Engineering humanoids that grasp, learn from human and experience, and perceive time

Tamim Asfour

http://www.humanoids.kit.edu

Institute for Anthropomatics and Robotics, High Performance Humanoid Technologies
On Dualities, Force and Time in Robotics

Tamim Asfour

http://www.humanoids.kit.edu
My team

Humanoids@KIT
Chiara’s robot Tomy

- **Tomy**: 1200 parts, 7 motors, 250 EURO
- Tomy assembled by Chiara
- **Chiara**: 9 years old
- **Tomy’s skills**: speech interface, kinesthetic teaching, annotating motions sequences via speech, control via smart phone, upper body tracking and imitation, ... lots of fun!
Humanoids in the real world

- Engineering Humanoids
- Grasping and manipulation
- Learning for human observation
- Natural Interaction and communication

© SFB 588
ARMAR-III in the RoboKITCHen

45 minutes, more than 2000 times since February 3, 2008
Combining action, vision and haptics for grasping

Initial object hypotheses

Generate hypotheses based on Color, Geometric primitives and Saliency

Hypothesis 49 is chosen for verification by pushing
Integrating language, planning and execution with OACs
The ARMAR Architecture (inspired by Xperience)

- **High-level:**
  - Natural language understanding, reasoning and planning

- **Mid-Level:**
  - MemoryX: mediator between sensorimotor data and symbolic knowledge

- **Low-Level:**
  - Execution
  - Hardware Abstraction Layer (HAL)
  - ArmAR-YARP bridge
Task execution with OACs

- **OAC library** as part of long-term memory
- Each OAC consists of:
  - ID
  - Specific parameters
  - Preconditions for planning
  - Effects
  - Link to a hierarchical statechart
  - Statistics about execution
- **Instantiated OACs** in the working memory for the current task
NLU, Planning and Bootstrapping mechanisms

Speech Command

Language Understanding

Plan goal

Object affordances

Replacement Manager

Replace unknown objects

Find valid object locations

Replacement Strategies

Plan Execution Monitor

Object not found

Action execution fails

PKS planner

Replace

Execute

Replan

Adjusted goal
The ArmarX Software

- Event-driven component-based robot software development environment
- Open Source robot software development environment

Code and documentation

- Source code: https://gitlab.com/ArmarX
- Documentation: https://armarx.humanoids.kit.edu
Loco-manipulation tasks on WALK-MAN

- Semi-public demo at project review
- MultiSense SL stereo camera

Task: Remove the pipe
What's next?

- **SecondHands: A robot assistant for industrial maintenance**
  - 5 years project in Horizon 2020 (2015 – 2020)
  - Ocado, KIT, Sapienza, EPFL, UCL
- Provide help to maintenance technicians in a warehouse environment
- Advancement in the automation of the relatively unexplored domain of production machine maintenance
- Reduction of production machinery maintenance costs
ARMAR robot technologies in warehouses

Payload: ca. 11 kg
Weight: ca. 25 kg
Payload-to-weight: 0.44

<table>
<thead>
<tr>
<th>Joint</th>
<th>Max. Torque</th>
<th>speed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>176 Nm</td>
<td>79°/s</td>
</tr>
<tr>
<td>2</td>
<td>176 Nm</td>
<td>79°/s</td>
</tr>
<tr>
<td>3</td>
<td>176 Nm</td>
<td>79°/s</td>
</tr>
<tr>
<td>4</td>
<td>100 Nm</td>
<td>132°/s</td>
</tr>
<tr>
<td>5</td>
<td>100 Nm</td>
<td>132°/s</td>
</tr>
<tr>
<td>6</td>
<td>100 Nm*</td>
<td>132°/s</td>
</tr>
<tr>
<td>7</td>
<td>34 Nm</td>
<td>206°/s</td>
</tr>
<tr>
<td>8</td>
<td>34 Nm</td>
<td>206°/s</td>
</tr>
</tbody>
</table>
ARMAR robot technologies in warehouses
Maintenance objects/tools

- Object/tools models
  - AllanKey.xml
  - AllanKey2.xml
  - AllanKey3.xml
  - Cutter.xml
  - Flashlight.xml
  - Screwdriver-Red-smaller.xml
  - Screwdriver-cross.xml
  - Wrench.xml
  - Pliers.xml
  - ...

- See KIT object database
  http://object-database.humanoids.kit.edu
Learning from human observation

Reproduction of wiping DMPs encoding a transient and a periodic pattern on ARMAR-IIIb
Learning from observation – prepare the dough

"ARMAR, could you please prepare the dough for me?"
KIT whole-body human motion database

https://motion-database.humanoids.kit.edu/

Conversion of Human and Object Motions with the MMM Framework
The KIT whole-body human motion database
mocap → MMM → robot model → real robot
ARMAR-IV: Mechano-Informatics

- Torque controlled
- 3 on-board embedded PCs
- 76 Microcontroller
- 6 CAN Buses

- 63 DOF
  - 41 electrically-driven
  - 22 pneumatically-driven (Hand)

- 238 Sensors
  - 4 Cameras
  - 6 Microphones
  - 4 6D-force-torque sensors
  - 2 IMUs
  - 128 position (incremental and absolute), torque and temperature sensors in arm, leg and hip joints
  - 18 position (incremental and absolute) sensors in head joints
  - 14 load cells in the feet
  - 22 encoders in hand joints
  - 20 pressure sensors in hand actuators
  - ...

More than mechatronics

ARMAR-IV
made@KIT
70 kg
170 cm
ARMAR-IV

- 63 DOF
- Torque-controlled!

Multi-contact active compliance balancing controller
Duality
Duality - Boolean Algebra

\[\land \leftrightarrow \lor \quad 0 \leftrightarrow 1\]
\[a \leftrightarrow a \quad \overline{a} \leftrightarrow \overline{a}\]

\[a \land a \quad \overset{H3}{=} \quad (a \land a) \lor 0\]
\[a \lor a \quad \overset{H3}{=} \quad (a \lor a) \land 1\]
\[a \land (a \lor \overline{a}) \quad \overset{H4}{=} \quad (a \lor a) \land (a \lor \overline{a})\]
\[a \lor (a \land \overline{a}) \quad \overset{H4}{=} \quad (a \lor a) \land (a \lor \overline{a})\]

\[a \land 1 \quad \overset{H4}{=} \quad a \lor 0\]

\[a \quad \overset{H4}{=} \quad q.e.d.\]
\[a \quad \overset{H3}{=} \quad q.e.d.\]
The duality of grasping and balancing

Equilibrium is reached by balancing similar sets of forces

Ground reaction forces ↔ Fingertip forces
Weight of the body (CM) ↔ Weight of the object (CM)
Torques on the joints ↔ Torques on the joints
The duality of grasping and balancing

Concepts of grasping can be applied to loco-manipulation

\[ G^T T = J_H \dot{\Theta} \]
\[ J_H^T \lambda_f = \tau \]
\[ -G \lambda_f = W \]
\[ \lambda_f \in \mathcal{F} \]

Balance  \( \leftrightarrow \)  Stable grasp
Step planning  \( \leftrightarrow \)  Grasp synthesis
On the Duality of grasping and balancing

- Selection of support pose
- Selection of contact points
- Classification of support poses possibilities
- Grasp selection
- Grasp synthesis
- Grasping taxonomies

Applications of grasping taxonomies
- Benchmark to test robot hand abilities
- Simplify grasp synthesis
- Inspire hand design
- Optimization of synergies: Formulation of dexterity/functionality as number of achievable grasps for maximization
- Guide autonomous grasp selection

M. R. Cutkosky, 1989
N. Kamakura, 1989
T. Feix et al, 2009
Bollock et al. 2013
Taxonomy of whole-body poses

18 standing poses
18 kneeling poses
10 resting poses

Total: 46 classes

Borras and Asfour, IROS 2015
Taxonomy of whole-body poses

Contact types:
- Plane: (hand or foot)
- Line: (arm or knee)
- Hold: (hand)

Single foot support:
- Single knee support:
- Double feet support:
- Double knee support:
- Combined double support:

Humanoids@KIT

Borras and Asfour, IROS 2015
Taxonomy of whole-body poses

Lines represent transitions

Contact types:
- Plane (hand or foot)
- Contact (arm or knee)
- Hold (hand)

Support types:
- Single foot support
- Single knee support
- Double feet support
- Double knee support
- Combined double support

Support categories:
- Less stable - More mobility
  - Single
  - Double
  - Triple
  - Quadruple

Area of contact:
- Less area of contact
- More area of contact

States:
- Standing
- Kneeling
- Resting
Validation of the taxonomy

- Analyses of different human loco-manipulation tasks with supports
- Reference model of the human body (Master Motor Map: MMM) with 104 DOF
- Motion capture data mapped to reference model of the human body (MMM)
- Automatic segmentation to detect support poses and transitions
- Automatic generation of a taxonomy of the poses and their transitions in the motion data
Analysis of pose transitions

Going upstairs with a handle

Detection of support contacts highlighted in red

Subject swings left leg with a right foot – right hand support pose
Analysis of whole-body loco-manipulation tasks

Walking Upstairs
Data-driven validation of the taxonomy

- Total of **121** motions processed
  - **Locomotion**
    - Upstairs/downstairs with handle
    - Walk with handle
    - Walk avoiding obstacles using hand supports
  - **Loco-manipulation**
    - Lean to reach/place/wipe
    - Bimanual pick and place of big objects
  - **Balancing**
    - push recovery
    - recovery due to lost balance
  - **Kneeling motions**
    - 4.5% of poses missed (all double foot supports (the looping edges))
Whole-body motion based on the taxonomy

- **n-gram language model**: Statistical approach to learning conditional transition probabilities between whole-body shape poses.
Software and documentation: MMM, Motion DB

- KIT Whole-Body Motion Database
  - https://motion-database.humanoids.kit.edu

- MMM:
  - https://gitlab.com/mastermotormap/mmmcore
  - https://gitlab.com/mastermotormap/mmmtools

- Dokumentation:
  - http://mmm.humanoids.kit.edu
  - https://motion-database.humanoids.kit.edu/faq

- KIT Object database
  - http://h2t-projects.webarchiv.kit.edu/Projects/ObjectModelsWebUI/
Lessons learnt in 16 years

Robotics is the science of integration

The “X” in robotics

- It is not the “X” in Self-X (self-organization, self-repair, self-refinement, ...)
- It is not the “X” in Co-X (co-habiter, co-worker, co-protector, ...)
- It is not the state variable in dynamical systems

Unfortunately, it is the value by which we have to speed up robot movies to make robots behave/move in a human-like way

\[ X > 1 \]
It’s all about Force

Force
Role of force

- Force is key element for interaction with the physical world.

- Human infant motor control studies also indicate that early manipulation relies on contact and force, with other senses being incorporated in control later in the development.

Claim:

- Objects, agents and their actions can be described based on a new concept of sensorimotor force fields (SFF) that provides a unified representation and computational mechanism for solving robotics tasks (grasping and manipulation, balancing, ...).

- SSF result from the integrated mapping of action and sensory modalities such as position, pressure, tactile, audio and vision sensory data to the force space.
Action and agent

Action
- Action represented by the force fields that generate it
- Dynamic systems; attractor landscapes

Agents = Embodiment
- Sensorimotor maps of the body schema; tool use
- Based on proprioception (and vision and haptic)
Perception - Physical laws

- Duality between force and position has been demonstrated in the robotics in the form of position-force control mechanisms (Newton’s law)

\[ F = m \cdot \ddot{x} \]

- Pressure is the amount of force acting perpendicularly per unit area

\[ F = A \cdot P \]

- Haptic: contact, pressure, proprioception, temperature, vibration
  - Superposition?

- Audio: loudness proportional to force (e.g. knocking)
Perception - Physical laws

- Vision
  - Depth $\rightarrow$ position
  - Colors
  - Intensity
  - Saliency
  - Attention
  - Motion $\rightarrow$ see action
  - Shape features
  - ....
Sensorimotor Force Fields (SFF)

From X to force and torque!
From pixels, voxels, taxels, ... to forces and torques
SFF - force4all

- Co-joint Object-Action representation in the force space (force-based OACs) → Robot “machine code” in the force space

Research questions:
- Definition of laws and rules for mapping of different sensory modalities into SFF.
- Mathematical and algorithmic modeling of SSFs.
- Operators and arithmetics for SFF: interaction of different SSFs resulting from different sensory modalities or action.
- Formulation of robotics tasks based on the SSF representation
- Compilers from natural language task description to the force space
- Which robotics tasks? Grasping, Balancing, ...
It is not only about Newton forces

- Mental forces
- Logical forces
- Theoretical forces
- Physical forces
- ...
Force in Japanese culture

force 理力
心理 mental
論理 logical
理論 theoretical
Force-based Human-EXO interface

- Feel the muscle activation (non invasive)
- Learn human-suit interaction force pattern and use them for motion prediction
ARMAR-5: Interface to the human body
Other examples

- SFFs for grasp recognition and reproduction

- A. Kheddar (CNRS/LIRMM) and A. Argyros (FORTH)
  - Towards Force Sensing From Vision: Observing hand-object interactions to infer manipulation forces, CVPR 2015

- Gentiane Venture
  - Emotion recognition based on force
Important

- It is not only about the EXO interfaces
- It is not only about physical forces (contact forces, ....)
- It is about the force space as unifying representation for sensorimotor experience and cognitive capabilities
Time
Time is vital

Knowing
- Knowledge hierarchies
- Episodic memory (what, where, when), forgetting
- Time-based: Past recall, future imagination

Doing
- Short-term: Fluency in HRI (e.g. turn taking)
- Long-term: constraints in action planning, habits.
- Multiple tasks coordination

Being
- Self identity over time
- Low level consciousness: perceive internal, environment changes
- High-level consciousness: link self to historical times

This is who I am!
Time in Robotics

- **Past**
  - Experience

- **Present**
  - Current world state

- **Future**
  - Prediction

- Time is fundamental for the implementation of **episodic memories**
KIT Manipulation Action Dataset

In total 70 demonstrations of 8 different manipulation actions

- cut
- mix
- drink
- pick-place
- pour
- put-in
- put-on
- take-down

10 Different Objects
Level I: Semantic Segmentation

Hierarchical Segmentation

Hand-Object Relation: No Contact | Bowl in hand | No Contact

Object-Object Relation: Approach | Lift-up | Place | Release | Withdraw

Semantic Distance Profile

- bowl to human motion
- bowl to sponge
- bowl to table
- sponge to table
Level II: Motion Segmentation

Hierarchical Segmentation

Hand-Object Relation:  No Contact  Bowl in hand  No Contact

Object-Object Relation:  Approach  Lift-up  Place  Release  Withdraw

Trajectory of the Bowl
Perception of Time: Put-on Action

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.9</td>
<td>2.1</td>
<td>2.8</td>
<td>1.7</td>
<td>1.8</td>
<td>1.4 sec</td>
</tr>
<tr>
<td></td>
<td>0.4</td>
<td>1.0</td>
<td>0.9</td>
<td>0.4</td>
<td>0.7</td>
<td>2.5 sec</td>
</tr>
<tr>
<td></td>
<td>0.4</td>
<td>0.6</td>
<td>1.2</td>
<td>0.4</td>
<td>1.2</td>
<td>3.1 sec</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>0.9</td>
<td>1.0</td>
<td>0.1</td>
<td>1.1</td>
<td>3.0 sec</td>
</tr>
<tr>
<td></td>
<td>2.8</td>
<td>3.2</td>
<td>3.0</td>
<td>2.1</td>
<td>3.1</td>
<td>Mean: 2.1 sec, Std: 0.9</td>
</tr>
</tbody>
</table>

Δt
Perception of Time: Human Demonstration

![Diagram showing time in seconds for different actions such as cut, pour, drink, put-on, take-down, put-in, pick & place, and mixing.]
Perception of Time: ARMAR-4 Imitation
Perception of Time: Psychological Experiments

- Psychological experiments support our new semantic action segmentation hypothesis
- Collaboration with the University of Groningen (Hedderik van Rijn, Experimental Psychology & Statistical Methods and Psychometrics)

Temporal Scaling

Semantic Segmentation (Level I)

Motion Segmentation (Level II)

Human Demonstration

R (spoon, bowl): Inside
R (spoon, bowl): Above

Key Frame: 3.7
Key Frame: 8.76

Semantic Segments
Position in mm

Time in seconds

400
200
-400
0
2
4
6
8
10
12
14

X
Y
Z

Sub-segments

Trajectory Dictionary

Based on the Dynamic Time Warping Distance Measure

Trajectory Histogram

Periodic Pattern

Humanoids@KIT

CLASS-WISE AVERAGE PERIODICITY MEASURES.

<table>
<thead>
<tr>
<th>Class</th>
<th>Stir</th>
<th>Pick</th>
<th>Place</th>
<th>Put In</th>
<th>Take Down</th>
<th>Put On</th>
<th>Drink</th>
<th>Pour</th>
<th>Cut</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.67</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.5</td>
</tr>
</tbody>
</table>
Temporal Scaling

**Enriched Manipulation Action Semantics for Robot Execution of Time Constrained Tasks**

Eren Erdal Aksoy, You Zhou, Mirko Wächter and Tamim Asfour

Institute for Anthropomatics and Robotics - High Performance Humanoid Technologies Lab (H2T)
Breakthroughs in robotics since ~2000 – my view

- Progress driven by

- „Cool“ new hardware
  - Robot mechatronics:
    - DLR/KUKA LWR, NAO, UR, iCub, youBot, FRANKA EMIKA, ... 

  - Sensors:
    - Kinect, ...

  - Computing power:
    - many-core systems, GPUs, ...

- Large amount of data (thanks to better hardware)
Thanks to ...

- **German Research Foundation (DFG)**
  - SFB 588  www.sfb588.uni-karlsruhe.de (2001 - 2012)
  - SPP 1527  autonomous-learning.org (2010 -)
  - SFB/TR 89  www.invasic.de (2009 -)

- **European Union**
  - IMAGINE  (2017-2020)
  - GRASP  www.grasp-project.eu (2008-2012)

- **Federal Ministry of Education and Research (BMBF)**
  - INOPRO  (2016-2021)

- **Karlsruhe Institute of Technology (KIT)**
  - Professorship “Humanoid Robotic Systems”
  - Heidelberg-Karlsruhe Research Partnership (HEiKA)
Thanks for your attention

May the force be with you!