The Karlsruhe ARMAR Humanoid Robot Family

Tamim Asfour, Rüdiger Dillmann, Nikolaus Vahrenkamp, Martin Do, Mirko Wächter, Christian Mandery, Peter Kaiser, Manfred Kröhnert and Markus Grotz

Abstract We present the development and evolution of the ARMAR humanoid robots, which have been developed to perform grasping and manipulation tasks in made-for-human environments, to learn actions and task knowledge from human observation and sensorimotor experience and to interact with humans in a natural way. We describe the mechatronics of the ARMAR robots, their grasping and learning capabilities and the underlying architecture.

1 Introduction

Humanoid robotics is an emerging and challenging research field, which has received significant attention during the past years and will continue to play a central role in robotics research and many applications of the 21st century. Regardless of the application area, one of the common problems tackled in humanoid robotics is the understanding of human-like information processing and the underlying mechanisms of the human brain in dealing with the real world. Considerable progress has been made in humanoid research resulting in a number of humanoid robots able to move and perform well-designed tasks. Over the past decade in humanoid research, an encouraging spectrum of science and technology has emerged that leads to the development of highly advanced humanoid mechatronic systems endowed with rich and complex sensorimotor capabilities. This includes WABOT-1, WABIAN-2 [52], ASIMO [59], HRP-2[38], HRP-4C [40], HRP4 [39], KOBIAN [86], Twendy-One [33], ARMAR [8], iCub [51], DB [11], CB [21], HUBO [46], Justin [83] and TORO [53], NAO[1], Toyota's partner robot[2], LOLA [47], REEM [3], DARwIn-

The authors are with the Institute for Anthropomatics and Robotics, Karlsruhe Institute of Technology, Germany, e-mail: asfour@kit.edu

[1] Aldebaran: www.aldebaran-robotics.com
[2] Toyota: http://en.wikipedia.org/wiki/Toyota_Partner_Robot
[3] PAL-Robotics: http://pal-robotics.com

OP[4], Robonaut [2], PETMAN and ATLAS[5], WALK-MAN [67], Valkyrie[6], and OceanOne [43] to name a few. Of major importance for advances of the field is without doubt the availability of reproducible humanoid robot systems such as HRP-2, NAO, iCub, HUBO and ATLAS which have been used in the last years as common hardware and software platforms to support humanoids research. Ambitious goals have been set for future humanoid robotics. They are expected to serve as companions and assistants for humans in daily life and as ultimate helpers in man-made and natural disasters. In 2050, a team of humanoid robots soccer players shall win against the winner of most recent World Cup [7]. DARPA announced in 2012 the next Grand Challenge in robotics: building robots which do things like humans in a world made for humans [8]. The research roadmap for humanoid robotics in the next years could be structured around two major groups of disciplines: driving disciplines which provide the ideas and motivation for research, and technological disciplines, which provide the necessary enabling technologies.

One motivation for building robot systems with human-like body morphology, i.e. humanoid robots, lies in the desire and need for versatile robot systems, which – similar to humans – are able to act and interact in made-for-human environments, to grasp and manipulate made-for-humans objects and tools. Another motivation is the fact that robots with human-like behaviors will contribute to more intuitive and fluent human-robot interaction as humans will better predict the robot's action if the robot behaves in a human-like way. From our perspective, the important research directions and challenges toward complete humanoids robot systems are:

- Engineering high performance humanoid mechatronics with human like performance regarding dexterity, speed, force capabilities, energy efficiency and compliance.
- Strategies for the realization of versatile grasping and manipulation capabilities in human-centered environments.
- Methods for learning motion primitives and task knowledge from human demonstration and sensorimotor experience to facilitate intuitive robot programming.
- Implementation of cognitive architectures which integrate perception, action, planning, learning and interaction.

In the following we describe our progress in the area of humanoid robotics research and development and present the ARMAR humanoid robot family and their capabilities regarding grasping and manipulation, learning from human observation and the underlying software architecture.

[4] Robotis: `http://www.robotis.com/xe/darwin_en`
[5] DARPA: `https://en.wikipedia.org/wiki/DARPA_Robotics_Challenge`
[6] NASA: `https://www.nasa.gov/feature/valkyrie`
[7] `RoboCup:http://www.robocup.org/a_brief_history_of_robocup`
[8] DRC: `https://en.wikipedia.org/wiki/DARPA_Robotics_Challenge`

2 The ARMAR robots

The ARMAR robots were developed within the Collaborative Research Center 588: Humanoid Robots - Learning and Cooperating Multimodal Robots (SFB 588), a large-scale research project funded by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) from 2001 until 2012 ([23]). All robots were realized as humanoid robot systems with all the perceptual and computational components which are needed to perform complex grasping, manipulation and learning tasks in a kitchen environment without any external sensors and without any off-board computing power.

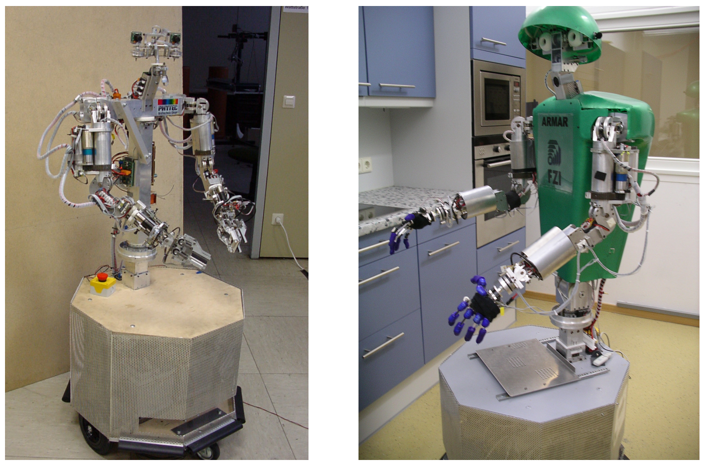

Fig. 1: The humanoid robot ARMAR-I[6], ARMAR-II[7].

In 2000, the first humanoid robot in Karlsruhe was built and named ARMAR [6], later ARMAR-I, which has 25 degrees of freedom (DoF) with two anthropomorphic arms each having 7 DoF, a head with 3 DoF equipped with a stereo camera system, two simple jaw grippers and a mobile wheel-driven platform with differential kinematics and a maximum velocity of about $1\,\mathrm{m/s}$. A torso with 4 DoF allows bending forward/backward and sidewards as well as a yaw rotation of about 330°. With a prismatic joint in the torso it was possible to increase the total height of the robot by $40\,\mathrm{cm}$. The weight of the robot was about $75\,\mathrm{kg}$ including batteries. The goal was to develop a humanoid robot, which is able to share its activity space with a human partner. A kitchen environment was selected as an application area. In ARMAR-I, position based control strategies, basic navigation using a 2D laser scanner and early motion coordination strategies between the different sub-systems were implemented. The control system of the robot was divided into the five subsystems head, left arm, right arm, torso and mobile platform, whereas each subsystem had its own software and hardware controller module, which were implemented in the Modular Controller Architecture (MCA) framework[9].

In 2002, the second version of the ARMAR series, namely ARMAR-II, was built [7, 24]. The kinematics in terms of body morphology and number of DoF was similar to ARMAR-I. The major difference was the replacement of the jaw-grippers with pneumatically driven five-finger fluid hands with 8 DoF each, which were originally developed as prosthesis hands [64, 65]. In addition, 6 DoF force/torque sensors were integrated in the wrists to allow the implementation of force and position-based control strategies. A three-level control architecture was developed to facilitate integration on conceptual and software integration. Within the architecture, a given task is decomposed into several sub-tasks, representing the sequence of actions the subsystems of the robot must perform to accomplish the task goal. The coordinated execution of a task, the scheduling and synchronization of sub-tasks based on logical conditions, external and internal events was realized using petri nets as described in [7].

Fig. 2: ARMAR-IIIa in the kitchen ([8, 10].

In 2006, the robot ARMAR-III was developed [8, 10], see Fig. 2. Based on the experience gained in developing the first robots, a new design was realized with improved mechatronics, sensor system and embedded control architecture with the goal of performing tasks in a kitchen environment. The upper body has been designed to be modular and light-weight while retaining similar size and proportion as an average person. For the locomotion, a mobile platform was used to allow for holonomic movability in the application area. Two years later, a second copy of the robot, ARMAR-IIIb, was developed and featured minor improvements of the embedded control architecture. Both robots have 43 DoF each. Each arm has 7 DoF and is equipped with 8 DoF five-finger hand and 6 DoF force/torque sensor in the wrist. Planar artificial skin pads [42] are mounted to the front and back of each shoulder and cylindrical artificial skin pads are mounted to the upper and lower arms. The head has 7 DoF. The eyes have a common tilt and can pan independently and

[9] www.mca2.org

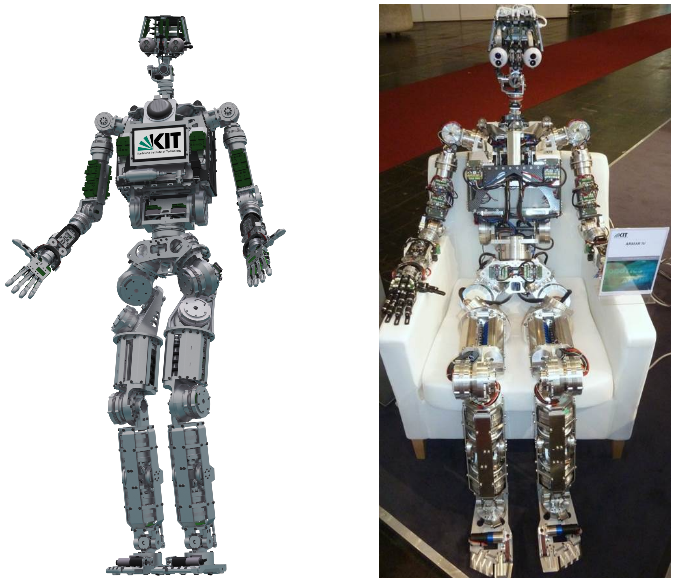

Fig. 3: The humanoid robot ARMAR-4 [9].

each eye is equipped with two digital color cameras, one with a wide-angle lens for peripheral vision and one with a narrow-angle lens for foveal vision to allow simple visuo-motor behaviors [10]. The torso and the mobile base have 3 DoF each. Similar to the previous versions, DC motors and harmonic drives were used in the arms, head, torso and mobile base. Linux with the Real-Time Application Interface RTAI/LXRT-Linux was used as operating system with MCA as software architecture. The computer architecture consists of industrial PCs and PC/104 systems connected via Gigabit Ethernet and 10 DSP/FPGA control units (UCoM) which communicate with the control PC via CAN bus. Details of the mechanical design of the robot are described in [1]. The total weight of the robot is about 140 kg with about 70 kg for two batteries. The software architecture of the robot is mainly implemented in the MCA. Since 2013, new software modules are implemented within the new software framework ArmarX[10].

In 2012, the humanoid robot ARMAR-4 was developed, a biped humanoid robot with 63 torque controlled DoF [9], see Fig. 3. The robot has 63 actuators, 214 sensors, 76 microcontrollers for low-level control, 3 PCs for perception, high-level control and balancing, a weight of 70 kg including batteries and total height of 170 cm. The head has 9 DoF and is equipped with two eyes and two cameras in each eye. The eyes execute independent pan and tilt movements. The visual system is mounted on a 5 DoF neck mechanism for lower and upper pitch, lower and upper roll and yaw. The arms have been developed to increase dexterity in bimanual manipulation tasks. Based on the human arm model developed in [3] and the manipulability

[10] `https://armarx.humanoids.kit.edu`

analysis [85] of the human arm kinematics, the inner shoulder joints (articulatio sternoclavicularis) was implemented in addition to the standard 7 DoF in humanoid arms. The design of the hand uses an improved version of the fluid actuators [29]. The hand has the size of a human hand and a weight of 450 g. Each finger of the hand has 2 DoF and the thumb can be moved in opposition, which results in 11 DoF per hand. The hand sensor system consists of absolute joint position encoders and air pressure sensors. The complete control electronics, sensors and valves have been integrated in the hand. In designing the legs, the basic requirements were to have a highly integrated design with advanced capabilities in terms of joint torques and human-like walking styles. Each leg has 7 DoF: 3 DoF in the hip, 1 in the knee, 2 in the ankle and 1 DoF at the toes. The structural parts are made of an aluminum alloy and optimized using FEM-based CAE-optimization. Joints of the arms and legs are realized by sensor-actuator units which integrate motor, harmonic drive, incremental and absolute joint position measurements, torque and temperature. The underlying software framework is ArmarX.

3 Grasping and Manipulation

One key research focus in the development of the ARMAR robots has always been the realization of versatile grasping and manipulation capabilities. While the first version ARMAR-I just had simple grippers, ARMAR-II was already equipped with an anthropomorphic five-finger hand with fluidic actuators. Since then, five-fingered hands were successfully used on all ARMAR robots. In combination with the anthropomorphic design of the arms, this design allowed performing dexterous single-arm and bimanual grasping tasks in kitchen environments. From a conceptual point of view we designed, implemented and evaluated a grasping pipeline which integrates the required perceptual and planning components.

3.1 The ARMAR Grasping Pipeline

The ARMAR grasping pipeline integrates grasp and motion planning approaches for single handed and bimanual tasks, and provides tools for intuitive programming of grasping actions on objects as well as strategies for execution of planned actions based on multi-sensory feedback, see Fig. 4. In the following, we describe several key components of the pipeline. The pipeline consists of different modules which are divided into offline and online modules. Offline components are used to pre-compute and generate information that is required during online processing. This involves the planning of feasible grasps based on prior object knowledge which is stored in an object database [57, 76]. Manipulability and reachability metrics are used to optimize the whole-body configuration and to support the online grasp selection [73, 69]. The hand-eye calibration module is used to deal with perception

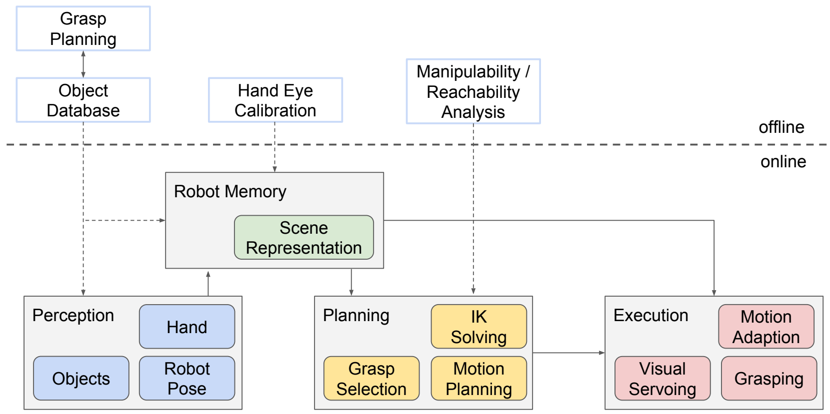

Fig. 4: The ARMAR grasping pipeline.

and action uncertainty [68]. The online components cover the robot memory, perception, planning, and execution components. The robot memory is used as a central component to exchange information between the different parts of the software architecture. This allows decoupling the perceptual processes from planning and action execution which in particular eases the exchange of existing and the integration of new components. The planning component is responsible for grasp selection [69], inverse kinematics [70, 74], and planning of collision free motions [71]. The execution of grasping and manipulation motions is realized with a position-based visual servoing approach [20], allowing to deal with inaccuracies and to adapt to changing environments in a reactive manner. In addition, visual collision detection [62] is performed to generate corrective movements during grasp execution. The online planning steps for an exemplary grasping task are depicted in Fig. 5. First, the reachability inversion approach [72] is used to determine suitable robot positions

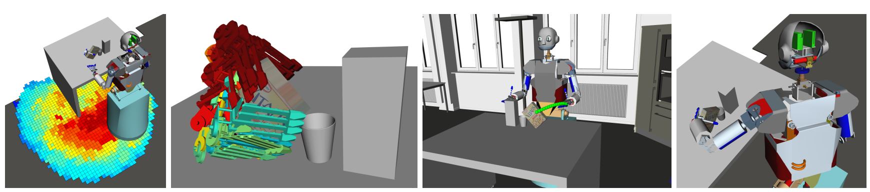

Fig. 5: Online planning of grasps based on precomputed manipulability information. From the left tp the right: 1) The 3D scene representation and a 2D map of potential robot positions for grasping. 2) Reachable grasps are colorized according to their manipulability. 3) A collision-free grasping motion. 4) The target grasping pose.

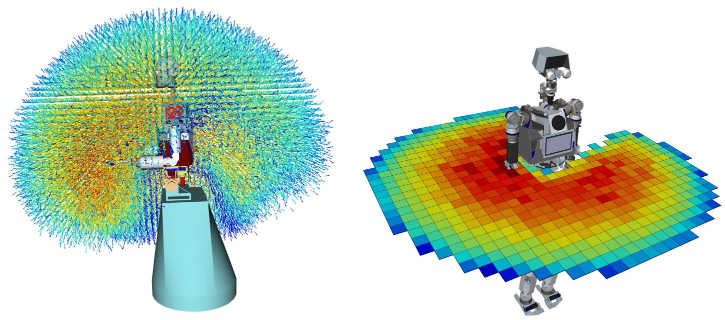

Fig. 6: Reachability and manipulability maps of ARMAR-III and ARMAR-4.

for grasping. Then, the most suitable grasp is selected based on the reachability and manipulability information for the corresponding arm. The IK-problem is solved when generating a collision-free whole-body target posture, which is used to plan a collision-free grasping motion with sampling-based motion planning approaches.

To support grasp selection and inverse kinematics tasks, we developed explicit representation of the robot's workspace as described in [69]. The robot workspace is represented by a 6D grid covering the positional and orientational parts of the Cartesian space. Each grid cell contains quality information for reachability and manipulability of a given kinematic chain of the robot. A visualization of this 6D space of the right end-effector of ARMAR-III and ARMAR-4 is shown in Fig. 6.

Object Models, Visual Object Recognition and Localization

For object recognition, localization and grasping, we use object models consisting of 3D geometric models, multi-view visual information and off-line generated grasps. Daily kitchen objects were scanned and stored in the KIT object model database [41] which is publicly accessible through a web interface[11]. The ARMAR grasping pipeline utilizes this data in multiple ways. Beside visualization purposes, the object models are used for grasp planning and collision detection. Additionally, geometric and visual information is used for object recognition and localization. We have developed two object recognition and localization systems for two classes of objects: objects that can be segmented globally, and objects exhibiting a sufficient amount of texture, allowing the application of methods using local texture features. Among the first class of objects are colored plastic dishes and cups, where we combine appearance-based methods, model-based methods and stereo vision for shape-based object recognition and localization [15]. Among the second class of objects are textured objects such as tetrapacks, boxes with any kind of food, or bottles, as

[11] `http://object-database.humanoids.kit.edu/`

Fig. 7: Object recognition of daily objects based on shape information and appearance features.

can be found in any kitchen. For object recognition and localization of such objects, we use a combination of the Harris corner detector and the SIFT descriptor [16].

Visually Guided Execution of Grasping Motions

To cope with the inaccuracies in perception, modeling and control, the execution of grasping motions is performed using a position-based visual servoing approach [77, 20]. By continuous visual observation of the current and the expected position of the end effector w.r.t. the target object, the motion can be adapted to follow the planned trajectory. An example of visually guided execution of single arm grasping motions is shown in Fig. 8. The position-based visual servoing approach has also been applied to loosely coupled bimanual tasks such as pouring juice and tightly coupled bimanual tasks, such as manipulating a wok, see Fig. 8, as well as for tasks involving physical interaction with the environment. In case of physical interaction tasks such as opening the fridge, see Fig. 9, we combine visual and force information to deal with uncertainties in perception and execution [82]. In addi-

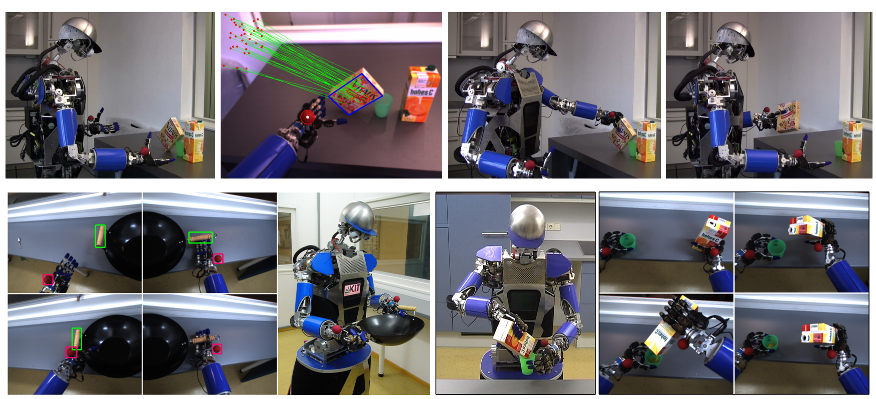

Fig. 8: Visual servoing based execution of planned single arm tasks (top) and bimanual task (bottom).

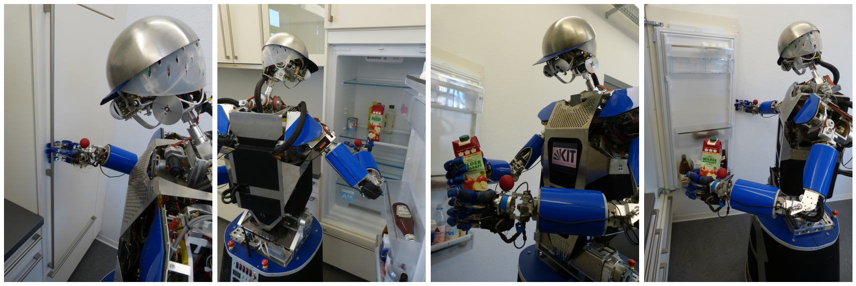

Fig. 9: Opening and closing the fridge and object hand-over from left hand to the right hand.

tion, we developed a mechanism for gaze selection during bimanual manipulation tasks, which accounts for accuracy requirements of the manipulation task by a new saliency measure which allow for computing an optimized viewing angle for visual servoing in bimanual grasping or re-grasping task [81].

Simultaneous Grasp and Motion Planning

As depicted in Fig. 4, grasping information is usually generated during an offline step and the information is stored in a database. In addition to such offline approaches, which we used e.g. in [56], we developed an integrated approach to simultaneously plan feasible grasps and collision-free grasping motions [71]. The *GraspRRT* planner combines the three main tasks needed for grasping an object: building a feasible grasp, solving the inverse kinematics problem and searching for a collision-free trajectory that moves the robot hand to the grasping pose. Therefore, RRT-based algorithms are used to build a tree of reachable and collision-free configurations. During RRT-generation, grasp hypotheses together with corresponding approach movements are computed. The quality of reachable grasping poses is evaluated online via grasp wrench space analysis. In addition to single-handed grasping motions, this approach is also capable of generating bimanual grasps together with corresponding dual arm grasping motions, see Fig. 10.

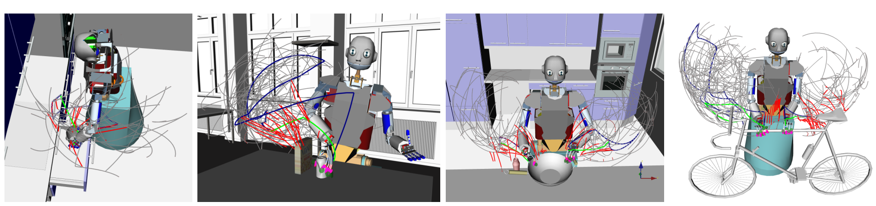

Fig. 10: The *GraspRRT* approach for single-handed and bimanual tasks.

Grasping Familiar Objects

The methods for object grasping and manipulation described above assume complete knowledge about the objects. To relax these assumptions, we started working on methods for grasping familiar objects, which exploit shape similarities of object parts and transfer grasping knowledge to similar objects [76]. To this end, we group similar objects to object categories and generate generalized grasping information with high transferability rates, indicating that such grasps can be applied to similar objects during online execution. We use a set of training models to generate grasping information that can be successfully applied to all objects in a category. In addition, objects are preprocessed in order to build a segmented mesh representation that allows performing part-based grasp planning and semantic information is connected to each segmented part. The approach allows grasping unknown objects if they belong to a trained object category and if the perception component is able to classify them correctly. The concept is divided into a grasp planning phase and an online phase in which grasping information is transferred to a novel object, see Fig. 11.

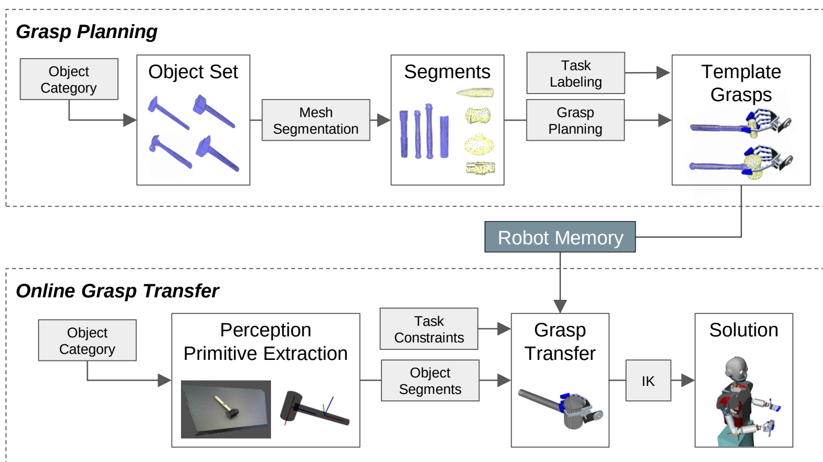

Fig. 11: Part-based grasp planning is performed on multiple objects of an object category. The resulting grasping information is evaluated according to the expected transferability to novel objects. The grasping information is applied online to novel objects which are segmented according to their RGB-D appearance.

In Fig. 12, a realistic use case with ARMAR-III is shown, where the robot is supposed to grasp a novel but familiar object. This means that the object mesh has not been used for generating the template grasps during offline grasp planning. Here, we assume that the object category (*hammer*) as well as the environment is known. Nevertheless, the actual shape of the object is not available but approximated based on visual information from geometric primitives. With this approach we are also

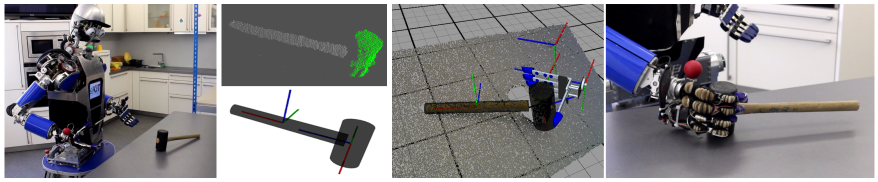

Fig. 12: A grasping task with a novel, but familiar object. The figure shows the perceived point cloud, the matched shape primitives, the transferred grasp and the xecution of the grasp on ARMAR-III.

able to identify and transfer task affordances as they are linked to the object parts. This information is used to transfer the generalized grasping information.

3.2 Grasping and Manipulation of Unknown Objects

Grasping Unknown Objects

In addition to grasping known and familiar objects we developed methods for active learning of objects by manipulation, which enables the discovery of objects in visual scenes and the accumulation of information about them. By making use of the robot's manipulation capabilities, we introduce additional information that can be utilized to reliably detect new, previously unknown objects and learn their visual appearance and finally grasp them using reactive grasping with corrective movements based on haptic feedback ([61, 60, 63]). In Fig. 13, ARMAR is in front of complex, cluttered scene with previously unknown objects. To segment objects in the scene, the robot generates object hypotheses based on the visual input following some heuristics. These initial object hypotheses may correspond to actual objects,

Fig. 13: ARMAR-III pushing an object in a cluttered scene to segment it.

but they may also cover only parts of an object or contain two or more different objects. In order to verify and correct these hypotheses, the robot interacts with the objects by pushing them to induce motion. This motion is then used as an additional cue that helps to verify or discard the object hypotheses, to remove parts of them that do not belong to the actual object and extend them to cover the object completely. For details of the approach, the reader is referred to [61].

Manipulation in Unknown Environments

To enable humanoid robots to perform grasping and manipulation tasks in unknown and/or dynamically changing environments, we investigated – inspired by the concepts of *affordances* [31] and *Object-Action Complexes (OACs)*, which was introduced by the PACO-PLUS project [12] and described in [44] – methods for the definition of objects and the environment in terms of action and interaction possibilities of the robot on them. In [18], affordance-based methods were applied in the context of haptic exploration on unknown objects. In [35, 37], we generalized the approach to *loco-manipulation*, i.e. tasks involving whole-body locomotion and manipulation of a humanoid robot, and developed a framework for the perception, extraction and validation of loco-manipulation affordances in unknown environments. Fig. 14 gives an overview over the perception and affordance extraction framework. RGB-D point clouds are used as input and are segmented into geometric primitives. Affordance hypotheses are associated with them, evaluated through affordance certainty functions and validated through physical interaction.

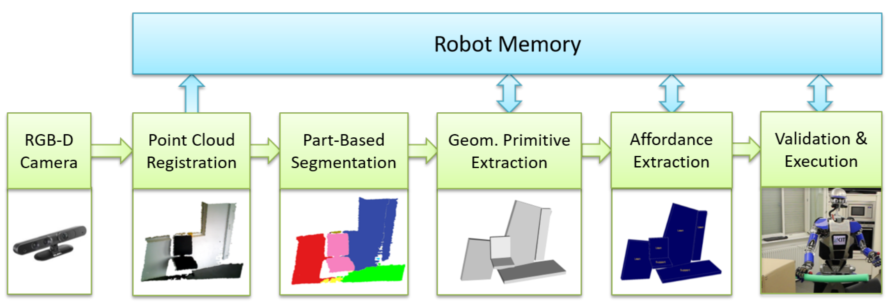

Fig. 14: The perception and affordance system.

Affordance certainty functions are hierarchically defined over the space of end-effector poses [34]. Higher-level affordances can be defined by combining lower-level affordances with additional properties of the corresponding primitive or with body-scaled properties of the perceiving agent. Fig. 15 shows visualizations of the different steps of the perception and affordance extraction system. The affordance-

[12] http://www.paco-plus.org

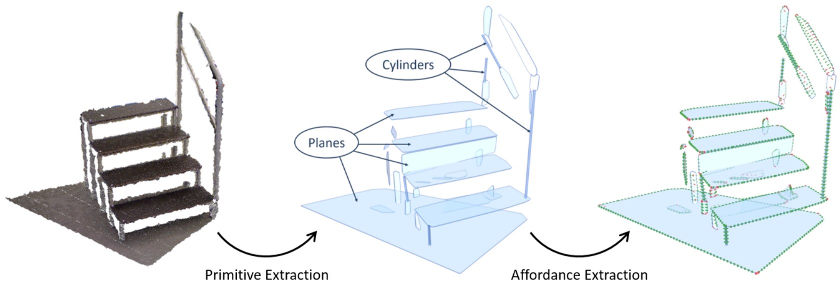

Fig. 15: Segmentation of the scene (left) into geometric primitives (center) and visualization of the affordance certainty function (right).

based system is implemented in the ArmarX and was evaluated in multiple scenarios on the humanoid robots ARMAR-III and ARMAR-4, as well as on WALK-MAN. Fig. 16 shows an evaluation scenario in which ARMAR-III detects a turnable valve. After detection, the inherently proposed end-effector poses are used for parameterizing a bimanual turning skill, which is then executed in order to turn the valve. In [36], a pilot interface was added to the system to allow high-level control of humanoid robots based on automatically proposed affordances and corresponding OACs for action execution.

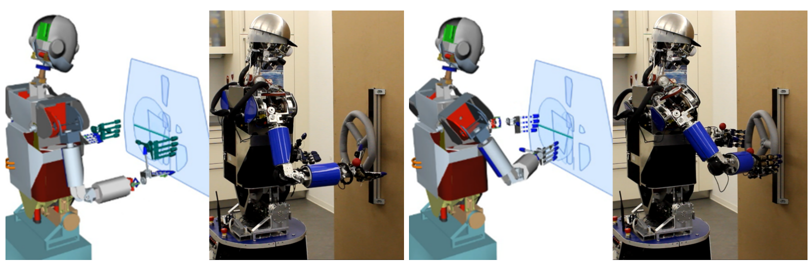

Fig. 16: Affordance-based bimanual valve turning.

4 Learning from Human Observation

To enhance the grasping and manipulation capabilities of humanoid robots in human-centred environments, we investigated how humanoid robot can acquire novel motor and task knowledge and adapt this knowledge to new situations by learning from human observation and experience. Promising way to achieve this goal are robot Programming by Demonstration (PbD) and imitation learning methods by which a robot learns new skills through human guidance [45, 22, 19]. These methods take inspiration from the way humans learn new skills by imitation to develop methods by which new skills can be transmitted to a robot. Learning task knowledge and control strategies for robots with high number of degrees of freedom which are deemed to interact in complex and variable environments, such as households is faced with two key challenges: first, the complexity of the tasks to be learned is such that pure trial and error learning would be too slow. PbD has proven thus to be a good approach to speed up learning by reducing the search space, while still allowing the robot to refine its model of the demonstration through exploration. Second, there should be a continuum between learning and control, so that control, strategies can adapt on the fly to drastic changes in the environment. In our research we address the three main research questions: observation of human demonstrations, generalization of the demonstration and reproduction of learned knowledge on the robot. In the following we describe our work in these areas.

4.1 Observation of Human Actions

The observation of human demonstrations of grasping and manipulation actions is accomplished by marker-less or marker-based human motion capture methods. In order to enable our humanoid robots to observe humans with its onboard stereo cameras, we developed various methods for the marker-less tracking of human body parts. For the tracking of the human upper body based on stereo color images, we presented in [13] a tracking algorithm which uses an adapted 3D model of the upper body model with 14 DoF (6 DoF for the base transformation, $2 \cdot 3$ for the shoulders, and $2 \cdot 1$ for the elbows), see Fig. 17. The model tracking is implemented using a particle filter algorithm which uses two different cues, an edge and a distance cue, in order to evaluate how well a given model configuration matches the current observations. Regarding the edge cue, observations are extracted from stereo images in the form of an edge map. These are compared against a set of contour points which

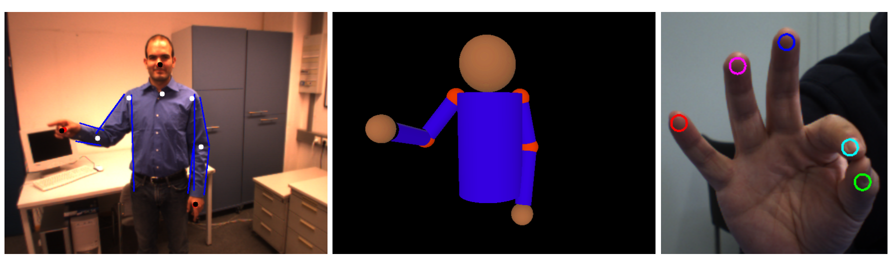

Fig. 17: Left: Upper body model projected on the image plane. Middle: Visualization of the 3D model used for upper body tracking using the method proposed in [13]. Right: Fingertip positions tracked using the method proposed in [25].

originate from the projection of the 3D model of the upper body onto the image plane. In addition to the edge cues, a distance cue is defined based on a stereo-based 3D hand and head tracking. To increase the robustness, the method implements a particle sampling in the vicinity of an inverse kinematics solution in order to facilitate the re-initialization of the procedure. With this system, we demonstrated an online imitation of human actions on the humanoid robot ARMAR-IIIb with an active vision system [14].

For the observation of human grasping actions, we presented in [25] a fingertip tracking approach that determines the fingertip positions in the foveal camera views based on circular image features. These features are tracked using a method which combines particle filtering with a mean shift algorithm. To enhance the feature extraction process, an edge map is generated by applying a multi-scale edge extraction technique. Based on this map, circular image features are extracted and formed into observations which are needed for the actual tracking algorithm. To increase the robustness of the algorithm, dynamical motion models are trained for the prediction of the finger displacements. An example of the resulting tracked fingertip positions is shown in Fig. 17.

However, marker-less motion capture methods are very limited regarding resolution, frame rate, and field of view, and are particularly sensitive to occlusions, which is unfavorable for the observation of manipulation actions. For the acquisition of accurate, high frequency motion data, we therefore employ a marker-based passive-optical Vicon MX motion capture system. The technique which is used here is described in [26, 50, 66] and relies on infrared cameras and artificial reflective markers. These markers are placed on 56 predefined anatomical landmarks of the human subject according to defined reference marker set. By triangulation, the Cartesian 3D locations of the markers can be determined. In order to gain information on the interaction of a human and the environment as well as the objects within, markers have been placed on environmental elements such as stairs, doors, and handrails as well as on objects which are subject to manipulation.

4.2 Unifying Representation of Human Motion

In order to enhance the comparability and facilitate the processing of human motion captured from distinct subjects with different approaches, a standardized interface based on the Master Motor Map (MMM) framework has been established. The MMM framework allows the unifying representation of human motion, which thus normalizes the observed motion from individual body characteristics and minimizes the efforts required to transfer motion between different embodiments. The MMM has been introduced in [12] and extended in [50, 66], and incorporates a whole-body reference model of the human body with kinematic specifications for 104 DoF: 6 DoF cover the model pose and 52 DoF are distributed on arms, legs, head, eyes and torso, which is shown in the left and middle of Fig. 18. The remaining DoFs are used to model the hand (23 DoFs each), shown right in Fig. 18. Besides

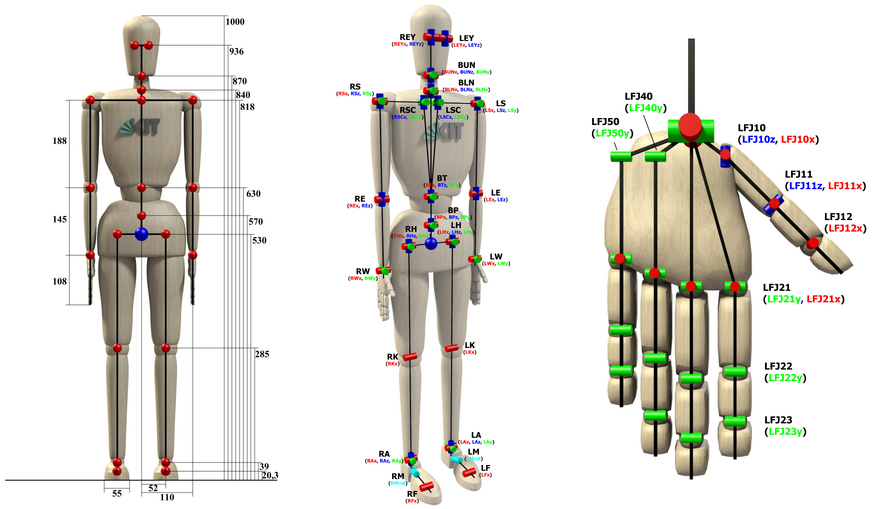

Fig. 18: The kinematics of the MMM reference model. Left side: Segment lengths (given in $\frac{1}{1000}$ of the total body height). Middle: Rotational joints, starting from the root node at the pelvis mid-point (blue sphere). Right: MMM model of the human hand with 23 DoF.

the kinematic specification of the reference model, the limb segments are enriched with proper body segment properties, such as mass distribution, segment length, and moment of inertia in order to compute whole-body dynamics. These anthropomorphic properties are defined and scaled according to linear equations which link these properties to global parameters such as height and weight of the whole body. Using the reference model of the human body, any motion capture data can be represented using the MMM, as long as an appropriate converter of input data to the MMM representation has been implemented, and MMM data can be used for various purposes, such as retargeting to a humanoid robot or action recognition. This is depicted in Fig. 19. To map a captured motion to the MMM, virtual markers whose arrangement correspond to the reference marker set are placed on the MMM model. By solving an optimization problem in order to minimize the distances between the virtual markers and the corresponding physical marker positions, joint angles for the model are reconstructed from motion capture data. For more details, the reader is referred to [50].

The captured raw motion data as well as post-processed motions in the MMM format are stored in our KIT Whole-Body Human Motion Database[13]. This database serves as a key element for a wide variety of research questions related e.g. to human motion analysis, imitation learning, action recognition, and motion generation in robotics. In contrast to previous approaches, the motion data in our database does not only considers motions of single human subjects, but is focused on the interac-

[13] `https://motion-database.humanoids.kit.edu`

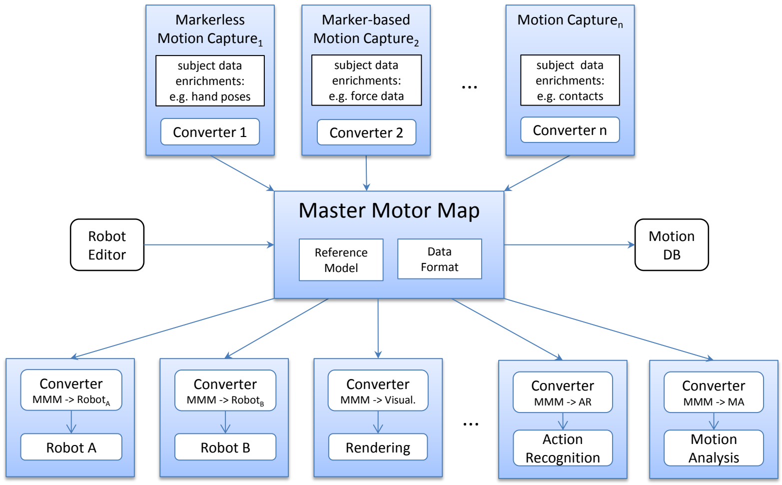

Fig. 19: The Master Motor Map (MMM) framework [50]: The human motion capture process is decoupled from motion analysis and reproduction components by providing a reference model of the human body and a unifying data format with corresponding interfaces. This allows to unify a wide variety of input data sources and to make motion capture data available in a transparent way.

tion between the human, objects, and the environment, which is crucial for the understanding of human actions, and for the reproduction of such tasks on ARMAR. Fig. 20 shows an example for such a motion task. Additionally, motion data for the interaction between several agents, both humans and robots, are available. Up to now, our motion database comprises motion data of a total run length of 28 hours captured from 146 different subjects (48% male, 17% female), and also includes motion data from the well-known CMU Graphics Lab Motion Capture Database[14]. For further details on the KIT Whole-Body Human Motion Database, the interested reader is referred to [50, 49].

4.3 Segmentation of Demonstrations

Recordings of human demonstrations are usually a seamless stream of actions without known meaning. In order to extract the meaning of demonstrations further processing and analysis of human motion is necessary. To this end, we introduced several algorithms and methods for the segmentation and labeling of motion data on the semantic level based on contact relations between human subjects, objects, and environment (see [80, 48]) up to the trajectory level by analyzing the position, ve-

[14] `http://mocap.cs.cmu.edu/`

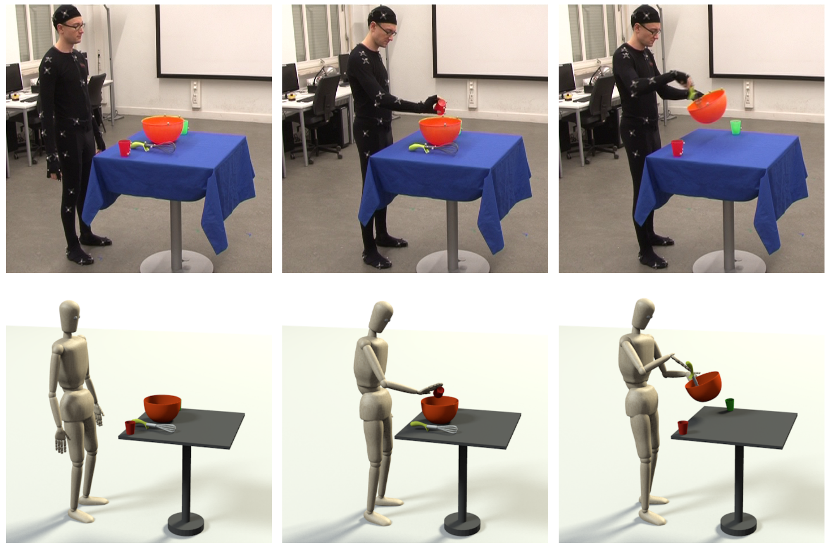

Fig. 20: Preparing the dough: Snapshots and MMM reconstruction from motion capture for a manipulation task with four environmental objects.

locity, and acceleration profiles in order to identify meaningful key points between motion segments (see [78, 4, 87]).

In [80], we employ a hierarchical two-level segmentation approach to extract manipulation actions from a human demonstration. On the first level, semantic segments are extracted based on object relation changes where contact/non-contact information is used to determine the borders of semantic segments. On the second level, the resulting semantic segments are further analyzed and subdivided into more granular segments. To achieve this, a method for assessing the characteristics of a motion within a semantic segment based on the motion acceleration profile is used. The motion is recursively divided by searching iteratively with a sliding window for the best split to find new segments. For each position of the sliding window, the trajectory left and right of the center of the window is evaluated. A score function describes the motion characteristic of the trajectory segments by incorporating the dynamics of the motion. A good segmentation point means that the difference between the score values of the left and right sub-segments is high, i.e. the difference in motion dynamics is high. On the next recursion level, the segments left and right of segmentation point are analyzed again until a minimum segment length is reached or the difference between the score function values is under a given threshold. Based on this sub-segmentation it is possible to find segments that differ in dynamics of the motion but have no observable effects on the environment and thus to allow distinguishing different actions, such as shaking a bottle or tossing a bottle. In Fig. 21 on the left, the different processing steps from demonstration to segmentation are

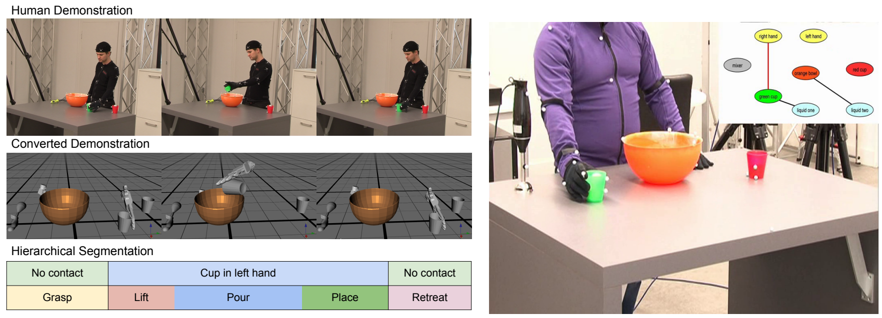

Fig. 21: Hierarchical segmentation of human demonstration based on contact relations and motion characteristics.

depicted. On the right a snapshot of a demonstration and the relation-graph at that moment is shown.

4.4 Representation of Actions

A core element which is crucial for the ability to learn goal-directed actions from observations is the representation of segmented actions, which are able to encode relevant movement characteristics extracted from the observed demonstrations and on the other hand allows generalization to different contexts. In [5], an approach using Hidden Markov Models (HMM) has been proposed in order to learn dual-arm tasks from multiple demonstrations. Based on characteristic key points which are extracted from the demonstrated trajectory an HMM is trained to encode a demonstration. To infer a generalized representation of the demonstrated task a state sequence is determined based on common key points, respectively states which are common to all trained HMMs. The represented task can be reproduced by interpolating between these states. To address the increasing kinematic complexity of humanoid robots and the required goal-directedness in robot grasping and manipulation, the concept of Dynamic Movement Primitives (DMP) was introduced in [32], which became a popular and widely used movement representation. A DMP incorporates linear dynamical systems, also referred to as transformation system, which is perturbed by nonlinear force term in order to encode arbitrary complex movements. A canonical system which drives the evolution of the system allows the integration and synchronization of multiple transformation systems and, thus, facilitates the encoding of movements even in higher dimensions. For discrete actions, the canonical system takes the form of a point attractor, which ensures that the transformation system converges to the goal within a time duration specified by a constant temporal parameter. This structure allows the parametrization of a DMP with new target positions and makes this representation temporally scalable in order

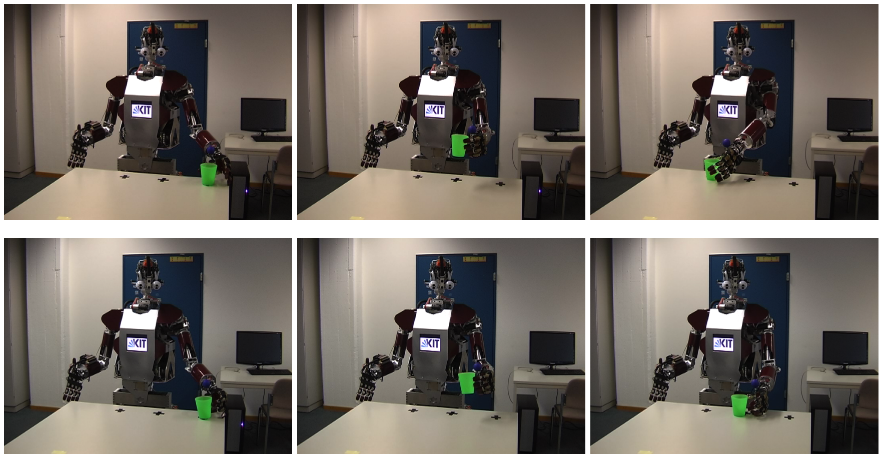

Fig. 22: ARMAR-IIIb performing pick-and-place actions using DMPs learnt from human observation.

to adapt a learned action to a new context. In [55], the feasibility and the generalization capabilities of DMPs regarding the learning, representation, and goal-directed reproduction of discrete actions such as pick-and-place (see Fig. 22), pouring, sliding (see Fig. 23), and pointing are shown.

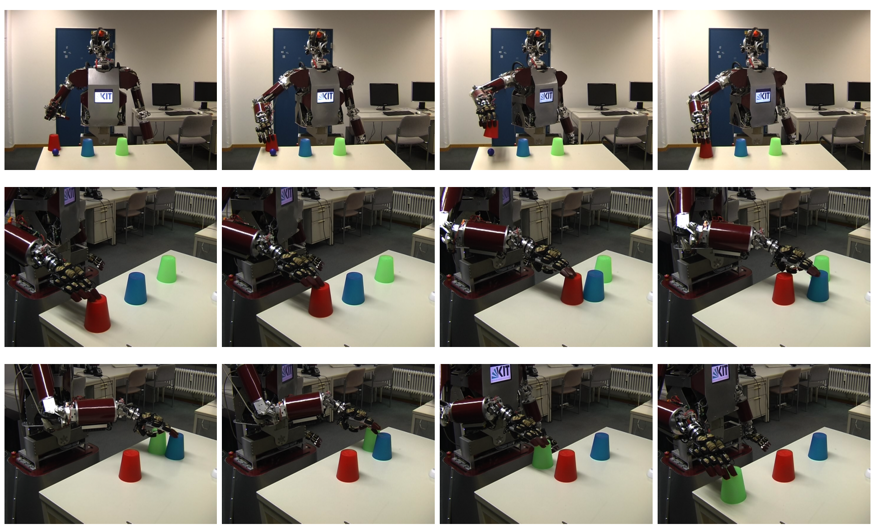

Fig. 23: ARMAR-IIIb playing shell game by executing sliding actions learned from human observation.

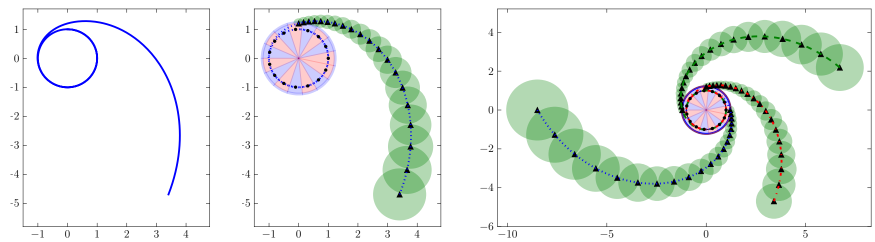

Fig. 24: Two dimensional canonical system: One dimension represents the current phase of the periodic motion while the other dimension corresponds to the distance of the current state towards the periodic pattern.

To enable a DMP to represent a periodic action where the start and end points of the corresponding trajectories are connected, the point attractor in the canonical system has to be replaced with a limit-cycle attractor. The temporal scale is replaced by a frequency parameter which can be used to control the periodic evolution of the transformation systems. Experiments showing the capability to encode periodic actions have been conducted in [30], where different table wiping actions DMPs have been learned in an online manner. In [28], an extension of the DMP formulation using a two dimensional canonical system is presented in order to achieve a common representation of discrete and periodic actions within a single DMP (see Fig. 24). This DMP formulation has been used on the ARMAR robots to encode periodic wiping actions and their corresponding discrete transient behaviours.

4.5 Learning from experience

Learning from sensorimotor experience and interaction with the world is fundamental for continuous grounding, adaptation, refinement and augmentation of actions learned from human demonstration. In this context, we introduced in the Xperience project [15] and in [84] the concept of Structural Bootstrapping – an idea taken from child language acquisition research - to define a generative mechanism that uses existing robot experience together with new observations to supplement the robot's knowledge with missing information about, object-, action- as well as planning-relevant entities. Thus, Structural Bootstrapping can be seen as a method of building generative models, leveraging existing experience to predict unexplored action effects and to focus the hypothesis space for learning novel concepts. Following this concept, an experience-based learning cycle for the wiping task was developed to learn generative models describing the relation between object properties and action parameters from sensorimotor experience and to use these models to predict the consequences of actions using internal simulation [27]. The learning cycle is shown

[15] www.xperience.org

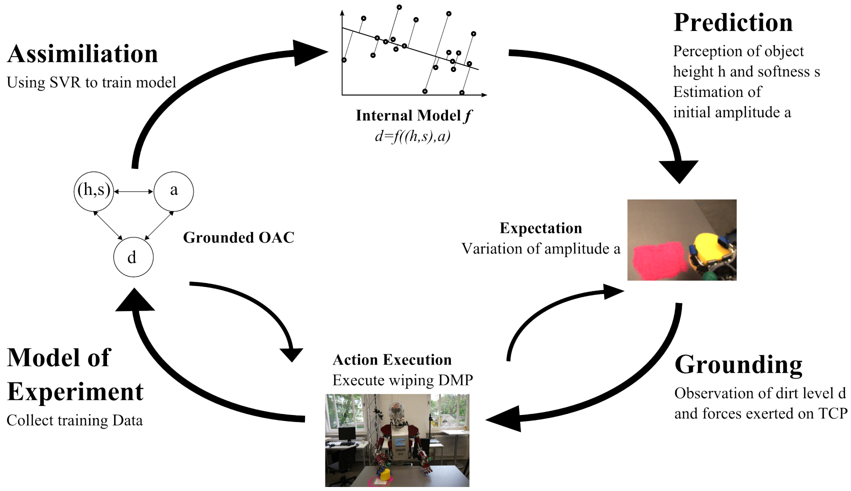

Fig. 25: The experience-based learning cycle for the wiping task.

in Fig. 25 and consists of four stages. The initial stage of this cycle is the exploration stage. Given generalized representations of objects and actions, the robot explores the scene in order to obtain instantiations of the object and the action. In the context of wiping, an object representation is inferred by manipulating and observing the object in hand, e.g. height and softness of a wiping tool are determined based on a squeezing action. A motion primitive encoding a wiping action and the object representation form the basis of an experiment which is conducted in the subsequent stage to create data from which concrete experience can be generated. The robot applies the wiping action and observes its effect on object, environment, and on itself, see Fig. 26. Experimental data is generated from these observations and used to ground and adapt the representations. In the modeling stage, knowledge in the form of internal models is extracted from experiential data. In subsequent iterations of the learning cycle, these models can be used to determine the expected effect and to constrain and control the exploration of the action parameter space within the repeated experiment based on novel perceived object representations. This allows the goal-directed creation of training data which is needed for the re-grounding of the representations and the revision of the internal models. Hence, this learning cycle allows the continuous acquisition, validation, and refinement of internal knowledge in long term association through exploration and predictive reasoning.

5 The ARMAR Architecture

A complex system like a humanoid robot imposes complex and versatile requirements on its control and cognitive architecture. With the requirements of a humanoid

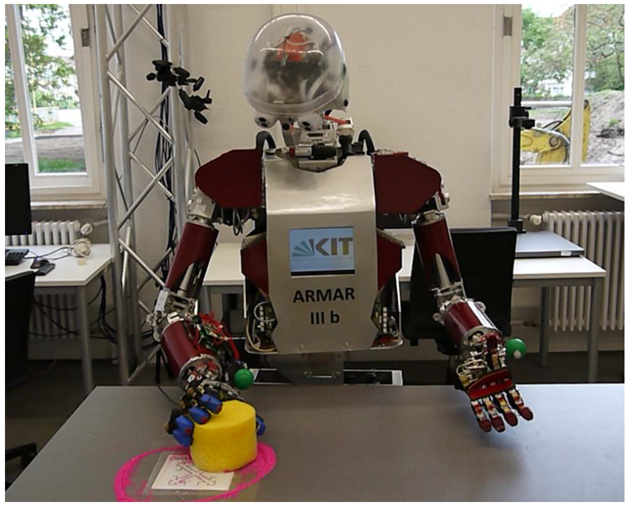

Fig. 26: ARMAR-IIIb wiping the table with a sponge.

robot in mind, the architecture shown in Fig. 27 has been implemented in the robot software development environment *ArmarX*, see also [75] and[79]. The architecture consists of three levels: the sensorimotor low-level, a memory system as the mid-level, and the reasoning and planning high-level. The low-level incorporates a hardware abstraction component and the ArmarX statecharts ([79]) for the execution of sensorimotor skills. The statecharts combine sensorimotor data of the low-level and processed data of the memory layer in a task-specific way to implement higher level skills such as grasp, lift, open, wipe, etc. The mid-level consists mainly of the memory structure MemoryX, which is a mediator between the low- and high-level and is equipped with processing capabilities like object recognition, object detection, hand tracking, self-localization, etc. MemoryX itself consists of three different types of memory: The prior knowledge, the long-term memory and the working memory. The prior knowledge contains information that is given by the developer like object classes with 3D models and e.g. pre-computed grasps or other offline generated knowledge. In the long-term memory, the robot stores information about common object places, object-action relations and associations, i.e. Object-Action Complexes (OACs), which should remain persistent but might change over time due to new experience. The third memory type is the working memory. In this volatile memory, the robot stores all the information it knows about the current world state like currently present objects and their position. The information stored in the MemoryX is the basis for the execution of complex actions that require processed sensor information and for the reasoning on the high-level of the architecture. At the high-level, the symbolic reasoning is realized to enable natural communication with the human for solving complex tasks. The three main components of the high-level are the Language Understanding (LU) component, the Replacement component (RM) and the Plan Execution Monitor component (PEM). The LU component transforms

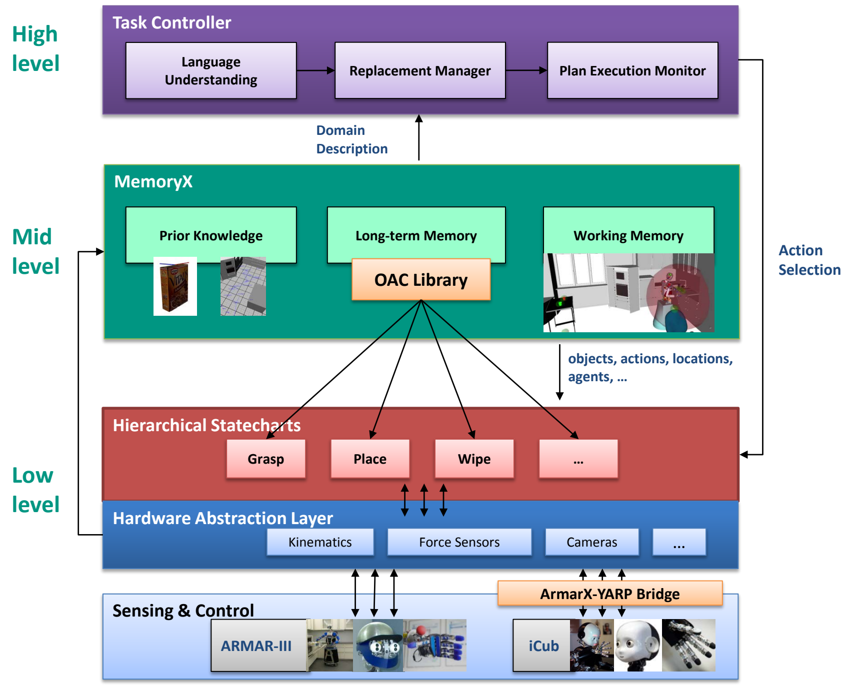

Fig. 27: The ARMAR robot architecture.

freely spoken text based on knowledge bases and the current memory content into a machine understandable representation and produces for example goals for the planning engine of the PEM (see [54]). The RM is a preliminary feasibility check for a given task as it checks e.g. for availability of all objects and their locations required for the task execution. In case of missing objects, it suggests objects and location replacements based on several replacement strategies. Upon success of the RM, the PEM is employed to find a symbolic plan for the current task. Once a plan is found, it executes and monitors the symbolic actions based on the OAC library of the long-term memory. The OAC library serves as a mediator between the PEM on the high-level and statecharts of the low-level. The PEM receives the available actions for solving a given task from the OAC library. To this end, the OAC library provides the action parameters as well as the required preconditions for execution and the expected effects. These effects are used by the PEM for the planning as well as for verifying the results of an action after execution. Each OAC in the library is also equipped with a link to the executing statechart, which is triggered by the PEM. The statechart in turn can read and store execution statistics in the OAC library to improve the performance over time. After fulfilling the task or encountering an unsolvable problem, the robot returns to an idle state and awaits new commands.

6 Conclusions and Future Plans

We introduced the ARMAR humanoid robot family, which has been developed in Karlsruhe, Germany. The robots have been developed with a special focus on the realization of daily tasks in human-centered environments. Since 2008, the ARMAR-III robot is able to perform complex tasks in a kitchen environment including vision-based grasping and manipulation of daily objects, opening/closing doors, loading the dishwasher, learning from human observation and interacting with humans using natural speech natural interaction and dialog management[16]. We described the mechatronics of the different robot generations and gave an overview over our research in the area of humanoid grasping and manipulation, learning from human observation and experience as well as the implementation of complete cognitive software architecture.

The development of the ARMAR humanoid robot family will continue in the directions of engineering high performance 24/7 humanoids able to predict, act in open-ended environments and learn from humans and own experience. Currently, we are working on wearable humanoids, a whole body exoskeleton for augmentation of human performance. A first prototype of the lower limb exoskeleton, the KIT-EXO-1 as part of ARMAR-5 is presented in [17] and shown in Fig. 28. With ARMAR-6, we are developing a humanoid robot which should provide assistance to maintenance technician in warehouse environments in a pro-active manner. The robot will be a second pair of hands that can assist the technician when he/she is in

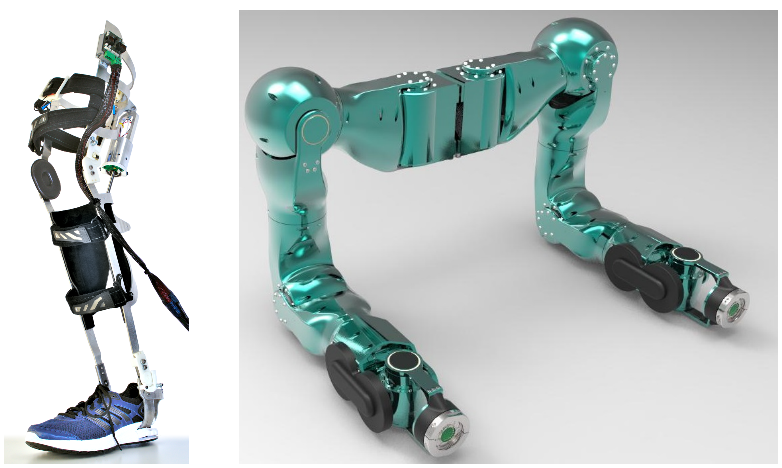

Fig. 28: Left: The KIT-EXO-1 as part of the wearable humanoid robot ARMAR-5. Right: The dual arm system of ARMAR-6.

[16] The demo in this video was given on February 3, 2008 for the evaluation panel of the humanoid robotic project (SFB 588) in Karlsruhe and more than 2200 times since this time. See `https://www.youtube.com/watch?v=SHMSyYLRQPM`

need of help. The dual-arm system of teh robot is described in [58] and shown in Fig. 28.

Acknowledgements The research leading to the development of the ARMAR humanoid robots has received funding from the the German Research Foundation (DFG: Deutsche Forschungsgemeinschaft) within the German Humanoid Research project SFB 588 and the European Union within the the EU Cognitive Systems and Robotic projects PACO-PLUS, Xperience, GRASP, WALK-MAN and KoroiBot as well as from the Karlsruhe Institute of Technology. We would like to thank all members of the H^2T lab and all students who contributed to this research in its different phases.

References

1. Albers, A., Ottnad, J., Sander, C.: Development of a new wrist for the next generation of the humanoid robot ARMAR. In: IEEE/RAS International Conference on Humanoid Robots (Humanoids), pp. 46–53 (2008)
2. Ambrose, R.O., Aldridge, H., Askew, R.S., Burridge, R.R., Bluethmann, W., Diftler, M., Lovchik, C., Magruder, D., Rehnmark, F.: Robonaut: NASA's space humanoid. Intelligent Systems and their Applications, IEEE **15**(4), 57–63 (2000). DOI 10.1109/5254.867913
3. Asfour, T.: Sensormotorische Koordination für die Handlungsausführung eines humanoiden Roboters (in German). Ph.D. thesis, Universität Karlsruhe (TH), Karlsruhe, Germany (2003)
4. Asfour, T., Azad, P., Gyarfas, F., Dillmann, R.: Imitation learning of dual-arm manipulation tasks in humanoid robots. International Journal of Humanoid Robotics **5**(2), 183–202 (2008)
5. Asfour, T., Azad, P., Gyarfas, F., Dillmann, R.: Imitation Learning of Dual-Arm Manipulation Tasks in Humanoid Robots. International Journal of Humanoid Robotics **5**(2), 183–202 (2008)
6. Asfour, T., Berns, K., Dillmann, R.: The humanoid robot armar: Design and control. In: The 1st IEEE-RAS International Conference on Humanoid Robots, pp. 7–8 (2000)
7. Asfour, T., Ly, D.N., Regenstein, K., Dillmann, R.: Coordinated task execution for humanoid robots. In: The International Symposium on Experimental Robots, pp. 259–267 (2004)
8. Asfour, T., Regenstein, K., Azad, P., Schröder, J., Vahrenkamp, N., Dillmann, R.: ARMAR-III: An Integrated Humanoid Platform for Sensory-Motor Control. In: IEEE/RAS International Conference on Humanoid Robots (Humanoids), pp. 169–175. Genova, Italy (2006)
9. Asfour, T., Schill, J., Peters, H., Klas, C., Bücker, J., Sander, C., Schulz, S., Kargov, A., Werner, T., Bartenbach, V.: Armar-4: A 63 dof torque controlled humanoid robot. In: IEEE/RAS International Conference on Humanoid Robots (Humanoids), pp. 390–396 (2013)
10. Asfour, T., Welke, K., Azad, P., Ude, A., Dillmann, R.: The Karlsruhe Humanoid Head. In: IEEE/RAS International Conference on Humanoid Robots (Humanoids), pp. 447–453. Daejeon, Korea (2008)
11. Atkeson, C., Hale, J., Pollick, F., Riley, M., Kotosaka, S., Schaul, S., Shibata, T., Tevatia, G., Ude, A., Vijayakumar, S., Kawato, E., Kawato, M.: Using humanoid robots to study human behavior. Intelligent Systems and their Applications, IEEE **15**(4), 46–56 (2000). DOI 10.1109/5254.867912
12. Azad, P., Asfour, T., Dillmann, R.: Toward an unified representation for imitation of human motion on humanoids. In: IEEE International Conference on Robotics and Automation (ICRA), pp. 2558–2563 (2007)
13. Azad, P., Asfour, T., Dillmann, R.: Robust Real-time Stereo-based Markerless Human Motion Capture. In: IEEE/RAS International Conference on Humanoid Robots (Humanoids), pp. 700–707. Daejeon, Korea (2008)
14. Azad, P., Asfour, T., Dillmann, R.: Robust real-time stereo-based markerless human motion capture. In: IEEE/RAS International Conference on Humanoid Robots (Humanoids), pp. 700–707 (2008)

15. Azad, P., Asfour, T., Dillmann, R.: Accurate shape-based 6-dof pose estimation of single-colored objects. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 2690–2695 (2009)
16. Azad, P., Asfour, T., Dillmann, R.: Combining harris interest points and the sift descriptor for fast scale-invariant object recognition. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 4275–4280 (2009)
17. Beil, J., Perner, G., Asfour, T.: Design and Control of a Lower Limb Exoskeleton (KIT-EXO-1). In: International Conference on Rehabilitation Robotics (ICORR), pp. 119–124. Singapore (2015)
18. Bierbaum, A., Rambow, M., Asfour, T., Dillmann, R.: Grasp affordances from multi-fingered tactile exploration using dynamic potential fields. In: IEEE/RAS International Conference on Humanoid Robots (Humanoids), pp. 168–174 (2009)
19. Billard, A., Calinon, S., Dillmann, R., Schaal, S.: Survey: Robot Programming by Demonstration (2008)
20. Böge, C., Vahrenkamp, N., Asfour, T., Dillmann, R.: Visual servoing for single and dual arm manipulation tasks in humanoid robots. at - Automatisierungstechnik **60**(5), 309–317 (2012)
21. Cheng, G., Hyon, S., Morimoto, J., Ude, A., Colvin, G., Scroggin, W., Jacobsen, S.: CB: A Humanoid Research Platform for Exploring NeuroScience. In: IEEE/RAS International Conference on Humanoid Robots (Humanoids), pp. 182–187 (2006). DOI 10.1109/ICHR.2006.321382
22. Dillmann, R.: Teaching and learning of robot tasks via observation of human performance. Robotics and Autonomous Systems **47**(2-3), 109–116 (2004)
23. Dillmann, R., Asfour, T.: Collaborative research center on humanoid robots (sfb 588). KI - Zeitschrift Künstliche Intelligenz **4**, 26–28 (2008)
24. Dillmann, R., Becher, R., Steinhaus, P.: Armar II - a learning and cooperative multimodal humanoid robot system. I. J. Humanoid Robotics **1**(1), 143–155 (2004)
25. Do, M., Asfour, T., Dillmann, R.: Particle filter-based fingertip tracking with circular hough transform features. In: IAPR Machine Vision Applications (2011)
26. Do, M., Azad, P., Asfour, T., Dillmann, R.: Imitation of human motion on a humanoid robot using nonlinear optimization. In: IEEE/RAS International Conference on Humanoid Robots (Humanoids), pp. 545–552 (2008)
27. Do, M., Schill, J., Ernesti, J., Asfour, T.: Learn to Wipe: A Case Study of Structural Bootstrapping from Sensorimotor Experience. In: IEEE International Conference on Robotics and Automation (ICRA) (2014)
28. Ernesti, J., Righetti, L., Do, M., Asfour, T., Schaal, S.: Encoding of Periodic and their Transient Motions by a Single Dynamic Movement Primitive. In: IEEE/RAS International Conference on Humanoid Robots (Humanoids), pp. 57–64. Osaka, Japan (2012)
29. Gaiser, I., Schulz, S., Breitwieser, H., Bretthauer, G.: Enhanced Flexible Fluidic actuators for biologically inspired lightweight robots with inherent compliance. In: IEEE International Conference on Robotics and Biomimetics (ROBIO), pp. 1423–1428 (2010)
30. Gams, A., Do, M., Ude, A., Asfour, T., Dillmann, R.: On-Line Periodic Movement and Force-Profile Learning for Adaptation to New Surfaces. In: IEEE/RAS International Conference on Humanoid Robots (Humanoids). Nashville, USA (2010)
31. Gibson, J.J.: The Ecological Approach to Visual Perception (1978)
32. Ijspeert, A.J., Nakanishi, J., Schaal, S.: Learning Attractor Landscapes for Learning Motor Primitives. In: Conference on Neural Information Processing Systems, pp. 1523–1530 (2002)
33. Iwata, H., Sugano, S.: Design of human symbiotic robot TWENDY-ONE. In: IEEE International Conference on Robotics and Automation (ICRA), pp. 580–586 (2009). DOI 10.1109/ROBOT.2009.5152702
34. Kaiser, P., Aksoy, E.E., Grotz, M., Asfour, T.: Towards a hierarchy of whole-body loco-manipulation affordances. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 2839–2846 (2016)
35. Kaiser, P., Gonzalez-Aguirre, D., Schültje, F., Borràs, J., Vahrenkamp, N., Asfour, T.: Extracting Whole-Body Affordances from Multimodal Exploration. In: IEEE-RAS International Conference on Humanoid Robots (Humanoids), pp. 1036–1043 (2014)

36. Kaiser, P., Kanoulas, D., Grotz, M., Muratore, L., Rocchi, A., Hoffman, E.M., Tsagarakis, N.G., Asfour, T.: An affordance-based pilot interface for high-level control of humanoid robots in supervised autonomy. In: IEEE/RAS International Conference on Humanoid Robots (Humanoids) (2016)
37. Kaiser, P., Vahrenkamp, N., Schültje, F., Borràs, J., Asfour, T.: Extraction of whole-body affordances for loco-manipulation tasks. International Journal of Humanoid Robotics (IJHR) (2015)
38. Kaneko, K., Kanehiro, F., Kajita, S., Hirukawa, H., Kawasaki, T., Hirata, M., Akachi, K., Isozumi, T.: Humanoid robot HRP-2. In: IEEE International Conference on Robotics and Automation (ICRA), vol. 2, pp. 1083–1090 Vol.2 (2004). DOI 10.1109/ROBOT.2004.1307969
39. Kaneko, K., Kanehiro, F., Morisawa, M., Akachi, K., Miyamori, G., Hayashi, A., Kanehira, N.: Humanoid robot HRP-4 - Humanoid robotics platform with lightweight and slim body. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 4400–4407 (2011). DOI 10.1109/IROS.2011.6094465
40. Kaneko, K., Kanehiro, F., Morisawa, M., Miura, K., Nakaoka, S., Kajita, S.: Cybernetic human HRP-4C. In: IEEE/RAS International Conference on Humanoid Robots (Humanoids), pp. 7–14 (2009). DOI 10.1109/ICHR.2009.5379537
41. Kasper, A., Xue, Z., Dillmann, R.: The KIT object models database: An object model database for object recognition, localization and manipulation in service robotics. The International Journal of Robotics Research **31**(8), 927–934 (2012)
42. Kerpa, O., Weiss, K., Worn, H.: Development of a flexible tactile sensor system for a humanoid robot. In: Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003) (Cat. No.03CH37453), vol. 1, pp. 1–6 vol.1 (2003)
43. Khatib, O., Yeh, X., Brantner, G., Soe, B., Kim, B., Ganguly, S., Stuart, H., Wang, S., Cutkosky, M., Edsinger, A., Mullins, P., Barham, M., Voolstra, C.R., Salama, K.N., L'Hour, M., Creuze, V.: Ocean one: A robotic avatar for oceanic discovery. IEEE Robotics Automation Magazine **23**(4), 20–29 (2016). DOI 10.1109/MRA.2016.2613281
44. Krüger, N., Geib, C., Piater, J., Petrick, R., Steedman, M., Wörgötter, F., Ude, A., Asfour, T., Kraft, D., Omrcen, D., Agostini, A., Dillmann, R.: Object-Action Complexes: grounded abstractions of sensorimotor processes. Robotics and Autonomous Systems **59**, 740–757 (2011)
45. Kuniyoshi, Y., Inaba, M., Inoue, H.: Learning by watching: Extracting reusable task knowledge from visual observation of human performance. IEEE Transactions on Robotics and Automation **10**, 799–822 (1994)
46. Lee, J., Kim, J.Y., Park, I.W., Cho, B.K., Kim, M.S., Kim, I., Oh, J.H.: Development of a Humanoid Robot Platform HUBO FX-1. In: SICE-ICASE, 2006. International Joint Conference, pp. 1190–1194 (2006). DOI 10.1109/SICE.2006.315364
47. Lohmeier, S., Buschmann, T., Ulbrich, H.: Humanoid robot LOLA. In: IEEE International Conference on Robotics and Automation (ICRA), pp. 775–780 (2009). DOI 10.1109/ROBOT.2009.5152578
48. Mandery, C., Borràs, J., Jöchner, M., Asfour, T.: Analyzing whole-body pose transitions in multi-contact motions. In: IEEE/RAS International Conference on Humanoid Robots (Humanoids), pp. 1020–1027 (2015)
49. Mandery, C., Terlemez, O., Do, M., Vahrenkamp, N., Asfour, T.: The KIT whole-body human motion database. In: International Conference on Advanced Robotics (ICAR), pp. 329–336 (2015)
50. Mandery, C., Terlemez, O., Do, M., Vahrenkamp, N., Asfour, T.: Unifying representations and large-scale whole-body motion databases for studying human motion. IEEE Transactions on Robotics **32**(4), 796–809 (2016)
51. Metta, G., Sandini, G., Vernon, D., Natale, L., Nori, F.: The iCub humanoid robot: an open platform for research in embodied cognition. In: Proceedings of the 8th workshop on performance metrics for intelligent systems, pp. 50–56. ACM (2008)
52. Ogura, Y., Aikawa, H., Shimomura, K., Morishima, A., ok Lim, H., Takanishi, A.: Development of a new humanoid robot WABIAN-2. In: IEEE International Conference on Robotics and Automation (ICRA), pp. 76–81 (2006). DOI 10.1109/ROBOT.2006.1641164

53. Ott, C., Eiberger, O., Englsberger, J., Roa, M.A., Albu-Schäffer, A.: Hardware and Control Concept for an Experimental Bipedal Robot with Joint Torque Sensors. Journal of the Robotics Society of Japan **30**(4), 378–382 (2012)
54. Ovchinnikova, E., Wächter, M., Wittenbeck, V., Asfour, T.: Multi-purpose natural language understanding linked to sensorimotor experience in humanoid robots. In: IEEE/RAS International Conference on Humanoid Robots (Humanoids), pp. 365–372 (2015)
55. Pastor, P., Hoffmann, H., Asfour, T., Schaal, S.: Learning and Generalization of Motor Skills by Learning from Demonstration. In: IEEE International Conference on Robotics and Automation (ICRA). Kobe, Japan (2009)
56. Przybylski, M., Asfour, T., Dillmann, R.: Planning grasps for robotic hands using a novel object representation based on the medial axis transform. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1781–1788 (2011)
57. Przybylski, M., Wächter, M., Asfour, T., Dillmann, R.: A skeleton-based approach to grasp known objects with a humanoid robot. In: IEEE/RAS International Conference on Humanoid Robots (Humanoids), pp. 376–383 (2012)
58. Rader, S., Kaul, L., Fischbach, H., Vahrenkamp, N., Asfour, T.: Design of a High-Performance Humanoid Dual Arm System with Inner Shoulder Joints. In: IEEE/RAS International Conference on Humanoid Robots (Humanoids), pp. 523–529. Cancn, Mexico (2016)
59. Sakagami, Y., Watanabe, R., Aoyama, C., Matsunaga, S., Higaki, N., Fujimura, K.: The intelligent ASIMO: system overview and integration. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), vol. 3, pp. 2478–2483 vol.3 (2002). DOI 10.1109/IRDS.2002.1041641
60. Schiebener, D., Morimoto, J., Asfour, T., Ude, A.: Integrating visual perception and manipulation for autonomous learning of object representations. Adaptive Behavior **21**(5), 328–345 (2013)
61. Schiebener, D., Ude, A., Asfour, T.: Physical interaction for segmentation of unknown textured and non-textured rigid objects. In: IEEE International Conference on Robotics and Automation (ICRA). Hong Kong (2014)
62. Schiebener, D., Vahrenkamp, N., Asfour, T.: Visual collision detection for corrective movements during grasping on a humanoid robot. In: IEEE/RAS International Conference on Humanoid Robots (Humanoids), pp. 0–0 (2014)
63. Schill, J., Laaksonen, J., Przybylski, M., Kyrki, V., Asfour, T., Dillmann, R.: Learning continuous grasp stability for a humanoid robot hand based on tactile sensing. In: IEEE RAS EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob), pp. 1901–1906. Rome, Italy (2012)
64. Schulz, S., Pylatiuk, C., Bretthauer, G.: A new ultralight anthropomorphic hand. In: Proceedings 2001 ICRA. IEEE International Conference on Robotics and Automation (Cat. No.01CH37164), vol. 3, pp. 2437–2441 vol.3 (2001)
65. Schulz, S., Pylatiuk, C., Kargov, A., Oberle, R., Bretthauer, G.: Progress in the development of anthropomorphic fluidic hands for a humanoid robot. In: 4th IEEE/RAS International Conference on Humanoid Robots, 2004., vol. 2, pp. 566–575 (2004)
66. Terlemez, O., Ulbrich, S., Mandery, C., Do, M., Vahrenkamp, N., Asfour, T.: Master motor map (mmm) - framework and toolkit for capturing, representing, and reproducing human motion on humanoid robots. In: IEEE/RAS International Conference on Humanoid Robots (Humanoids), pp. 894–901 (2014)
67. Tsagarakis, N.G., Caldwell, D.G., Negrello, F., Choi, W., Baccelliere, L., Loc, V., Noorden, J., Muratore, L., Margan, A., Cardellino, A., et al.: Walk-man: A high-performance humanoid platform for realistic environments. Journal of Field Robotics (2016)
68. Ulbrich, S., Ruiz, V., Asfour, T., Torras, C., Dillmann, R.: Rapid learning of humanoid body schemas with kinematic bezier maps. In: IEEE/RAS International Conference on Humanoid Robots (Humanoids), pp. 0–0 (2009)
69. Vahrenkamp, N., Asfour, T.: Representing the robots workspace through constrained manipulability analysis. Autonomous Robots **38**(1), 17–30 (2015)
70. Vahrenkamp, N., Asfour, T., Dillmann, R.: Efficient inverse kinematics computation based on reachability analysis. International Journal of Humanoid Robotics **9**(4), 0–0 (2012)

71. Vahrenkamp, N., Asfour, T., Dillmann, R.: Simultaneous grasp and motion planning. IEEE Robotics and Automation Magazine **19**(2), 43–57 (2012)
72. Vahrenkamp, N., Asfour, T., Dillmann, R.: Robot placement based on reachability inversion. In: IEEE International Conference on Robotics and Automation (ICRA), pp. 0–0 (2013)
73. Vahrenkamp, N., Asfour, T., Metta, G., Sandini, G., Dillmann, R.: Manipulability analysis. In: Humanoids, pp. 568–573 (2012)
74. Vahrenkamp, N., Muth, D., Kaiser, P., Asfour, T.: IK-Map: an enhanced workspace representation to support inverse kinematics solvers. In: IEEE/RAS International Conference on Humanoid Robots (Humanoids), pp. 785–790 (2015)
75. Vahrenkamp, N., Wächter, M., Kröhnert, M., Welke, K., Asfour, T.: The robot software framework armarx. Information Technology **57**(2), 99–111 (2015)
76. Vahrenkamp, N., Westkamp, L., Yamanobe, N., Aksoy, E.E., Asfour, T.: Part-based grasp planning for familiar objects. In: IEEE/RAS International Conference on Humanoid Robots (Humanoids), pp. 919–925 (2016)
77. Vahrenkamp, N., Wieland, S., Azad, P., Gonzalez-Aguirre, D., Asfour, T., Dillmann, R.: Visual servoing for humanoid grasping and manipulation tasks. In: IEEE/RAS International Conference on Humanoid Robots (Humanoids), pp. 406–412 (2008)
78. Wächter, M., Asfour, T.: Hierarchical segmentation of manipulation actions based on object relations and motion characteristics. In: International Conference on Advanced Robotics (ICAR), pp. 549–556 (2015)
79. Wächter, M., Ottenhaus, S., Kröhnert, M., Vahrenkamp, N., Asfour, T.: The armarx statechart concept: Graphical programming of robot behaviour. Frontiers in Robotics and AI **3**(33) (2016)
80. Wächter, M., Schulz, S., Asfour, T., Aksoy, E., Wörgötter, F., Dillmann, R.: Action sequence reproduction based on automatic segmentation and Object-Action Complexes. In: IEEE/RAS International Conference on Humanoid Robots (Humanoids), pp. 189–195 (2013)
81. Welke, K., Schiebener, D., Asfour, T., Dillmann, R.: Gaze selection during manipulation tasks. In: IEEE International Conference on Robotics and Automation (ICRA). Karlsruhe, Germany (2013)
82. Wieland, S., Gonzalez-Aguirre, D., Vahrenkamp, N., Asfour, T., Dillmann, R.: Combining force and visual feedback for physical interaction tasks in humanoid robots. In: Humanoid Robots, 2009. Humanoids 2009. 9th IEEE-RAS International Conference on, pp. 439–446 (2009)
83. Wimböck, T., Borst, C., Albu-Schäffer, A., Ott, C., Schmidt, C., Fuchs, M., Friedl, W., Eiberger, O., Baumann, A., Beyer, A., Hirzinger, G.: DLR's Two-Handed Humanoid Justin: System Design, Integration, and Control. at-Automatisierungstechnik **58**(11), 622–628 (2010)
84. Wörgötter, F., Geib, C., Tamosiunaite, M., Aksoy, E.E., Piater, J., Xiong, H., Ude, A., Nemec, B., Kraft, D., Krüger, N., Wächter, M., Asfour, T.: Structural bootstrapping - a novel concept for the fast acquisition of action-knowledge. IEEE Transactions on Autonomous Mental Development **7**(2), 140–154 (2015)
85. Yoshikawa, T.: Manipulability of Robotic Mechanisms. The International Journal of Robotics Research **4**(2), 3–9 (1985)
86. Zecca, M., Endo, N., Momoki, S., Itoh, K., Takanishi, A.: Design of the humanoid robot KOBIAN: Preliminary analysis of facial and whole body emotion expression capabilities. In: IEEE/RAS International Conference on Humanoid Robots (Humanoids), pp. 487–492 (2008). DOI 10.1109/ICHR.2008.4755969
87. Zöllner, R., Asfour, T., Dillmann, R.: Programming by demonstration: Dual-arm manipulation tasks for humanoid robots. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (2004)