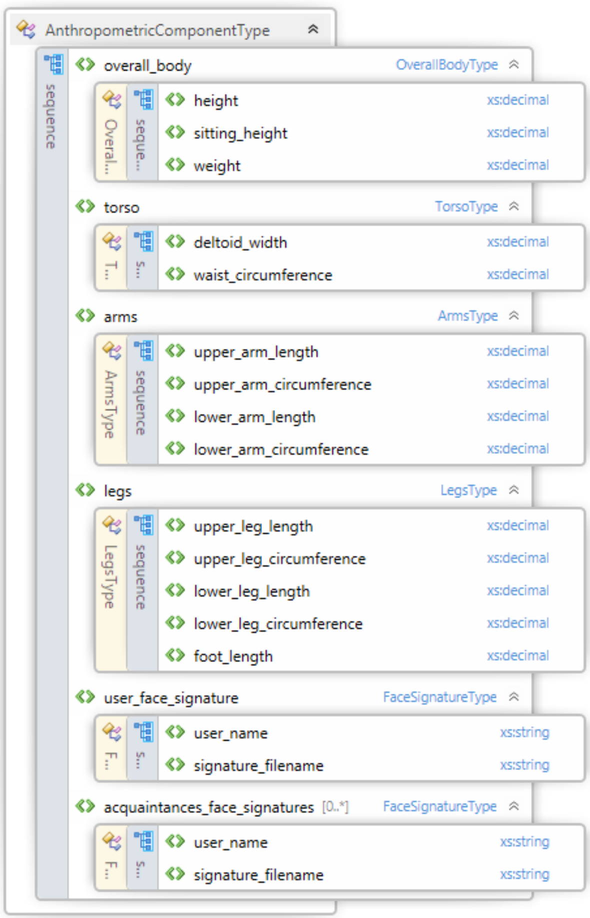


User-centric Robot Cognition

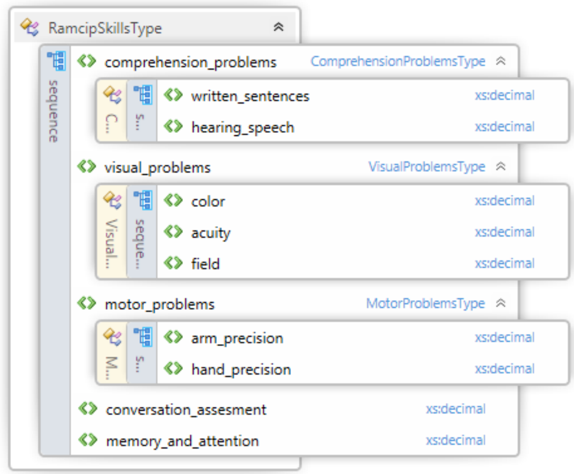
VUM-based knowledge representation

RAMCIP VUM main parts summary (1/2)

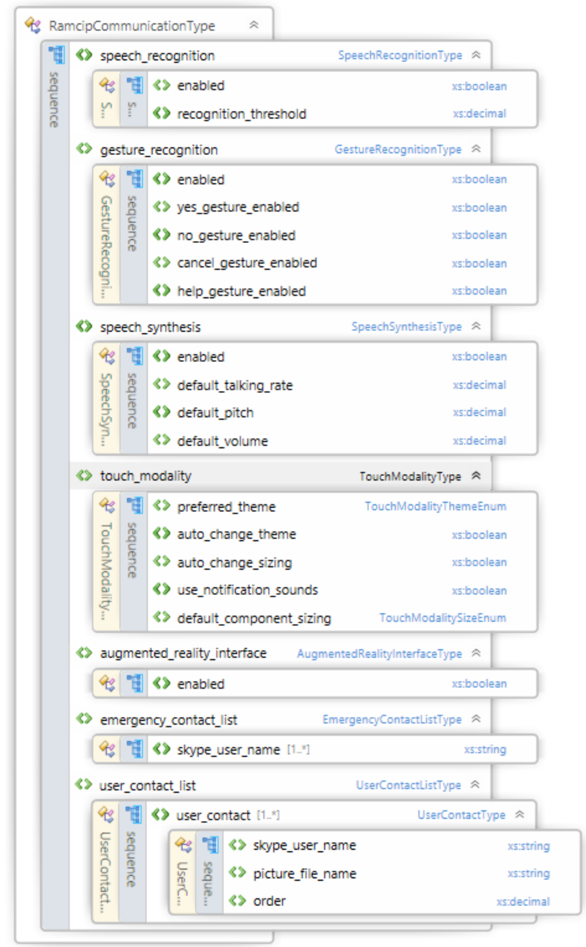
Anthropometric



Physical and Cognitive Skills



Communication Preferences

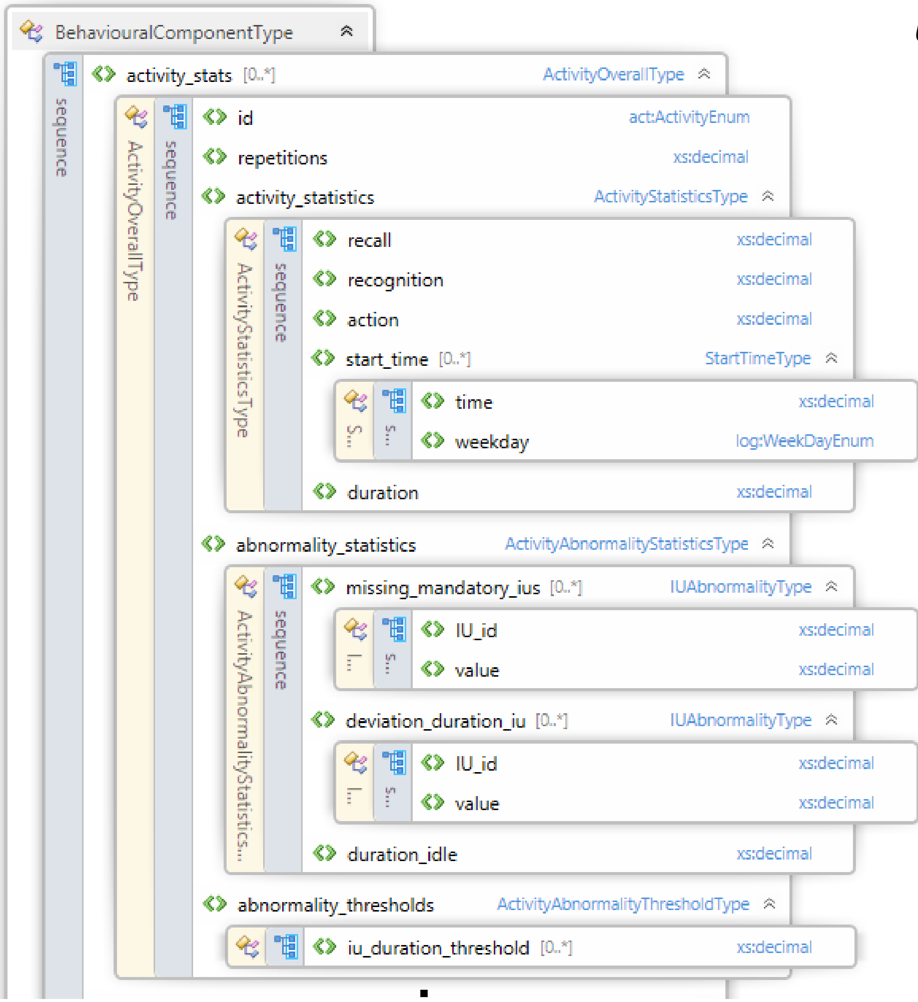


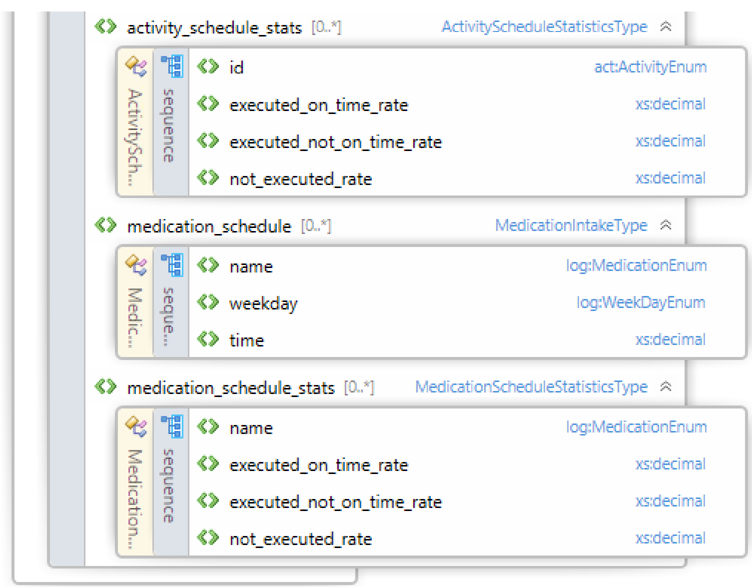
User-centric Robot Cognition

VUM-based knowledge representation

RAMCIP VUM main parts summary (2/2)

User Behavior





User-centric Robot Cognition
high-level assistive robot decision making

- *Aim:*
  - To provide a Decision Making strategy suitable to determine **when** and **how** the robot should intervene to assist the user
- *To achieve this, the robot should:*
  - be constantly aware about the user and the environment
  - act as a prompting system that **associates** the **robot's awareness about the user** with specific types of **robotic actions**
  - compensate the **partial sensor input** acquired during the monitoring and modeling of the daily human activities through vision
- *Our approach relies on Partially Observable Markov Decision Processes (**POMDP**)*[1]
  - POMDPs handle the partial observability of the environment
  - POMPDs are prompting systems based on the rule that:

The agent receives observations from the environment and decides on its actions

[1] Pineau, J., Thrun, S.: High-level robot behavior control using pomdps. In: AAAI 02 Workshop on Cognitive Robotics. Volume 107. (2002)

User-centric Robot Cognition

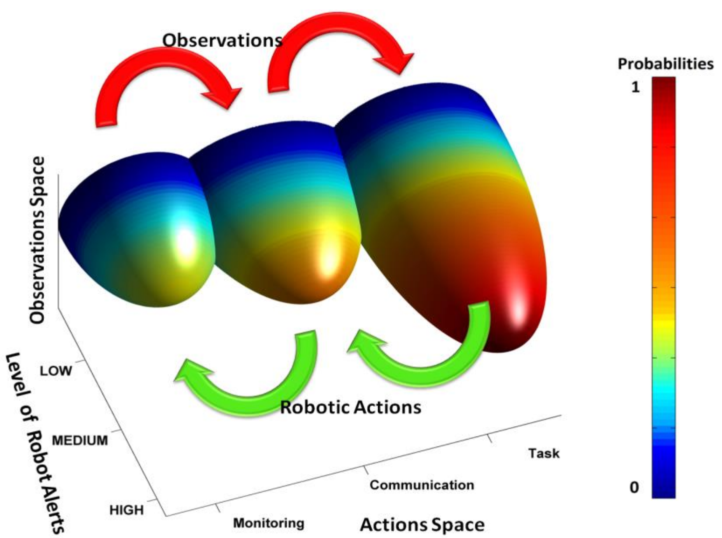
POMDPs in robot decision making

- *POMDP basic principles:*
  - **State space**: Determines the *condition of the environment*
  - **Actions space**: Comprises the *set of actions that the agent is able to perform* so as to interact with the environment
  - **Observations space**: Encloses the agent's *perception input* from the environment
  - **Rewards**: The restrictions space imposed by *penalizing or endorsing specific agent action* given the environment state

- *POMDP-based approach for a proactive domestic service robot (e.g. RAMCIP) :*
  - **States** correspond to the **robot alert levels** about the human and the environment
  - **Actions** comprise the set of robotic actions that the robot is able to perform so as to interact with the human and the environment
  - The robot **intervention actions** are associated with the robot's **levels of alert about the human and the environment**

User-centric Robot Cognition
POMDP design principle in RAMCIP

- Our POMDP-based decision making approach takes into account:
  - The **state space** based on robot levels of alert about the human and the environment
    - **High levels of alert** require *engagement with actual robot task* e.g. fetching tasks
    - **Medium levels of alert** require *engagement with communication tasks* e.g. dialogue
    - **Low levels of alert** require *monitoring of the human* e.g. tracking and activity recognition
  - The **action space;** actions triggered based on the current state of the robot and context
    - Robot task planning (navigation, manipulation, grasping, …)
    - Robot communication (dialogue, User Interphase, Augmented Reality display, …)
    - Robot monitoring (vision based human & environment tracking, activity recognition, …)
- A POMDP model is produced based on the principle:
  - ***Observations* tend to increase the levels of robot alert**
  - ***Actions* tend to decrease the levels of robot alert**
- A policy graph is computed:
  - Outlines a sequence of robotic actions for the denouement of the assistive scenario



User-centric Robot Cognition

POMDP design principle in RAMCIP

- **POMDP model generation (Novel in RAMCIP)**
  - FSM diagrams have been created as <u>**maps**</u> that constrain the POMDP models for the RAMCIP use cases
    - **Transition probabilities** among <u>directly connected states are</u> modeled with <u>increased values</u> normalized to the total number of states in the FSMs
    - **Observation probabilities** among <u>linked robotic actions</u> are also explicitly declared with <u>increased values</u> and the rest observation probabilities receive uniform values
    - **Rewards:**
      - Increased **positive** values assigned for the transition **from high to low states**
      - **Negative** values are assigned for the transition **from low to high states**

- **Partial observability trick** (from FSM to POMDP):
    - The produced policy graph resolves the assisting scenarios irrespectively the state that will be initiated
    - Additive value: <u>the system tends to transit in **states of low level of robot alerts**</u>
    - The FSM is the ideal scenario; *the current design can resolve the use cases by simultaneously considering the probability of appearance of **all the observations***

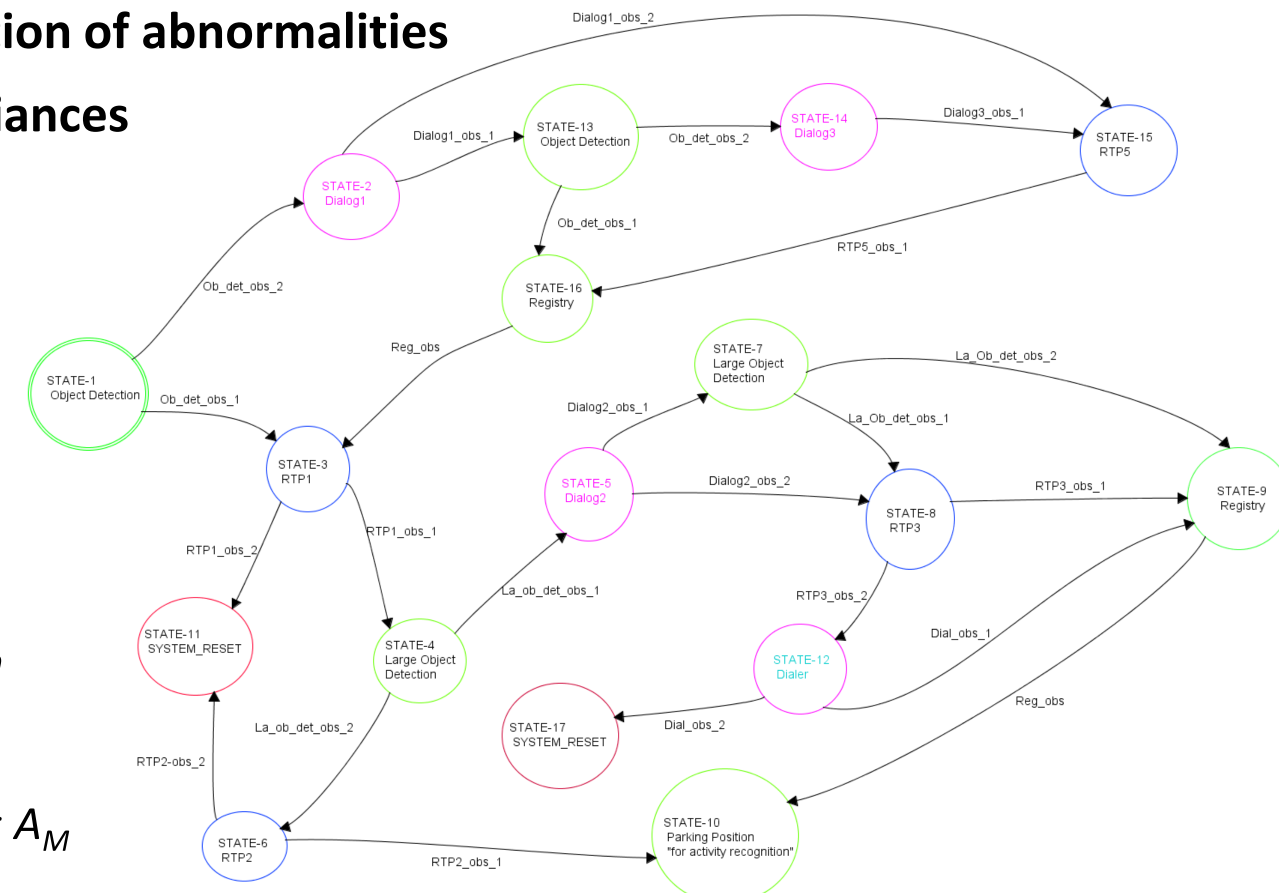
User-centric Robot Cognition

POMDP design principle in RAMCIP

- **Exemplified RAMCIP Scenario**

  Assistance upon detection of abnormalities related to electric appliances during cooking

- **FSM state diagram:**



- ***Blue:*** *Task actions* A_T
- ***Magenta:*** *Communication actions* A_C
- ***Green:*** *Monitoring actions* A_M

User-centric Robot Cognition
POMDP design principle in RAMCIP

- **Exemplified RAMCIP Scenario**

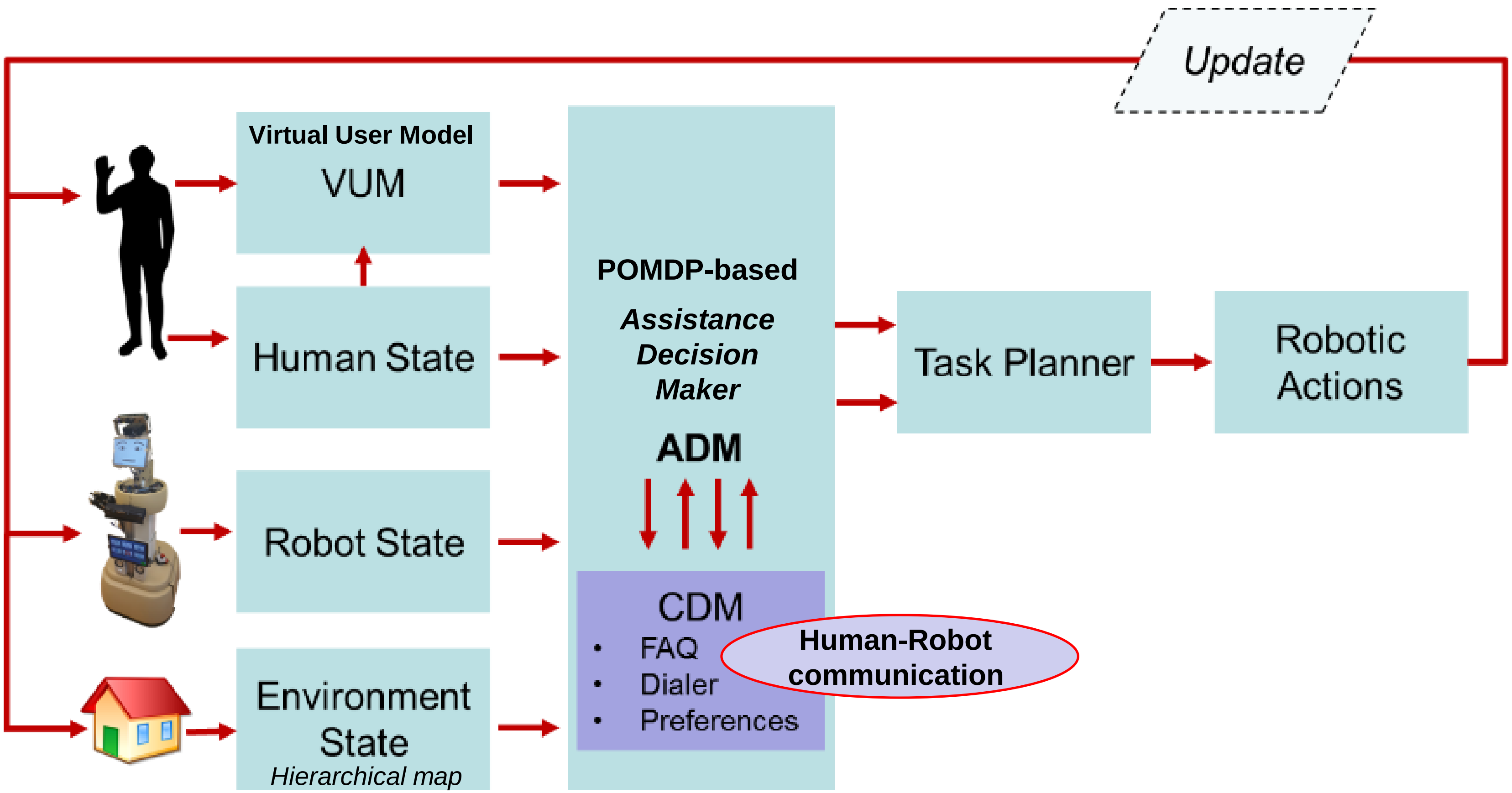
POMDP design

- **Conceptual grouping of states**
- **Clustering of robotic actions**

- **Table on the right:**
  - Mapping from FSM diagram to POMDP
    - Level of robot alerts (States)
    - Group of robotic actions (Actions)

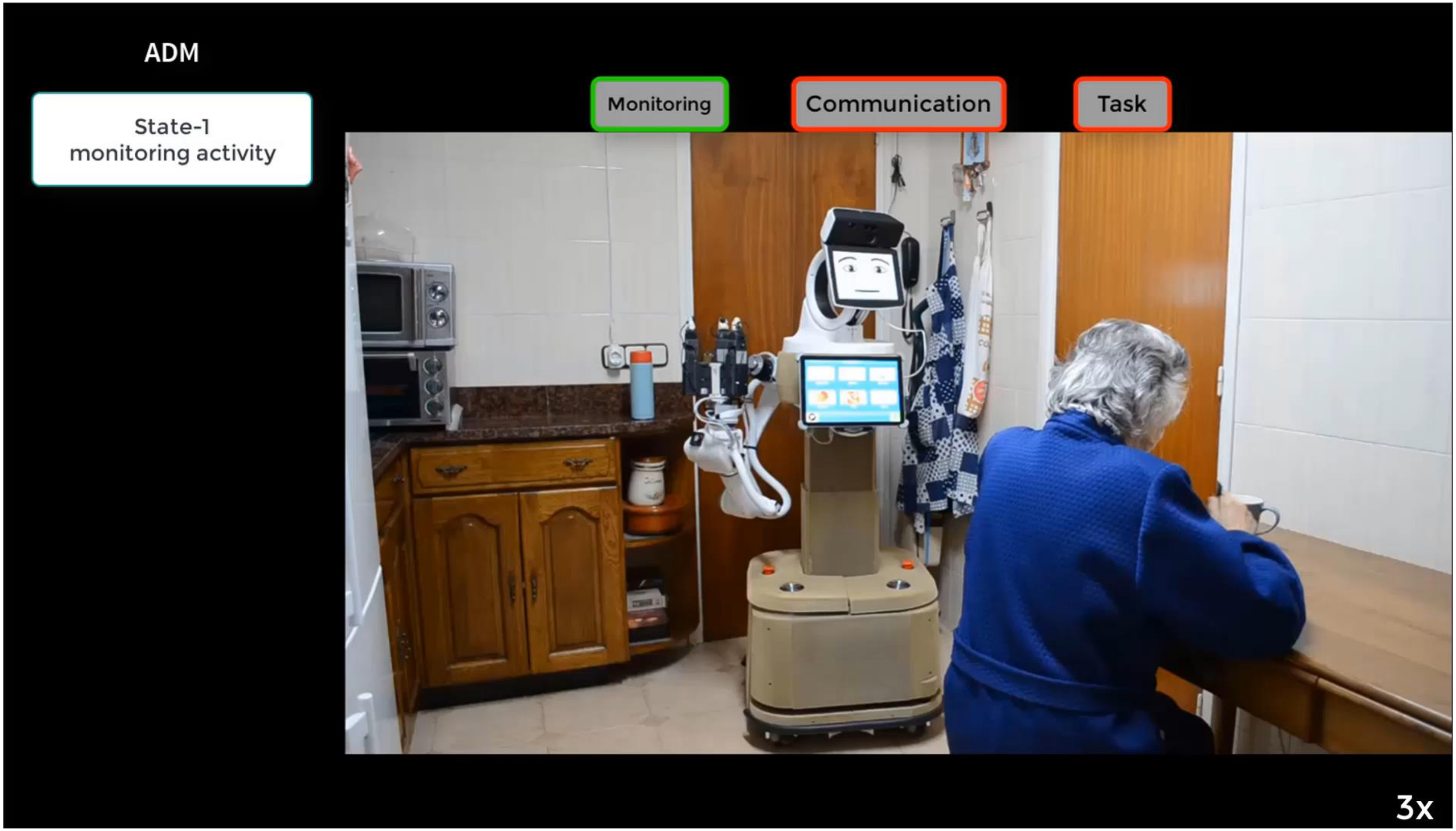
Levels of Robot Alert			Actions		
High	*Medium*	*Low*	*Task*	*Communication*	*Monitoring*
State-3	State-2	State-1	**RTP1**: Robot navigates to the parking position suitable to monitor the state of appliance	**Dialog1**: Robot communicates with human about some missing objects and asks if it should fetch them	**Object-Detection**: The SW component suitable to detect and recognize small objects
State-6	State-5	State-4	**RTP2**: Robot navigates to the parking position suitable to monitor the cooking activity	**Dialog2**: Robot communicates with human about forgetting to turn off an appliance and asks if it should close it	**Large object detection**: The SW component suitable to recognize the state of large articulated objects
State-8	State-12	State-7	**RTP3**: Robot plans the actions for navigation and manipulation of appliance	**Dialog3**: Robot informs the human that it will go manipulate the appliance	**Registry**: The SW component suitable to register the incidents
State-15	State-14	State-9	**RTP4**: The robot fetches the missing objects	**Dialer**: The robot failed to turn off the appliance and notifies for external help	**Parking Position**: The SW component suitable to switch the robot in monitoring state where the human and environment are observed
—	—	State-10	—	—	—
—	—	State-13	—	—	—
—	—	State-16	—	—	—

User-centric Robot Cognition
POMDP-based decision making in RAMCIP



User-centric Robot Cognition

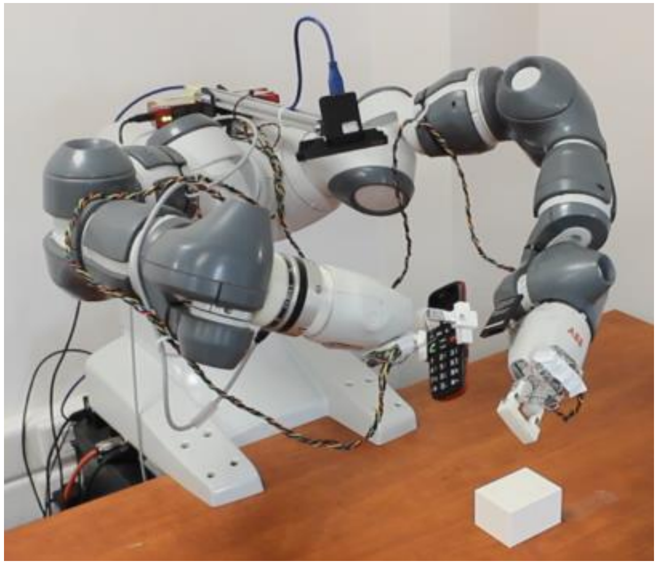
POMDP-based decision making in RAMCIP



Applications

- Professional service robots

@agile manufacturing



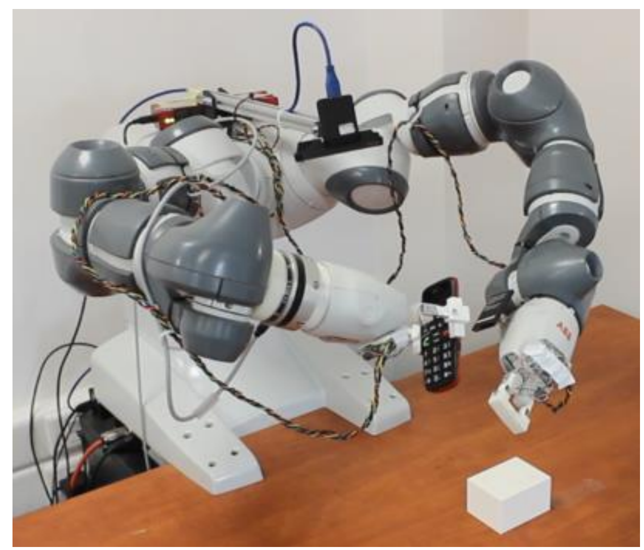
Professional service robots
@agile manufacturing

Robot perception in learning by demonstration tasks

- Key challenge:

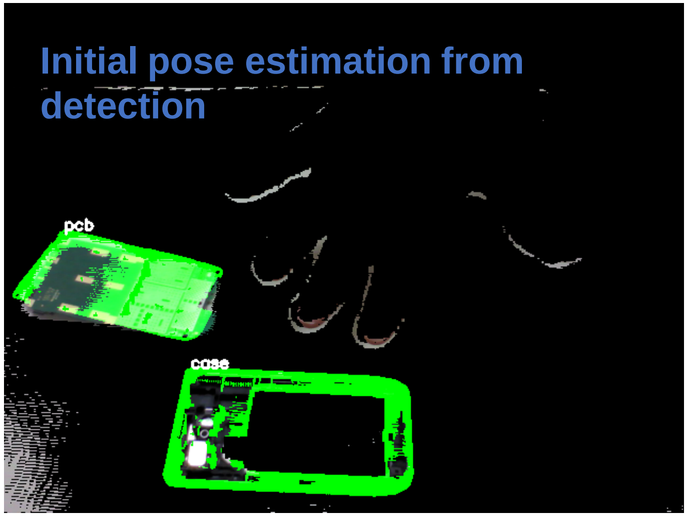
Hand-object detection and tracking in 3D

…through commercial RGB-D sensors



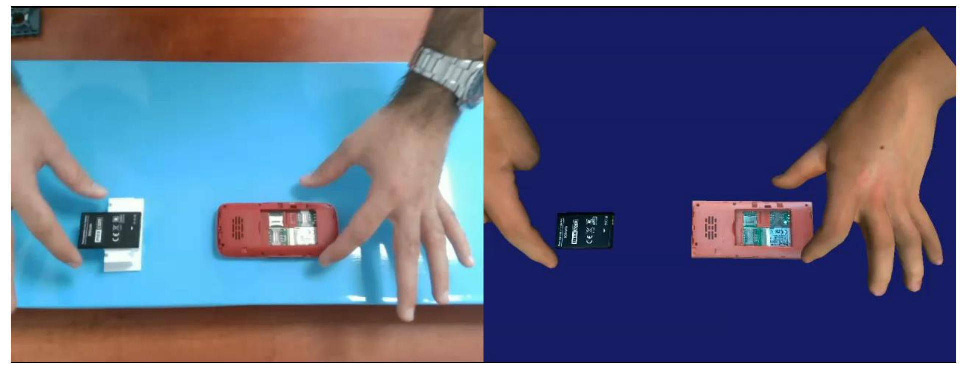
Learning by demonstration
hand-object detection and tracking in 3D

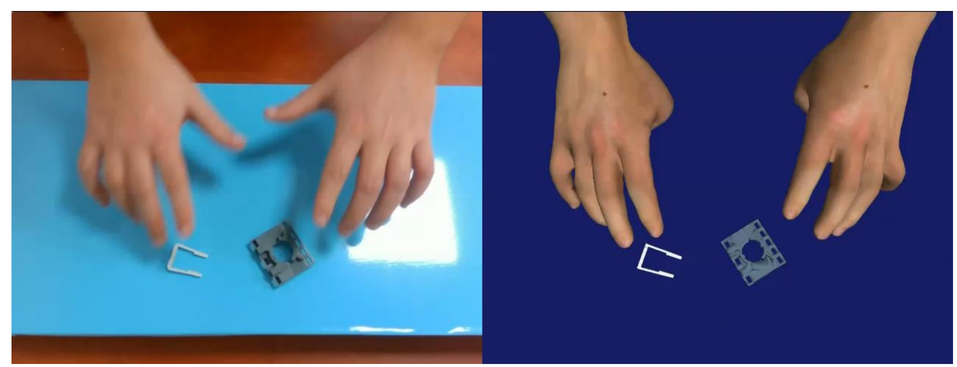
- **Input:** RGBD data from common commercial sensor
- **Object Detection** (6DoF pose) is performed based on sparse auto-encoders for feature extraction and Hough Forests for classification
- **3D CAD models** are employed for both **training** the object detector and **performing** hand-object tracking
  - **6 DoF** for the models of the assembly parts
  - **42 DoF** for the hand models
- **Coarse hand detection** of an open configuration is performed



Learning by demonstration
hand-object detection and tracking in 3D

- **Hand-Object Tracking** implementation using Particle Swarm Optimization (PSO)
  - Detection results are used for initializing the tracker
  - Building upon existing approaches on hand tracking in order to perform joint hand – object tracking
  - **Addressing deformable objects**, as well
  - **Optimization Time:** 0.6 sec per frame





Learning by demonstration

key-frames extraction

Key-frames:
Important states of the demonstrated assembly
Folding Assembly Example

Key-frame information **stored in** **XML format**

General information:

- Scenario id and current step
- Object(s) id involved in the demonstration phase
- Relative timestamp

Kinematics & Motion information:

- Object pose coordinates (position & orientation, 6 DOF)
- Hand pose (42 DOF)

Semantic information:

- User defined corresponding to assembly states, e.g. *grasp, align*
- Automatic system suggestions, e.g. *aligned axes*

Dynamics information:

- Forces derived from the kinesthetic learning
- Grasping contact points
- Object deformation characteristics


```
<?xml version="1.0" encoding="UTF-8" standalone="true"?>
- <KeyFrame xsi:schemaLocation="http://www.SARAFunXML.com
  SARAFun_KeyFrame_XmlSpec_v02.xsd" xmlns:xsi="http://www.w3.org/2001/XMLSchema-
  instance" xmlns="http://www.SARAFunXML.com" t="25.4" idx="1" id="0">
  - <CurrentAction id="assembly.mpg">
      <Description>Putting one object over the other</Description>
    - <InvolvedObjects>
        <Object id="Obj1"/>
        <Object id="Obj2"/>
      </InvolvedObjects>
    - <VisualFeedback>
      - <CameraSensor id="RealSenseF200">
          <FrameRange fileList="RealSenseF200_Sequence.xml" idxLast="210" idxFirst="30"/>
        </CameraSensor>
      - <CameraSensor id="Xtion">
          <FrameRange fileList="Xtion_Sequence.xml" idxLast="220" idxFirst="40"/>
        </CameraSensor>
      </VisualFeedback>
    </CurrentAction>
  - <Objects>
    - <Object id="ObjA" name="Mobile Phone PCB">
        <MeshFile>mobile_phone_pcb.obj</MeshFile>
      - <PoseState>
          <Position z="-0.36945" y="-0.0175897" x="-0.125605"/>
          <YPR rotz="-1.73068" roty="-0.679461" rotx="0.0018003"/>
        </PoseState>
        <Deformation>NotYetDefined</Deformation>
      </Object>
    - <Object id="ObjB" name="Mobile Phone Case">
        <MeshFile>mobile_phone_case.obj</MeshFile>
      - <PoseState>
          <Position z="-0.317434" y="-0.0832089" x="-0.0241354"/>
          <YPR rotz="-0.0524788" roty="0.0192357" rotx="-0.723375"/>
        </PoseState>
        <Deformation>NotYetDefined</Deformation>
      </Object>
    </Objects>
  - <Instructor>
    - <Hand id="LeftHand" name="Instructors left Hand">
```

Learning by demonstration
key-frames extraction

- **Key-frames:** Important states of the demonstrated assembly
  - Finite State Machines (FSMs) or Behavioral Trees (BTs) employ the extracted Key-frames and their information to automatically generate the robot's assembly program

    Proposed Key-frame extraction approach [1]
  - Employing hand-object tracking in 3D for kinematic information extraction and automatic Key-frame identification based on **semantic graphs** from image sequences
    - Extending past approaches focusing on 2D RGB images [2]
    - Detecting Key-frames as structural changes of the semantic graph
    - Post processing (and instructor's feedback) for identifying the assembly state (e.g. grasp, contact, etc.) corresponding to each Key-frame for construction of the assembly FSM or BT

[1] Grigorios S. Piperagkas, Ioannis Mariolis, Dimosthenis Ioannidis, Dimitrios Tzovaras:Key-frame Extraction With Semantic Graphs in Assembly Processes. **IEEE Robotics and Automation Letters** 2(3): 1264-1271 (2017)

[2] Aksoy et al. Learning the semantics of object–action relations by observation." Int. Journal of Robotics Research 2011,

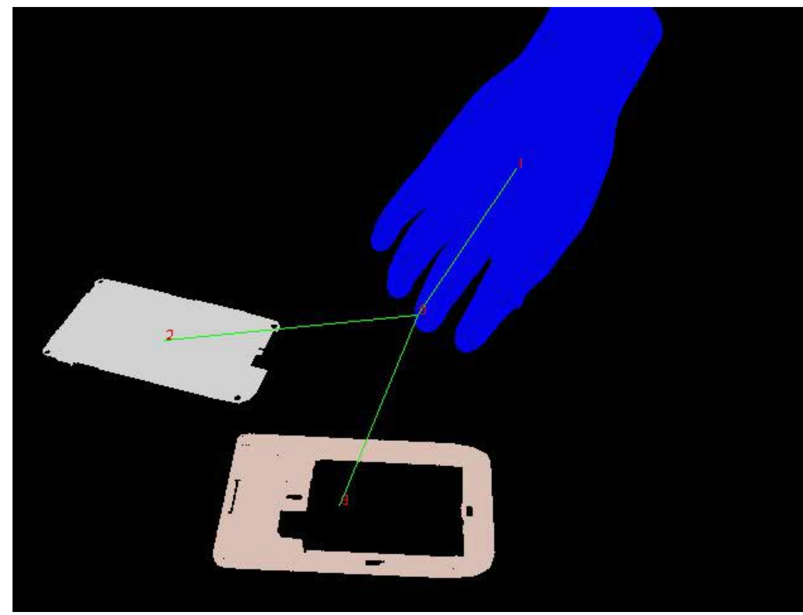
Learning by demonstration
automatic key-frame identification

Proposed method: Semantic Graphs[1]

- Hand and objects segmentation
  - Using output from tracking module

- Scanning of each labeled input image horizontally and vertically, to count the relations between objects and/or hands
  - The sequence is analyzed semantically, by labeling the graph edges as “absent -11”, “not touching-0”, “touching-2” and “overlapping-1”

- Construction of semantic graph
  - **compressed graph of derivatives** of actions/states which define the core of the sequence



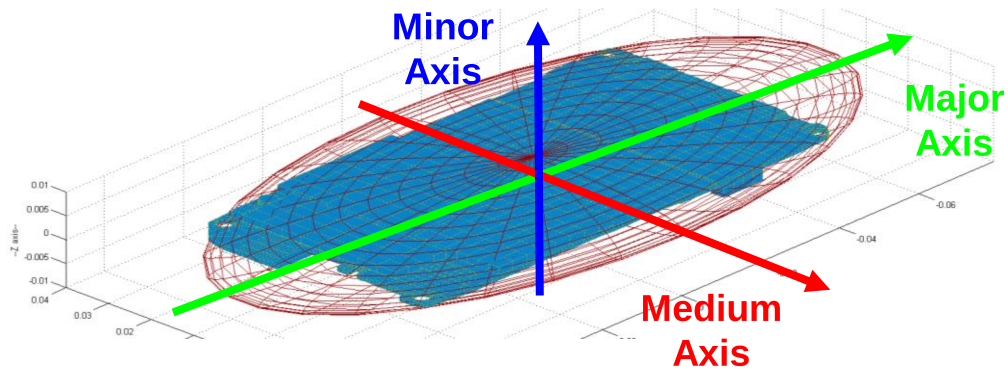
[1] Grigorios S. Piperagkas, Ioannis Mariolis, Dimosthenis Ioannidis, Dimitrios Tzovaras:Key-frame Extraction With Semantic Graphs in Assembly Processes. **IEEE Robotics and Automation Letters** 2(3): 1264-1271 (2017)

Learning by demonstration
automatic key-frame identification

New features using 3D information from tracker

Novel modeling approach in 3D, based on ellipsoids

- 3D Ellipsoids are automatically **fitted to the objects' CAD models**, at initialization stage



- Solution of **Minimum Volume Covering Ellipsoid** problem, by exploiting a *Dual Reduced Newton* convex optimization algorithm, yields a precise ellipsoid for each object
- Processing of manipulation actions is now **processing of relations between ellipsoids** in 3D
  - using **3D pose and position** of objects/hand estimated from **hand-object tracking** algorithm
  - **2D rendered images** are also employed
- Analytical **computation of free margin** between ellipsoids: ability to **track touching or overlapping in 3D**

Learning by demonstration

automatic key-frame identification

Semantic Graph definitions using the new 3D features[1]

- 5 node labels
  - 0: Background
  - 1: Right Hand
  - 2: Assembly Part1
  - 3: Assembly Part2
  - 4: Left Hand
- 11 edge labels indicating
  - No relation
  - Touching
  - Overlapping
  - Alignment
  - Combinations of the above

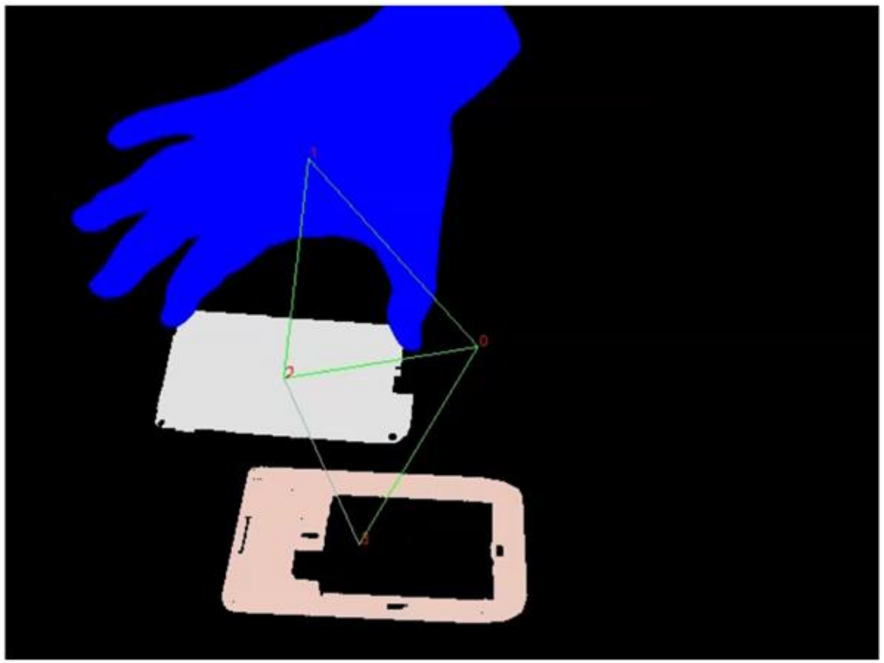
Edge labels	Relations between nodes n_i and n_j
0	i,j nodes present – No relation between them
1	i overlapping j
2	i touching j
3	Ellipsoid's i major axis is parallel to ellipsoid's j selected axis
4	Ellipsoid's i medium axis is parallel to ellipsoid's j selected axis
5	Ellipsoid's i minor axis is parallel to ellipsoid's j selected axis
6	All ellipsoid's i axes are parallel to selected ellipsoid's j axes
7	Ellipsoids i and j are touching and have one axis parallel
8	Ellipsoids i and j are touching and have all axes parallel
9	Ellipsoid i is overlapping j and they have one axis parallel
10	Ellipsoid i is overlapping j and they have all axes parallel
11	i or j is absent from the current frame

[1] Grigorios S. Piperagkas, Ioannis Mariolis, Dimosthenis Ioannidis, Dimitrios Tzovaras: Key-frame Extraction With Semantic Graphs in Assembly Processes. **IEEE Robotics and Automation Letters** 2(3): 1264-1271 (2017)

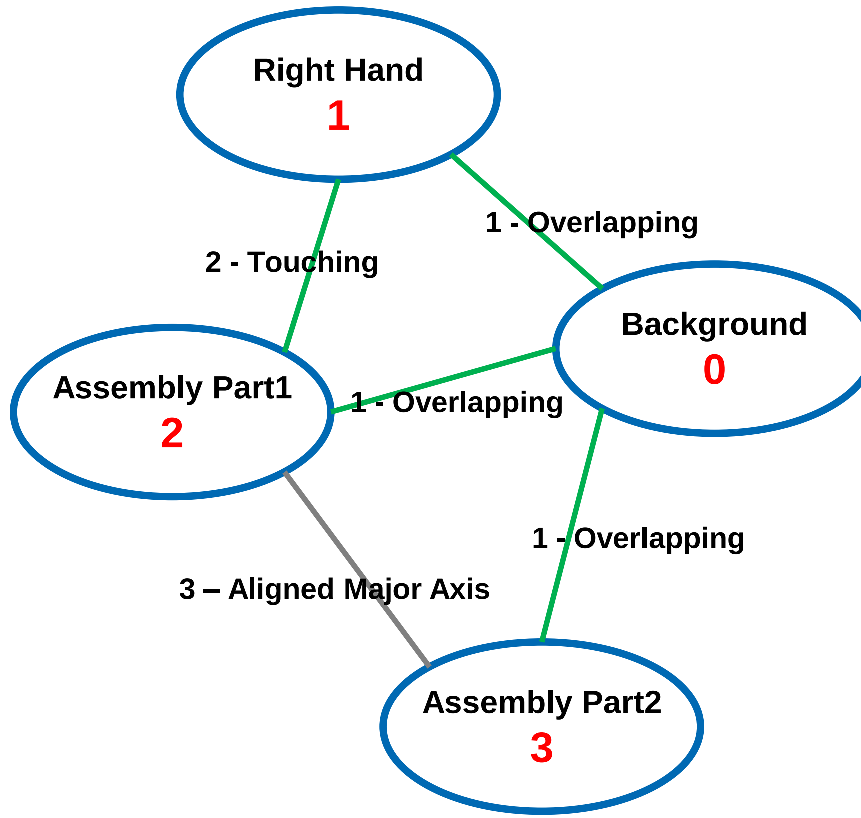
Learning by demonstration

automatic key-frame identification

Demonstrated Scene Frame



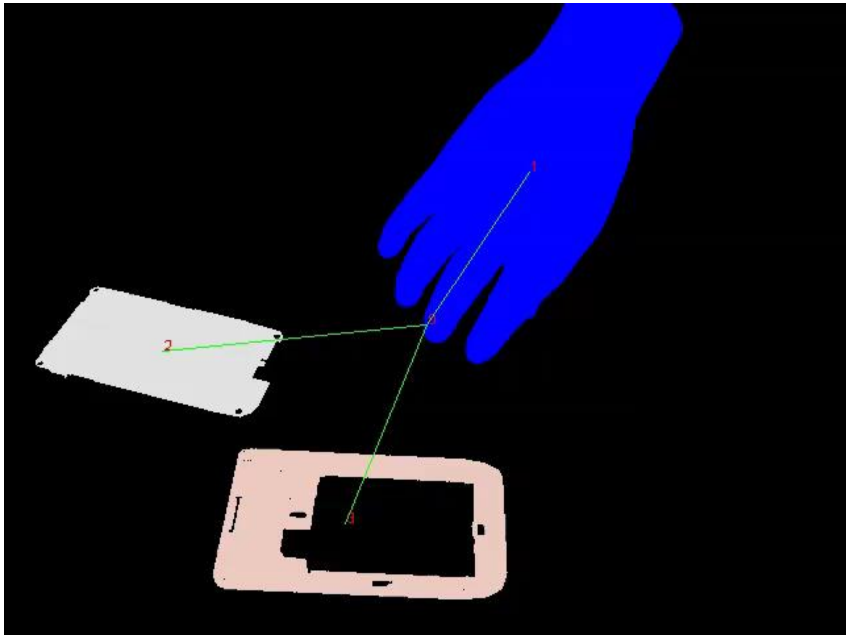
Corresponding Semantic Graph



Learning by demonstration

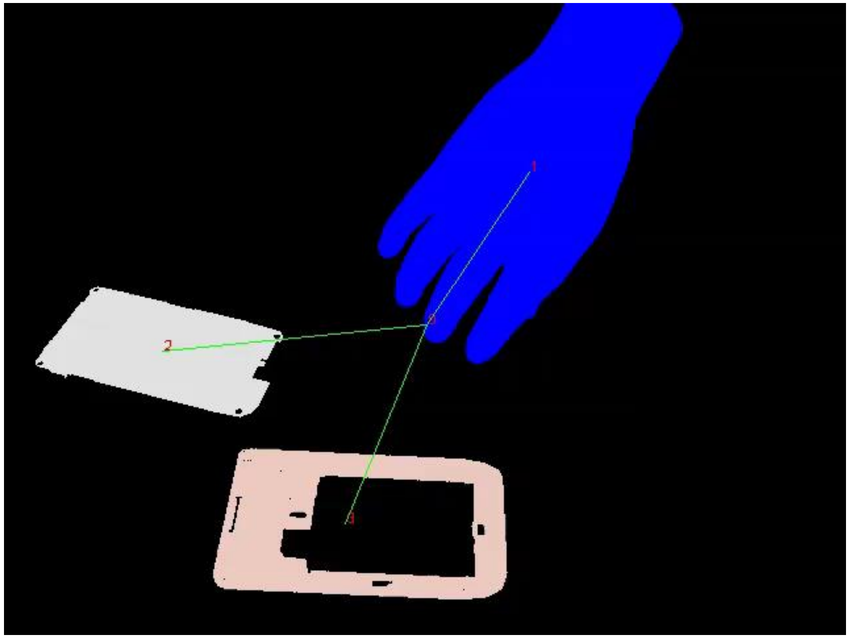
automatic key-frame identification

using only 2D information



Extracted semantic graphs

method extended to 3D using ellipsoids



Extracted semantic graphs

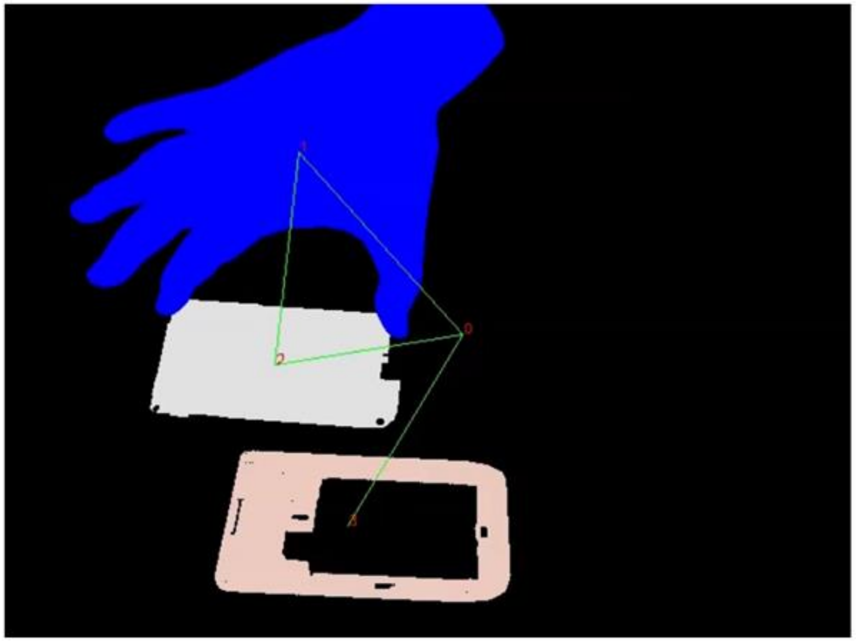
Learning by demonstration

automatic key-frame identification

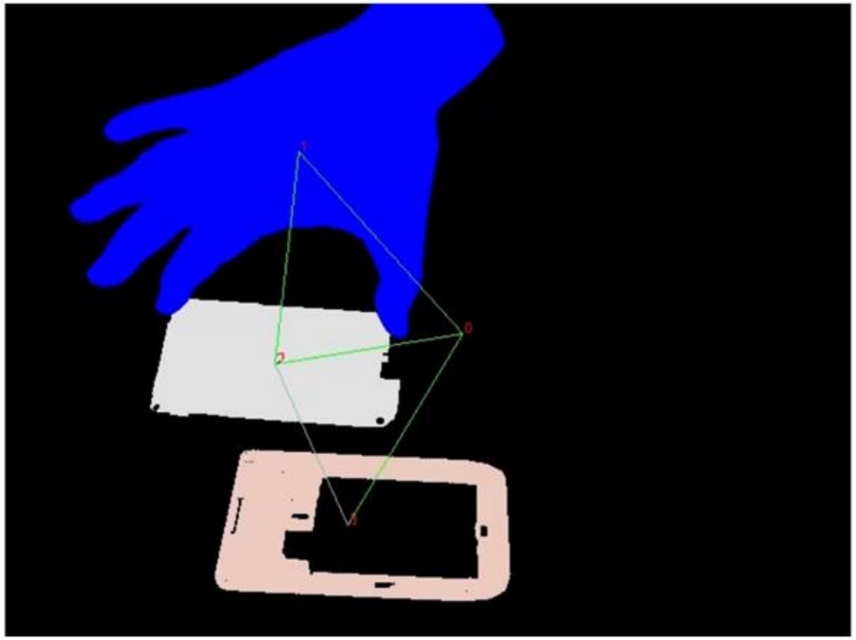
Parallel axes configurations can be detected in the 3D case

using only 2D information

method extended to
3D using ellipsoids



Extracted semantic graphs



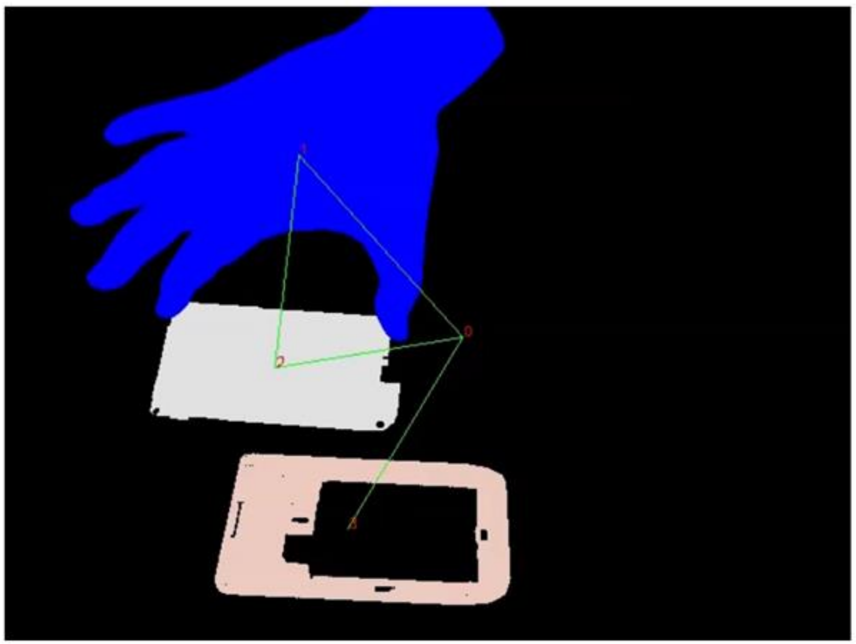
Extracted semantic graphs

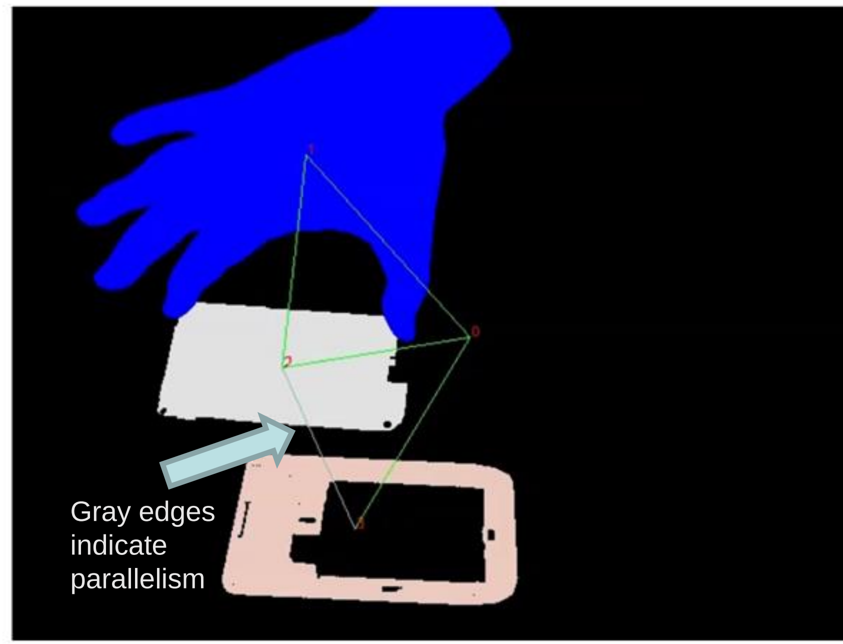
Learning by demonstration
automatic key-frame identification

Parallel axes configurations can be detected in the 3D case

using only 2D information

method extended to 3D using ellipsoids



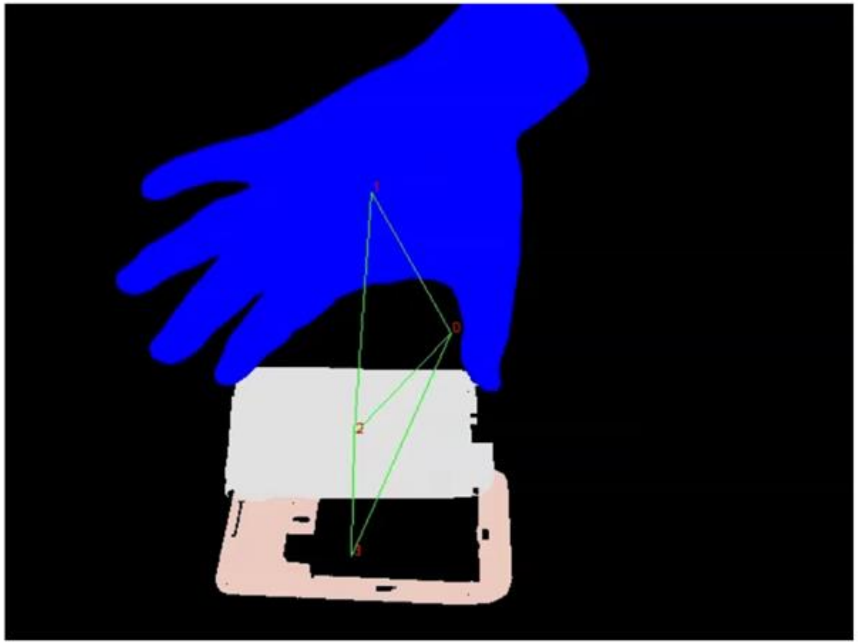
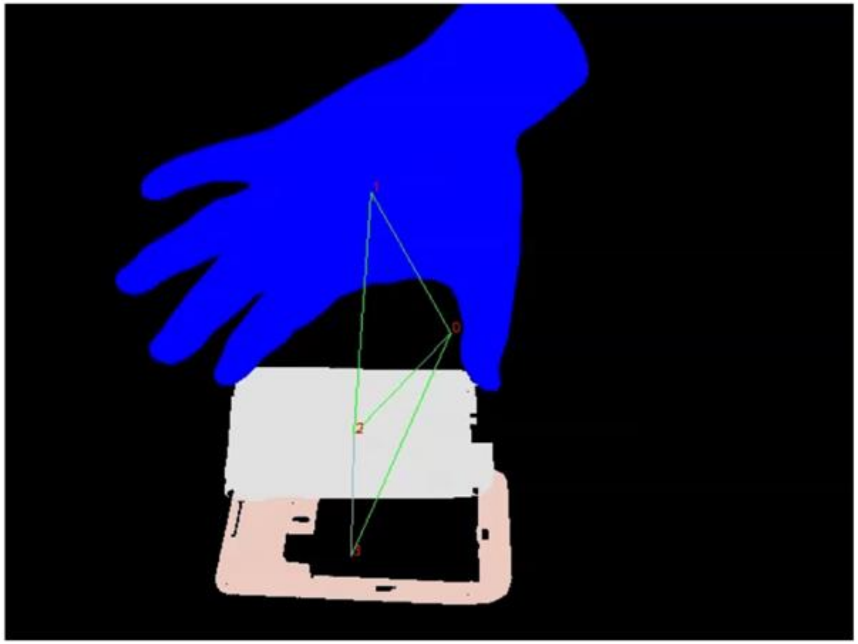


Extracted semantic graphs

Extracted semantic graphs

Learning by demonstration
automatic key-frame identification

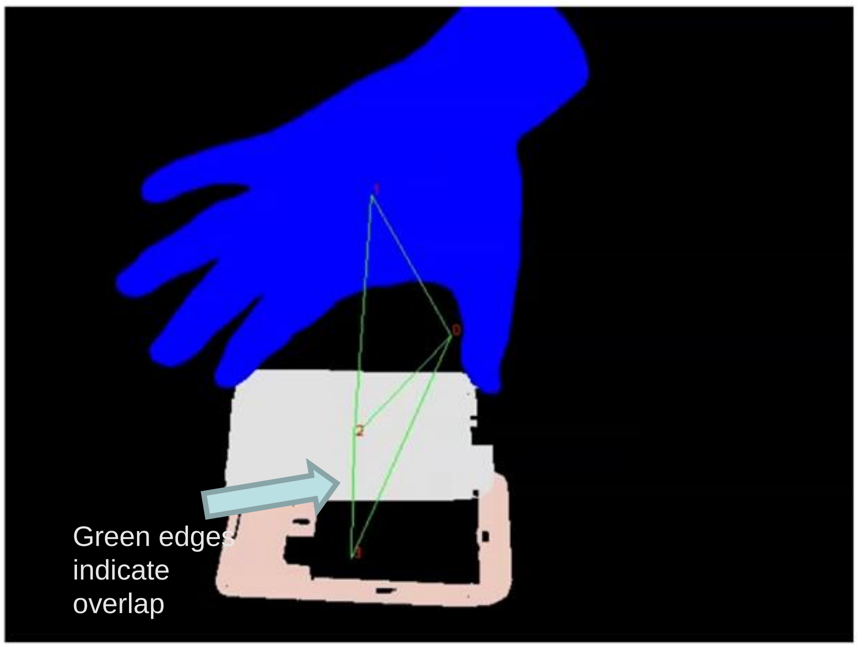
Erroneous touching or overlap detections can be avoided in the 3D case

using only 2D information	method extended to 3D using ellipsoids
	
Extracted semantic graphs	Extracted semantic graphs

Learning by demonstration
automatic key-frame identification

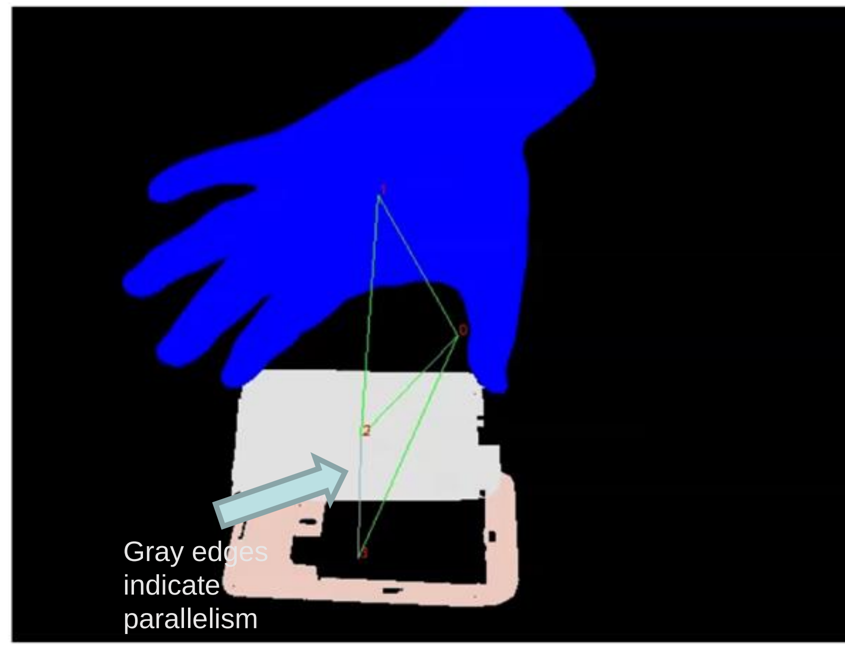
Erroneous touching or overlap detections can be avoided in the 3D case

using only 2D information



Extracted semantic graphs

method extended to
3D using ellipsoids



Extracted semantic graphs

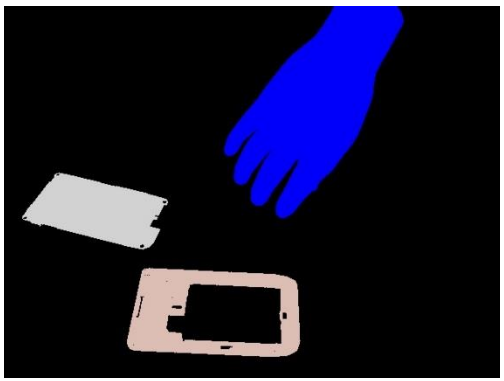
Information Technologies Institute

Learning by demonstration

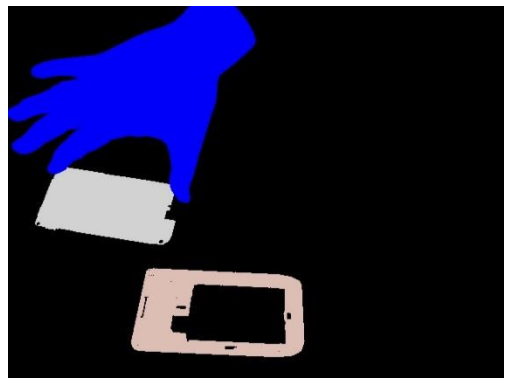
automatic key-frame identification

Automatically extracted key-frames in 3D based on changes of graphs

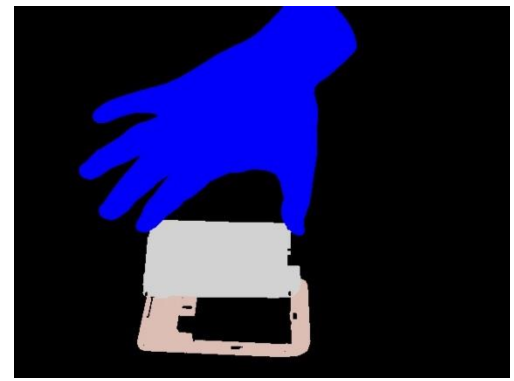
No touching



Hand - Obj1 touching



Obj1 – Obj2 1 axis parallel



Obj1 – Obj2 3 axes parallel

Obj1 – Obj2 overlap

Hand not touching

Learning by demonstration

automatic key-frame identification

Key-frame Extraction Evaluation

- 14 assembly demonstration experiments
  - 14 Key-frame sets automatically extracted
  - 14 Key-frame sets manually extracted
- Average number of extracted Key-frames
  - Automatic: 9.6
  - Manually: 7.3
- Average ratio r_{Kf} between automatically vs manually extracted Key-frames: 1.3
  - 2 additional Key-frames extracted by the system (object's orientation)
- Mean distance between manually and automatically selected key-frames
  - with respect to their position within each recorded sequence
  - Average distance of only **5.3** frames in a **183** frame sequence
- The system extracts almost the same amount of key-frame as the teacher
- The extracted frames are very close to those manually selected by the teacher

No.	Total Frames	Extracted key-frames			Mean Distance (std)	Mean Distance/Total Frames %
		Auto	Manual	r_{Kf}		
1	184	10	8	1.2	6.1 (10.0)	3.3
2	214	11	7	1.6	3.0 (4.8)	1.4
3	131	10	7	1.4	5.4 (7.3)	4.1
4	153	6	7	0.9	3.2 (3.9)	2.1
5	218	10	7	1.4	11.4 (19.0)	5.2
6	164	5	8	0.6	0.6 (1.2)	0.4
7	159	8	7	1.1	5.3 (11.0)	3.3
8	206	11	7	1.6	8.0 (12.8)	3.9
9	212	15	7	2.1	3.3 (6.3)	1.6
10	168	9	8	1.1	4.6 (8.6)	2.8
11	189	11	7	1.6	8.7 (13.8)	4.6
12	168	7	7	1	4.1 (7.9)	2.5
13	188	11	7	1.6	7.3 (11.2)	3.9
14	212	10	8	1.2	2.1 (4.3)	1.0
Total avg.	183	9.6	7.3	1.3	5.3 (10.3)	2.9

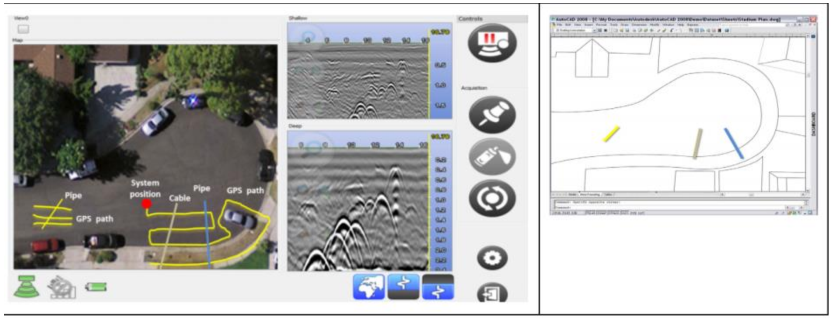
Applications

- Service robots **@field/construction sites**

Field service robots
@field/construction sites

- Autonomous subsurface mapping can be essential for many applications:
  - Landmine detection
  - Structured utilities detection
  - Buried infrastructures detection

- Current situation:
  - Semi-automatic procedure
  - Manual data collection
  - Data interpretation from experts
    - Semi-automatic annotation of the subsurface profile



Field service robots
@field/construction sites

- **Proposed approach [1]**
  - **3D underground mapping** with a **mobile robot** and a **GPR antenna array**

- Joint surface/subsurface mapping method overview
  - Surface
    - SLAM and constraints-based outdoor **path planning**, robot **navigation**
    - Graph-based stereo **visual odometry**
    - **General graph optimization (g2o)** for loop closure and localization refinements
  - Sub-surface
    - GPR data collection, signal pre-processing and **B-Scans formulation**
    - **Underground utility detection** through B-Scans processing
    - Underground map creation, **coupled with surface map**

*[1] G. Kouros, I. Kostavelis, E. Skartados, D. Giakoumis, A. Simi, G. Manacorda, D. Tzovaras, "3D Underground Mapping with a Mobile Robot and a GPR Antenna", **2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2018)**, Madrid, Spain, Oct 2018*

Vision-based Robot Navigation

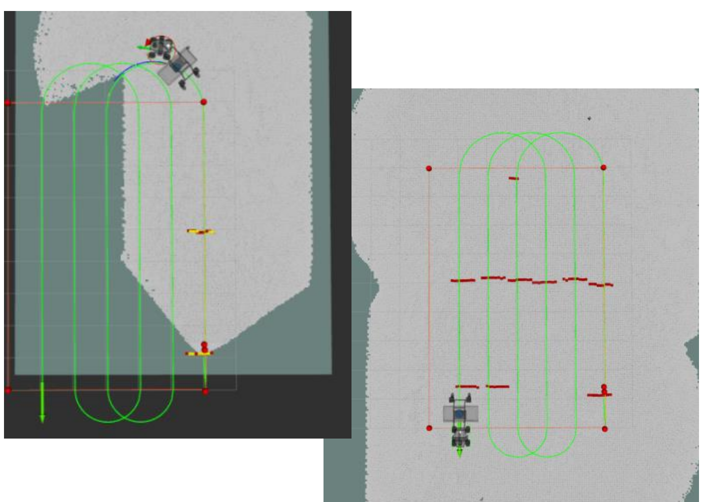
For outdoor/field robotics

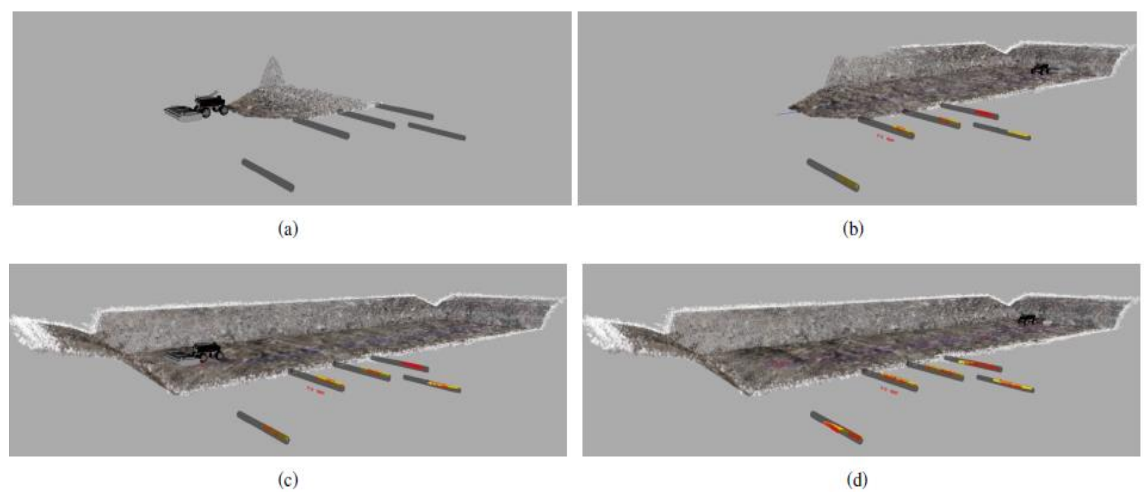
Step #1

- Navigation of a mobile robotic platform in outdoor environment
- **Aim: Autonomous robot path planning and navigation**

Core Technologies utilized

- Autonomous exploration of outdoor environment through
  - **SLAM** and constraints-based outdoor **path planning**
  - Stereo camera -based visual odometry for robot **localization**
  - Model Predictive Control (MPC) –based robot **navigation**





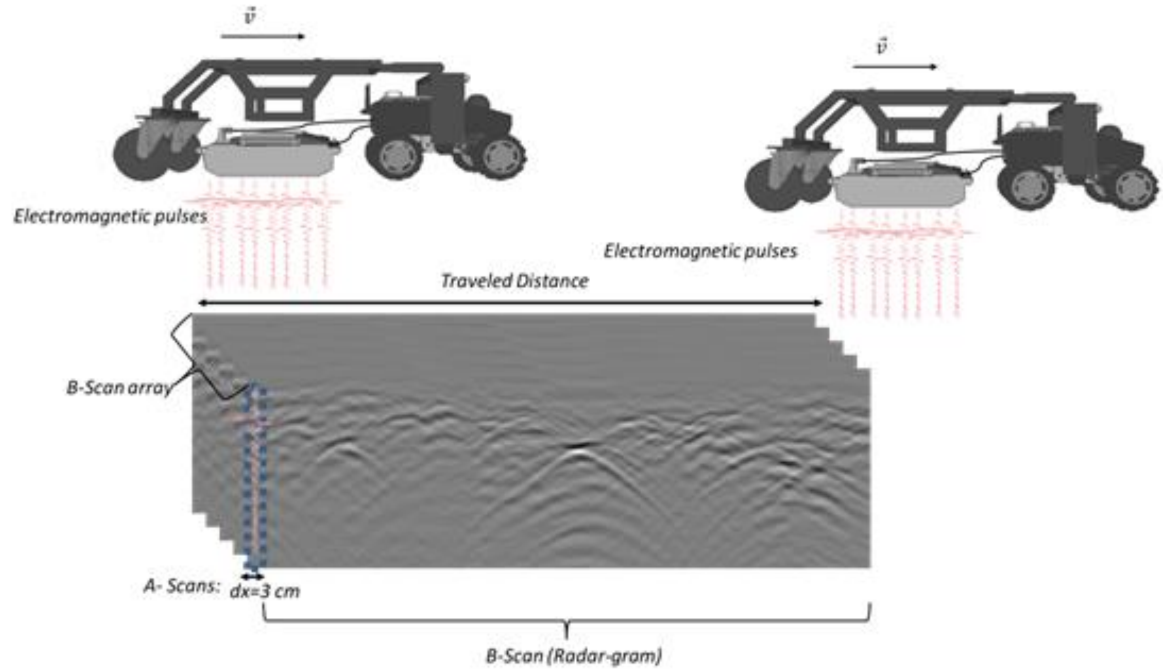
Underground mapping

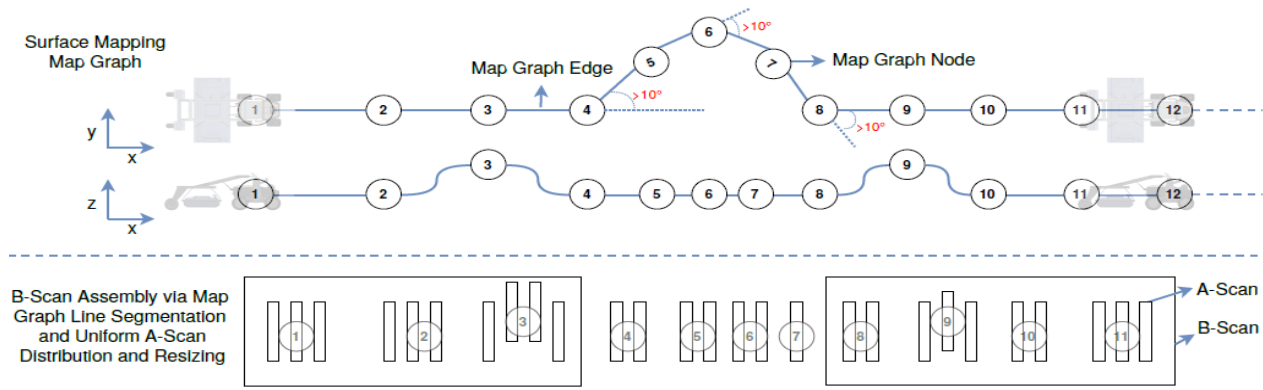
joint surface/subsurface mapping

- Along the rover motion, data are collected from the GPR antenna

- Collected A-Scans are registered to the localization graph

- B-Scan formulation corresponds to straight routes and is constrained by the robot's path

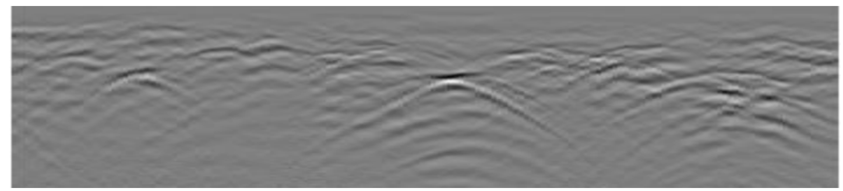


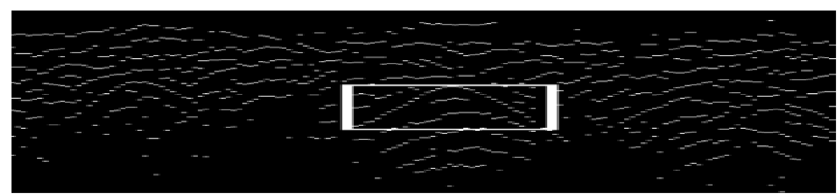


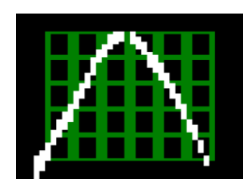
Underground mapping
underground utilities detection

Hyperbola patterns detection on B-Scans

- Two-step segmentation
- Isolation for salient regions
- Multidimensional HoG features
- SVM classification for hyperbola detection



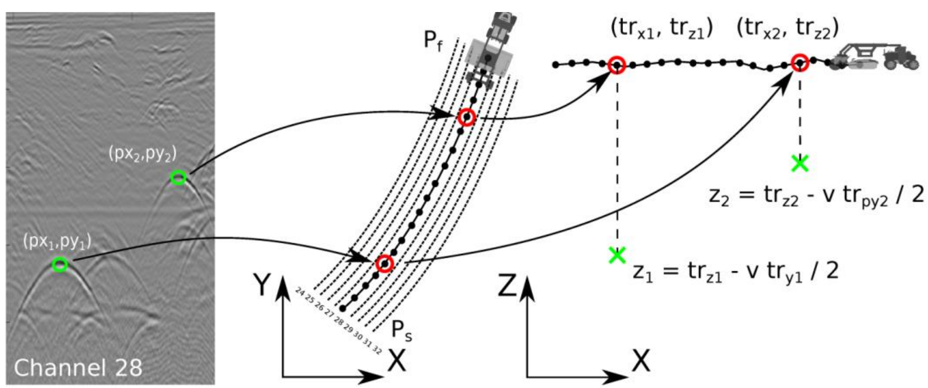




Underground mapping

3D reconstruction of underground environment

- Each A-Scan is registered to a node of the localization graph
- The depth of the detected apex is calculated using the propagation velocity ***v*** in the medium
- Apexes also inherit the transformation from the respective graph node
- The output is a sparse point cloud from the subsurface utilities

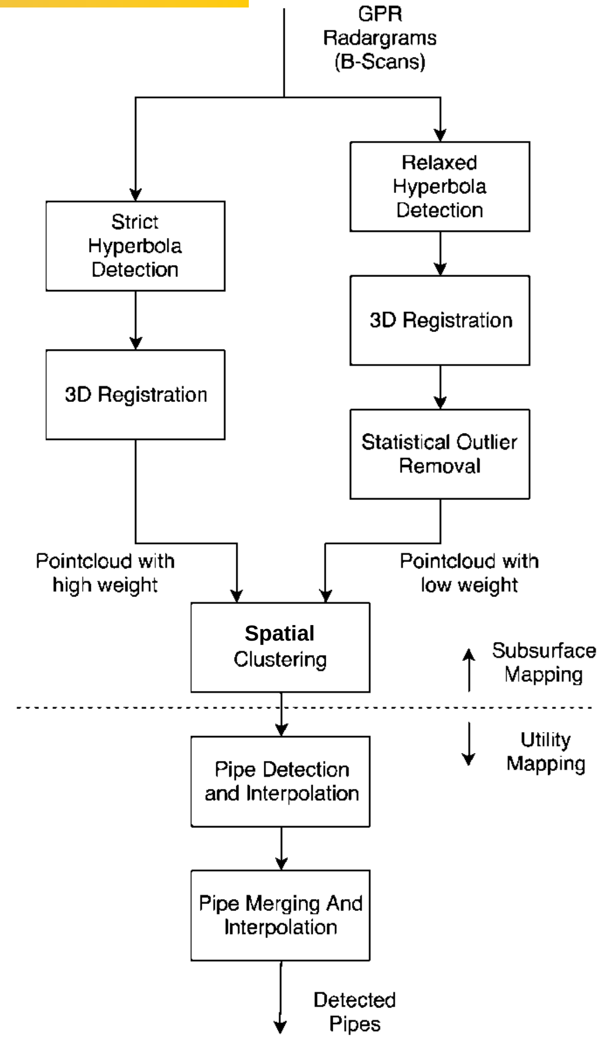


Underground mapping

underground utilities mapping

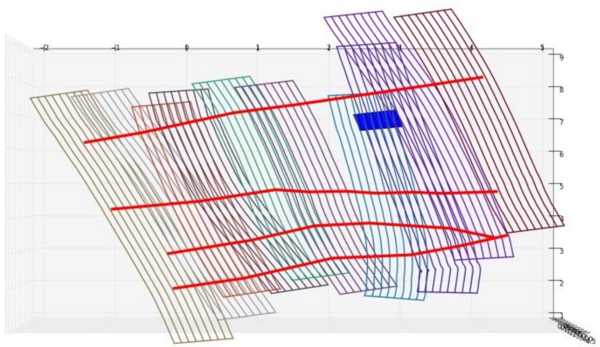
Identification of structured shapes in the subsurface

- Further processing on the point cloud stemming from the hyperbola apexes detection
- Outliers removal and density-based spatial clustering
- Registration with primitive geometrical shapes to isolate pipes, manholes etc.

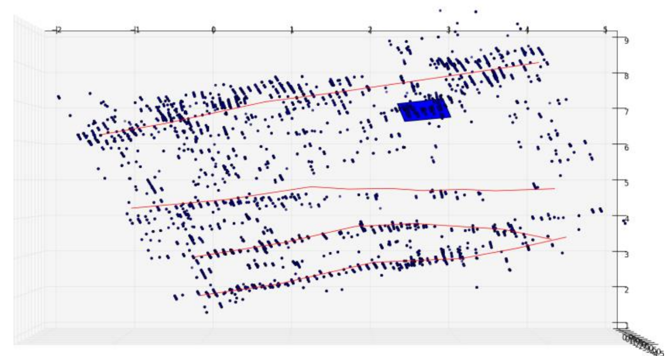


Underground mapping

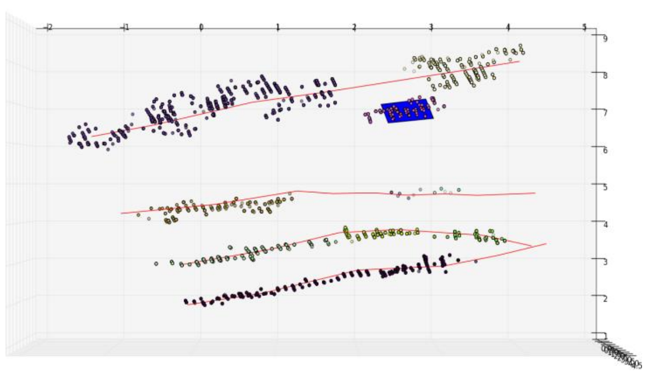
underground utilities mapping



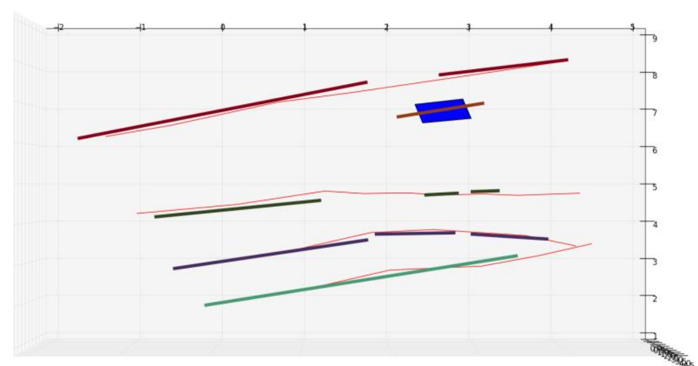
(a) *Robot trajectories; annotated pipes shown in red*



(b) *Detected hyperbolas*



(c) *Outliers removal & clustering*



(d) *Utility mapping*

Underground mapping

underground utilities mapping

BADGER Project: http://badger-robotics.eu/

*[1] G. Kouros, I. Kostavelis, E. Skartados, D. Giakoumis, A. Simi, G. Manacorda, D. Tzovaras, "3D Underground Mapping with a Mobile Robot and a GPR Antenna", **2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2018)**, Madrid, Spain, Oct 2018*

Conclusions

AI-Enhanced Computer Vision for Service Robots

- Future service robots are expected to provide assistance in a wide spectrum of diverse domains
  - **at home**, acting as personal, assistive service robots
  - in **agile manufacturing**, **exploration**, **construction** applications, and much more

- Technologies that advance **robot autonomy**, endorsing robots with better action and awareness are necessary
  - Robot perception, cognition, navigation and human-robot interaction capabilities play a key role in service robots
    - ***AI-enhanced computer vision techniques are key elements in this scope***

Conclusions

AI-Enhanced Computer Vision for Service Robots

- Key challenges for future service robots operating **in real homes, agile manufacturing, field/construction applications**
  - Environment mapping
  - Object recognition
  - Human tracking and activity recognition
  - Human-object interactions tracking for learning by demonstration
  - Affective human-robot communication
  - Autonomous navigation, localization and mapping
  - Robust, context-aware decision making

- AI-enhanced computer vision can help towards

  Methods robust enough for applications in real environments

Computer Vision and Signal Processing Techniques for Advanced Robotic Applications

Dr. Dimitrios Tzovaras

Dimitrios.Tzovaras@iti.gr