Developing systems with advanced perception, cognition, and interaction capabilities for learning a robotic assembly in one day

Dr. Dimitrios Tzovaras

Director of CERTH/ITI, Researcher Grade A'

Email: dimitrios.tzovaras@iti.gr

Teaching from Demonstration for Robotic Assembly Tasks

Problem Definition

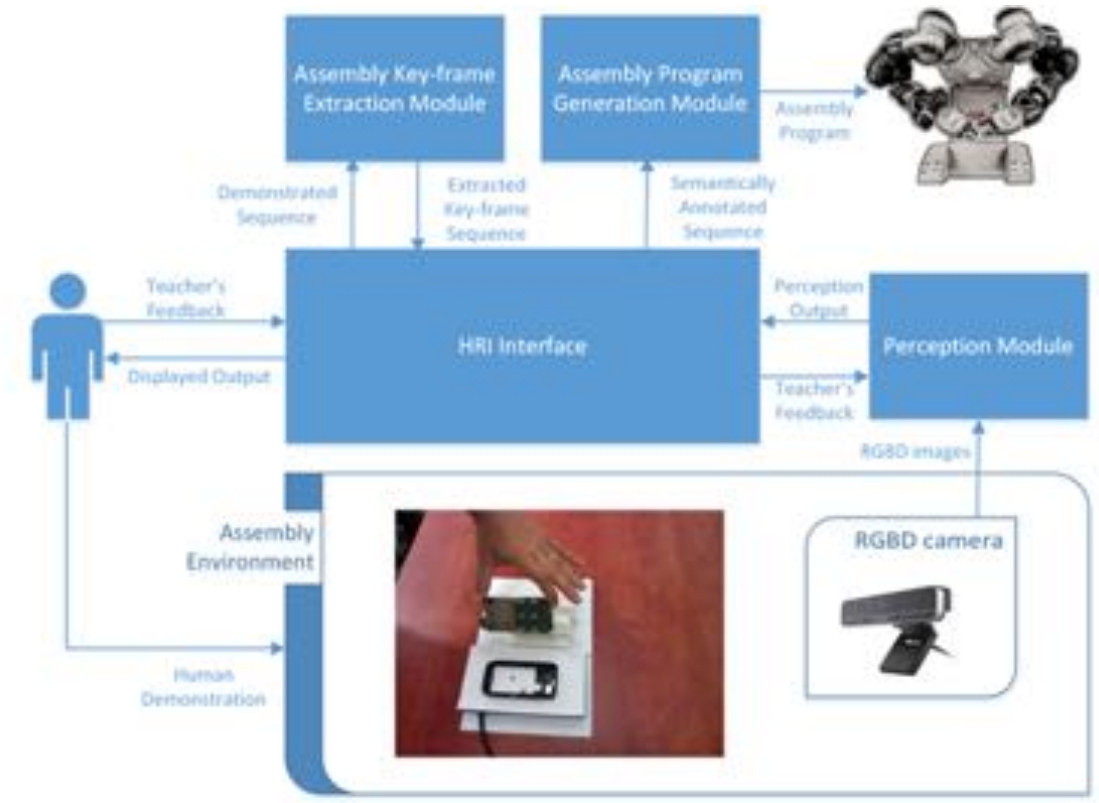
- Enable a non-expert user to teach a new assembly task to an industrial robot in less than a day
  - no explicit programming required

Motivation

- Even expensive products produced in large volumes are still assembled *manually* in low wage countries under *harsh conditions*

Approach

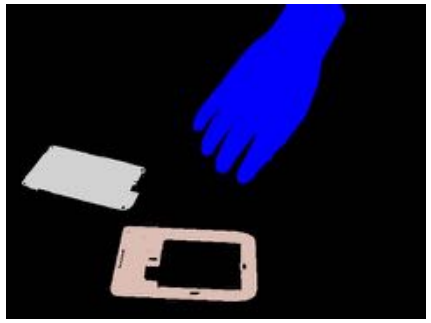
- Extend the robotic system with advanced perception and cognition abilities
- Develop a user-friendly Human Robot Interaction (HRI) interface
  - human operator demonstrates a task



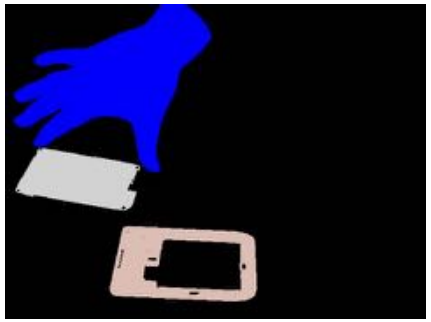
Overview of the proposed approach

Assembly Key-frame extraction: Automatic extraction

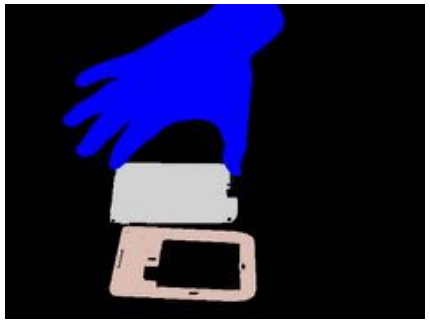
- Automatic Key-frame identification based on semantic graphs from image sequences[1]
- Employing 3D hand-object tracking results we can
  - automatically extract kinematics and motion information
  - perform more accurate and robust segmentation using 2D rendered images instead of watershed
  - extend the method to 3D data using ellipsoids to fit the object models[2] resulting to additional semantic relationships between the objects
- Implemented and tested in assembly video samples from RGBD data



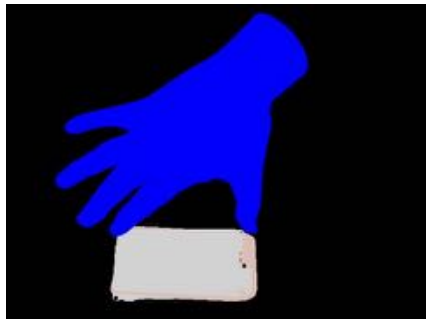
Key-frame 01



Key-frame 02



Key-frame 03



Key-frame 04

Segmented masks based on 2D rendered images constructed by the models of the tracked objects

1. *"Learning the semantics of object–action relations by observation." Int. Journal of Robotics Research 2011, Aksoy et al.*
2. *"Keyframe extraction with semantic graphs in assembly processes.", in IEEE Robotics and Automation Letters 2017, Piperagkas et al.*

Perception: Hand-Object Detection and Tracking in 3D

RGBD data are acquired

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

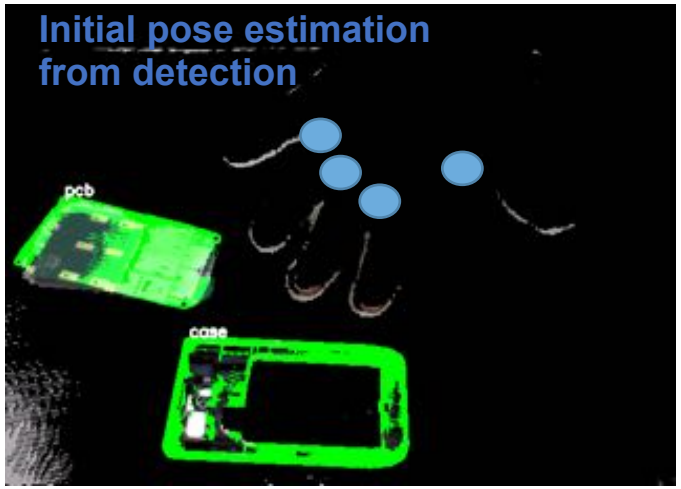
Hand-Object Tracking implementation using Particle Swarm Optimization (PSO)

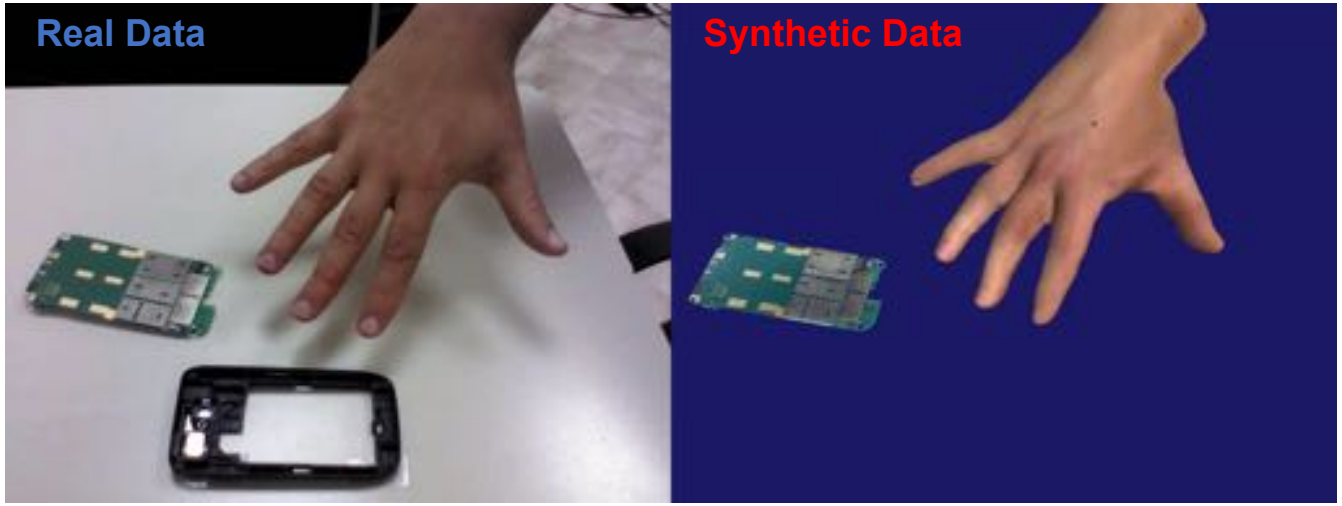
Based on hand tracking approaches in:

- "Real-Time Continuous Pose Recovery of Human Hands Using Convolutional Networks", Jonathan Tompson, Murphy Stein, Yann Lecun and Ken Perlin, SIGGRAPH'14
- "Efficient model-based 3d tracking of hand articulations using Kinect", I. Oikonomidis, N. Kyriazis and A. Argyros, BMVC 2014

Modified optimization for ***joint*** hand – object tracking

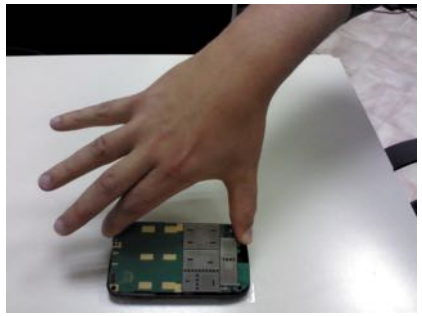
Optimization Time: 0.6 sec per frame





Assembly Key-frame Extraction: Definition of Key-frames



Key-frame information

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. *grasping*
- Automatic system suggestions, e.g. *aligned axes*

Dynamics information:

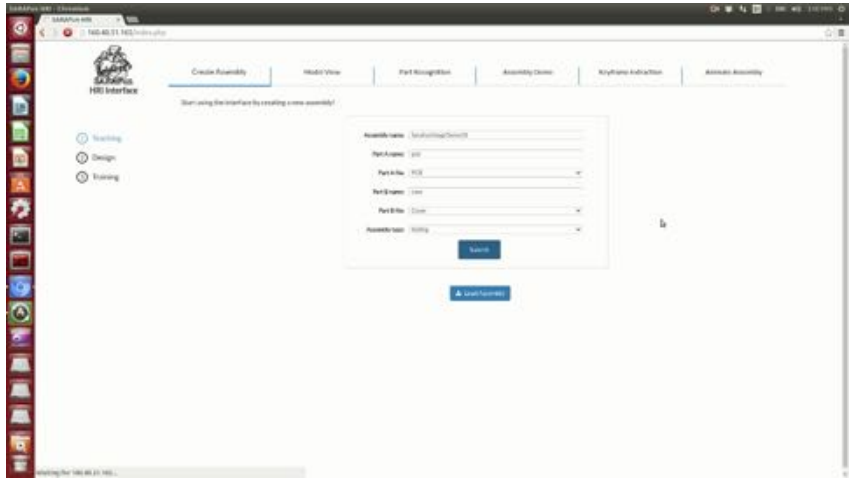
- Forces derived from the kinesthetic learning
- Grasping contact points
- Object deformation characteristics

XML format

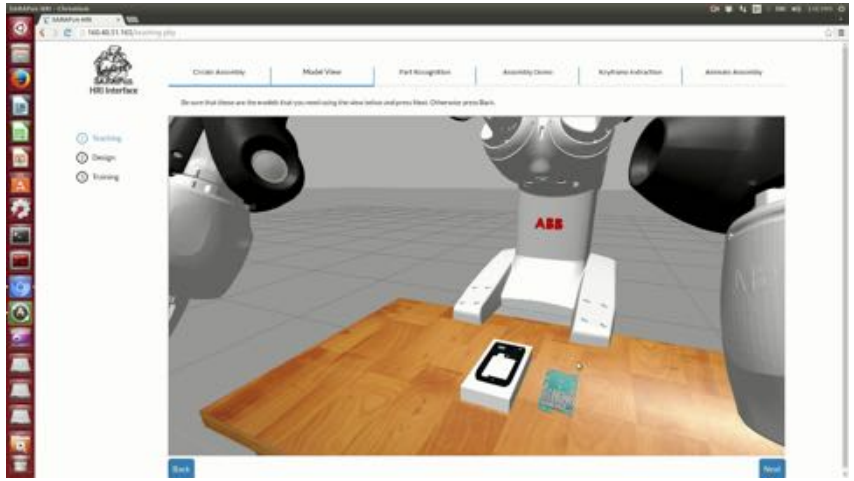
```
<?xml version="1.0" encoding="UTF-8" standalone="true"?>
- <KeyFrame xsi:schemaLocation="http://www.SARAFunXML.com
SARAFun_KeyFrame_XmlSpec_v02.xsd" xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance" xmlns="http://www.SARAFunXML.com" t="25.4" idx="1" id="0">
  - <CurrentAction id="assembly.mpg">
      <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">
```

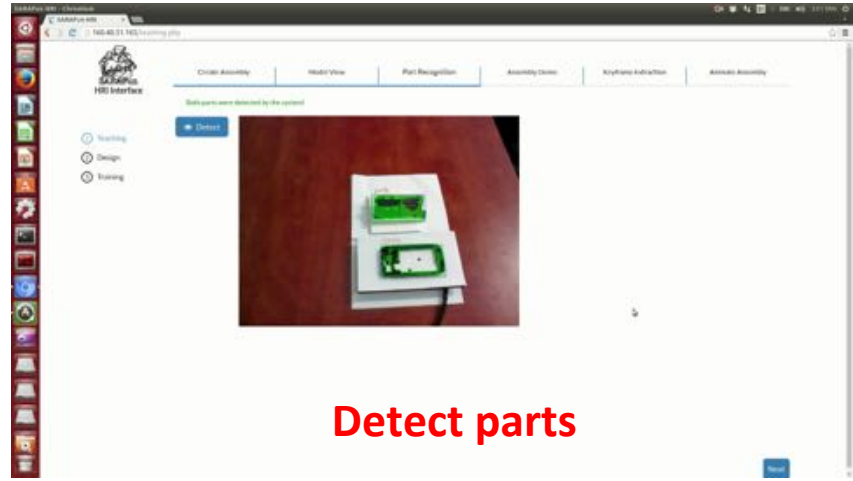
HRI interface : Teaching Example

Create new Assembly



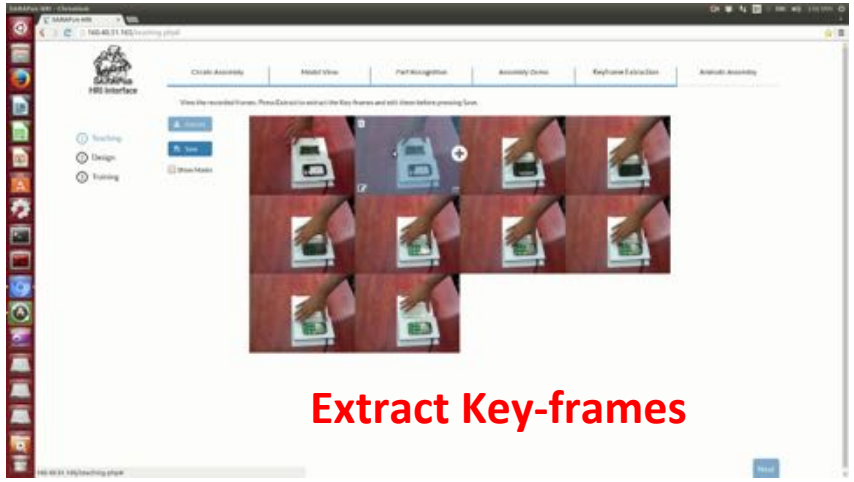
Preview CAD models



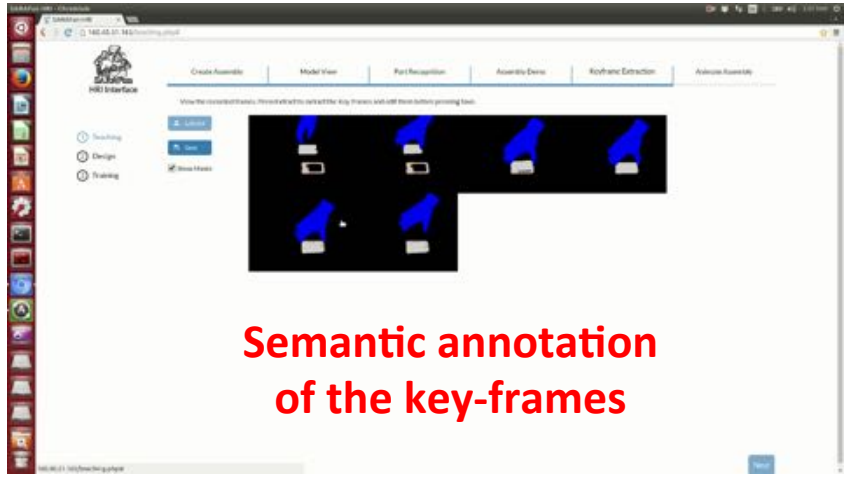


Detect parts

Detect hand and record assembly



Extract Key-frames



Semantic annotation of the key-frames

Assembly Program Generation

- A sequence of Key-frames is used for deriving an assembly program based on the associated semantic information
  - Sequential Function Charts
  - Behavior Trees
- Out of scope of this presentation

Future Work

- Test bi-manual assemblies
- Examine different types of assembly
  - Folding assembly
  - Insertion by deformation assembly
- Address assemblies with deformable parts

Thank you!