ARMAR-6: A High-Performance Humanoid for Human-Robot Collaboration in Real World Scenarios

Tamim Asfour, Mirko Wächter, Lukas Kaul, Samuel Rader, Pascal Weiner,
Simon Ottenhaus, Raphael Grimm, You Zhou, Markus Grotz, Fabian Paus

***Abstract*—A major goal in humanoid robotics is to enable safe and reliable human-robot collaboration in real-world scenarios. In this paper, we present ARMAR-6, a new high-performance humanoid robot for various tasks, including but not limited to grasping, mobile manipulation, integrated perception, bi-manual collaboration, compliant motion execution and natural language understanding. We describe how the requirements arising from these tasks influenced our major design decisions, resulting in vertical integration during the joint hardware and software development. In particular, the entire hardware including structure, the sensor-actuator-units, the low level controllers as well as the perception, the grasping and manipulation skills, the task coordination and the entire software architecture are all developed by one single team. Component interaction is facilitated by our software framework ArmarX that furthermore allows to seamlessly integrate and interchange third-party contributions. To showcase the robot's capabilities, we present its performance in a challenging industrial maintenance scenario that requires human-robot collaboration, where the robot autonomously recognizes the human's need of help and provides such help in a proactive way.**

***Index Terms*—Humanoid Robotics, Human-Robot-Collaboration, Software-Hardware Humanoid Robot Design.**

I. INTRODUCTION AND RELATED WORK

Humanoid robots have existed for decades in research labs all around the world. While there is a wealth of possible applications in practical use-cases, the inherent complexity of humanoid robots and the complexity of the tasks they are expected to perform has hindered their transition from research labs to real-world use. One of these task complexities is the need to work closely with humans in a safe, predictable, intuitive and productive manner. An important step towards the transition of humanoids into real-world applications is, therefore, the proliferation of technologies that allow robots to work cooperatively and safely with humans, which has led to the development of collaborative robots.

On the design-side, these technologies include light-weight robot design [1], anthropomorphic kinematics (e. g. [2], [3]), force and torque sensing [4], high-speed real-time bus systems (e. g. EtherCAT) and an extremely high level of system integration to fit the required functionalities into the desired form-factor.

The authors are with the High Performance Humanoid Technologies Lab, Institute for Anthropomatics and Robotics, Karlsruhe Institute of Technology (KIT), Germany.

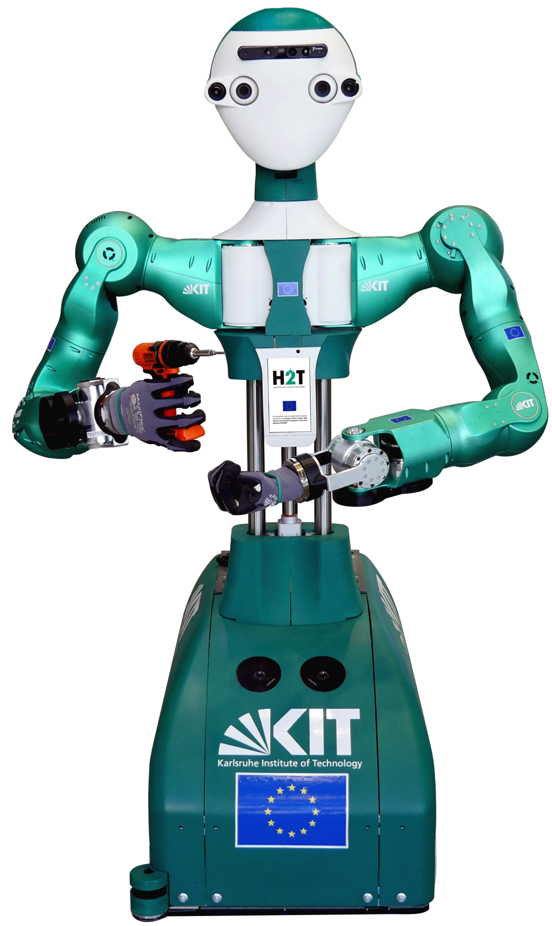

Figure 1. The ARMAR-6 robot: A collaborative humanoid robot.

These technologies have found their way into commercial products in the form of individual arms (e. g. by *Kuka*, *Universal Robots* and *Franka Emika*), dual arm systems (e. g. by *ABB* and *Kawada Industries*) and even complete humanoid robots such as *PAL Robotics*' REEM-C[1] and TALOS [5].

The design, however, is only one aspect of truly collaborative robots. Another crucial aspect are the cognitive abilities of the robot that are needed to correctly interpret situations and understand how to work together with humans, even in environments that are not completely known beforehand. A hardware platform that enables cognitive bimanual mobile

[1]http://reemc.pal-robotics.com/en/reemc/

manipulation capabilities needs to be equipped with various exteroceptive sensors and sufficient on-board computing resources, in addition to the body, arms and end-effectors. Mobility is another important requirement for such a robot, as it drastically expands the working radius, thereby enabling a vast amount of additional applications.

Several complete humanoid systems that integrate these capabilities for real-world use have been introduced recently. The list is long and we therefore limit our presentation to systems designed for dual-arm mobile manipulation. These include the HRP robots that have been applied to aircraft assembly tasks [6] and construction work [7], DLR's Rollin' Justin [8] and the fully torque-controlled TORO [9], as well as robots such as the WALK-MAN [10], the E2-DR [11] and CENTAURO [12] developed for rescue missions in man-made environments. Humanoid robots for highly specialized application areas in the real world include the Ocean One for underwater missions [13] and NASA's Valkyrie for extra-planetary deployment [14].

With ARMAR-6 we set out to advance the state of the art in mobile collaborative robots both in terms of *system design*, i. e. regarding physical capabilities and overall technological readiness for real-world applications, as well as on all aspects of *cognitive capabilities* necessary to accomplish challenging tasks beyond laboratory environments. The presented robot advances the state of the art with respect to kinematics, workspace and workload of the robot. To the best of the authors' knowledge, no other humanoid system exists that combines an arm reach of over 1 m with a carrying capacity of over 10 kg at full arm extension and limitless joints. ARMAR-6 is not only physically capable of demanding tasks but also equipped with the comprehensive sensor suite and the cognitive abilities that facilitate natural and safe collaboration with humans.

ARMAR-6, depicted in Figure 1, is the latest generation in the ARMAR humanoid robot family [15] and has been developed as robotic research platform within the *SecondHands*[2] project. The vision of the SecondHands project is to develop and enable a robot to literally provide a second pair of hands to a human worker in order to support, to increase safety and to enhance efficiency. The development is driven by an ambitious use-case, in which the robot is tasked to autonomously and proactively assist a maintenance technician with repair tasks on a material handling system in a highly automated customer fulfillment center of an online retailer. This particular scenario poses many requirements on the robotic system that challenged us to push the envelope of what is currently possible in collaborative robotics.

The deployment in an actual warehouse environment requires the robot to be extraordinarily robust and reliable. ARMAR-6 therefore incorporates all the experience from the construction and deployment of its predecessors regarding the choice of components, material selection, cable routing and software modules on all levels to realize a highly reliable system. In addition, we have designed the robot, both in terms of hardware and software, to be highly modular. Key components can be exchanged quickly, and subsystems of the robot such as the wheeled platform, the arms or the head, can easily be tested and used separately. Since some of the tasks involve handling heavy machine parts, e. g. the overhead covers of conveyor belts, the robot needs a comparatively high payload capacity. Concretely, we designed ARMAR-6 to handle up to 10 kg in each hand. The tasks require the ability of overhead manipulation well above 2 m as well as picking up objects from the floor. Mobility while carrying heavy objects and the ability to handle tools made for humans are other key requirements.

In terms of cognitive abilities, ARMAR-6 needs to be able to understand a variety of scenes, recognize human actions and intents, learn from human observation and its own experience, process the data produced by cameras, laser scanners and proprioceptive sensors while still acting fluently at human-level speeds. It needs to be able to apply large amounts of forces when needed and at the same time interact gently and predictably with humans, e. g. in the context of kinesthetic teaching.

All of these requirements can only be fulfilled when addressed jointly in the mechanical design of the robot, its actuators, control systems and computing hardware as well as the functional architecture and its efficient implementation in the software, which distributes tasks, ensures reliability and allows adaptation to new tasks and scenarios with little additional programming.

The control software has access to the sensors and actuators of the arms, head and mobile platform via the high-speed EtherCAT bus at a rate of 1 kHz, enabling fast and convenient prototyping of novel control methods. Robot control on all levels is enabled by well documented software interfaces implemented in the ArmarX software framework [16]. Moreover, the arms of the robot provide standardized interfaces complying with ISO 9409-1-50-7-M6 to quickly exchange end-effectors (e. g. different hands or grippers), provided they are 48 V compliant and have an EtherCAT interface.

We introduced the humanoid robot ARMAR-6 in [17]. In the present paper, we elaborate on the reasoning behind the design choices for developing a humanoid maintenance assistant for man-made environments and on the hardware and software co-design. We further put into perspective all our separately published contributions with respect to the overall design, control and software architecture of the ARMAR-6 robot.

In the remainder of this paper, we give an overview over the hardware and software that together make ARMAR-6 one of the most capable, fully integrated platforms for research on mobile compliant dual arm manipulation and human-robot collaboration in industrial environments. We describe the demonstration scenario and show how the various components and design choices, both in terms of hardware and software, enable ARMAR-6 to accomplish seamless human-robot interaction and collaboration during various tasks under challenging circumstances.

[2]https://secondhands.eu

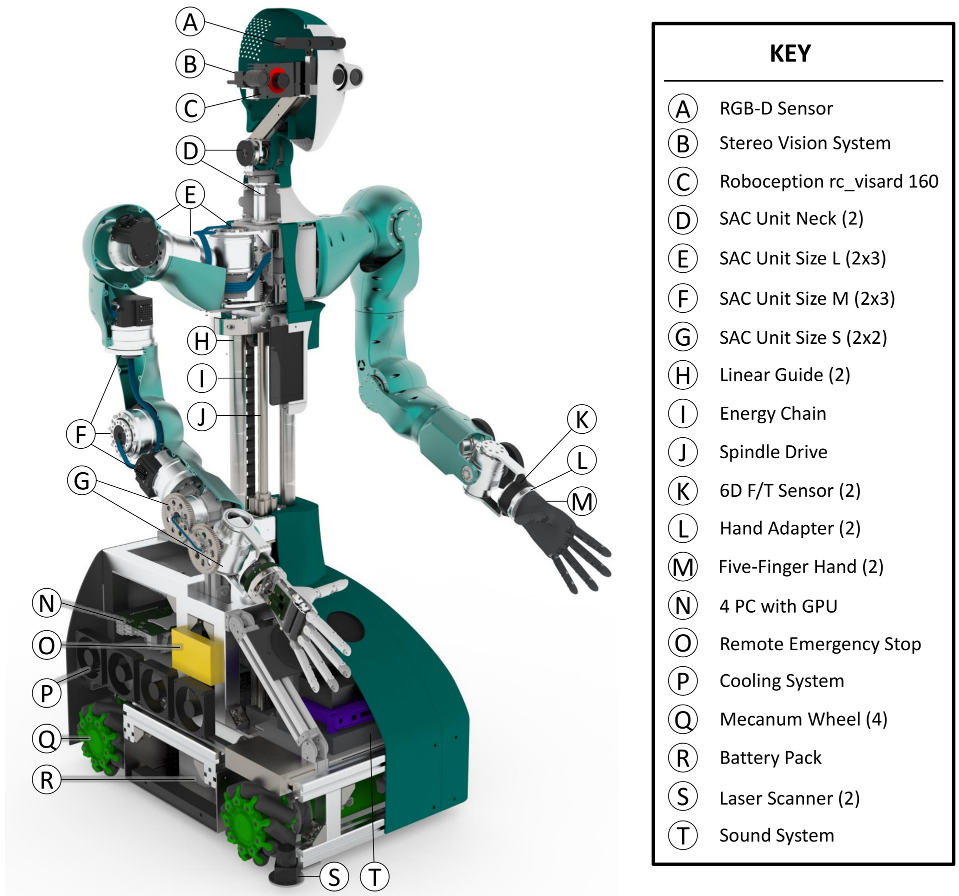

Figure 2. Visualization of essential internal components (A-T) of ARMAR-6.

II. Hardware & Software Co-Design for Embodied Intelligence

The broad set of requirements imposed by the challenging scenario necessitates a principled approach to the design of the robot's hardware and its control software. We chose a fundamentally *modular* design approach for both the hardware and software of ARMAR-6 to be able to split the design into manageable parts. On the hardware side, this modular approach means that each part of the robot can be operated separately and maintains a well-specified interface to integrate with the other components. On the software side, components can be distributed over different machines to ensure the high performance and responsiveness needed for collaborative tasks.

A. Overall Robot Design

ARMAR-6 has 28 active degrees of freedom (DoF), a maximum height of 192 cm and a weight of 160 kg (plus a 40 kg battery pack). On the highest level, the hardware of the robot is split into five different modular robotic systems, i. e. the dual arm system, the five-finger hands, the sensor head, the torso and the mobile platform. An overview over these components and their internal workings can be seen in Figure 2. The following subsections elaborate on the development of these parts as well as the software that controls them. A summary of the hardware is shown in Table I.

B. Dual Arm System with 16 Torque-Controlled DoF

Inspired by the human kinematics, each arm has 8 DoF, including a clavicle joint of the inner shoulder [19]. This additional joint significantly increases the bimanual workspace of the dual arm system [20]. In combination with a reach of 1.3 m and a large range of motion in each joint, the dual arm system has a total workspace of 8.4 m^3 and a bimanual workspace of 4.9 m^3. The structure of the arms follows an exoskeleton design approach: Hollow aluminum parts connect the joints mechanically and simultaneously serve as support structure and covers. This results in a stiff yet lightweight construction, enabling the robot to carry loads up to 20 kg at full arm extension.

To generate a required torque, each arm joint is actuated by a highly-integrated sensor-actuator-controller unit (SAC unit) [18]. These modular SAC units (Figure 3) include a brushless DC (BLDC) motor and a backlash-free Harmonic Drive reduction gearbox with all necessary bearings, as well as a comprehensive sensor suite, communication and motor control electronics. The incremental and absolute position encoders with an absolute accuracy of $0.1°$ and a resolution of 19 bit allow highly precise position and velocity control.

TABLE I
SPECIFICATIONS AND KEY HARDWARE COMPONENTS OF ARMAR-6

Weight				160 kg (without batteries)
Height				152-192 cm
Footprint				60 × 80 cm
Platform speed				1 m/s
Payload				10 kg per arm
Power consumption				460 W (nominal), 1050 W (peak)
DOF	Neck	2	Custom SAC-units based on brushed DC motors, Harmonic Drive reduction gears and custom motor controllers	Pitch and Yaw
	Arms	2×8	Custom SAC-units [18] based on Robodrive brushless DC motors, Harmonic Drive reduction gears and ELMO motor controllers	4 per shoulder, 2 per elbow, 2 per wrist
	Hands	2×14	Maxon Brushed DC motors	Underactuated five-finger hands with one motor for 12 finger joints and one for 2 thumb joints per hand
	Torso	1	Dunker Brushless DC motor with spindle drive and brake and ELMO motor controller	Linear actuator with 40 cm vertical travel
	Platform	4	Donkeymotion drive system with brushless DC motors and brake and ELMO motor controller	Omni-wheels for holonomic motion
Sensors	Head	1	Roboception rc_visard 160 stereo sensor	3D vision
		1	PrimeSense Carmine 1.09	RGB-D
		2	Point Grey Flea3 USB-3 cameras	Stereo vision
	SAC-Units	1	Renishaw AksIM MBA7/8	Absolute position
		1	AMS5306	Incremental position
		1	Texas Instruments ADS1220 ADC	Torque (strain gauge full bridge)
		1	Bosch BNO055 IMU	Acceleration and rotational rate
		5	Various temperature sensors	Temperature of different components
	Arms	2	ATI Mini 45 F/T sensor	6D force-torque sensing
	Torso	1	Waycon Draw Wire Sensor SX50	Absolute vertical position
	Platform	2	Hokuyo UST-10LX 2D lidars	Laser range finders for orientation and safety
Power Supply	Internal		NiMH battery with 38 Ah @ 48 V (1824 Wh)	Power-Autonomy for up to 4 h
	External		48 V power supply	Tethered power for development and testing
Computers (4)	Real-time		Mini-ITX, Core-i7, 32GB RAM, Ubuntu, ArmarX	RT-PREEMPT patch and SOEM EtherCAT master
	Vision			GeForce GTX-1080 GPU
	Speech			Roland USB sound card
	Planning			ArmarX master node
Communication	Internal		EtherCAT Bus	Real-time automation bus connecting all actuators to the realtime computer; 1 kHz bus update rate
	External		GigaBit Ethernet	Either via LAN, 2.4 GHz or 5 GHz WLAN via the internal router and switch
Robot Development Environment			ArmarX	Middleware, Comprehensive high- and low-level APIs
User Interface			ArmarX	GUI, natural language, various programming languages

To enable safe human-robot interaction, we pursue an active compliance approach. Therefore, each SAC unit includes a strain-gauge-based torque sensor with a sensitivity of more than $0.04\,\mathrm{Nm}$. Combined with fast real-time sensor signal transmission provided by the EtherCAT bus, this allows for precise torque control at a rate of $1\,\mathrm{kHz}$.

For highly integrated mechatronic systems, cabling is crucial for overall system reliability. We use slip rings, which allow continuous cable rotation between rotating and non-rotating parts of the joints. As a result, each arm has three limitless joints, namely the rotation of the shoulder, the upper arm and the forearm.

In addition to high integration and robustness, modularity and encapsulation are key design principles for the SAC units. Electrically, the daisy-chained SAC units expose minimal interfaces: The input consists of connectors for the communication and DC power bus as well as a separate emergency stop (e-stop) cable. The output consists of the corresponding three connections. In between units, the three cables (communication, power and e-stop) are routed inside the hollow structure, protecting them from mechanical stress and external interferences. The high modularity also simplifies the maintenance of the robot arm, since the SAC units can be easily exchanged if needed. To achieve or even surpass human performance with anthropomorphic appearance and kinematics, the units have been designed to be scalable: The three different SAC unit types provide peak torques between 63 and $176\,\mathrm{Nm}$ at maximum rotational speeds between 79 and $206\,^{\circ}/\mathrm{s}$.

The last SAC unit in the wrist is followed by an additional 6D force-torque (F/T) sensor, which provides the robot with haptic feedback during manipulation and human-robot interaction tasks. It is attached to a SCHUNK hand adapter that provides a well-defined mechanical and electronic interface. This allows to quickly exchange different end-effectors and to adapt the robot to new scenarios.

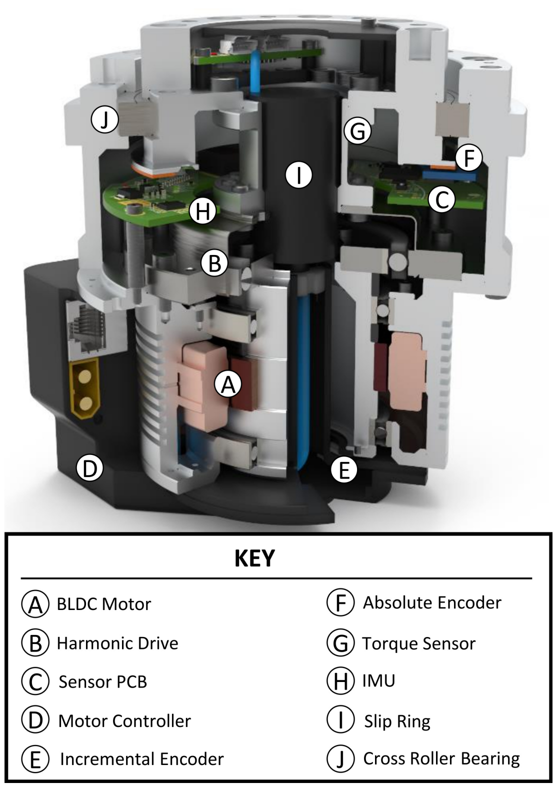

Figure 3. Rendered 3D cross-section of the mid-sized sensor-actuator-controller (SAC) unit. The dual arm system of ARMAR-6 has 16 highly integrated SAC units in three different sizes.

C. Underactuated Five-Finger Hands

For the operation in human-centered environments, ARMAR-6 is equipped with two five-fingered hands. The anthropomorphic shape and kinematics based on human proportions allow manipulation of different tools and other man-made objects. Each of the 14 finger joints (two in the thumb and three per finger) is supported by two ball bearings. Actuation of the joints is realized using Dyneema tendons guided in PTFE tubes for reduced friction.

The four fingers and the thumb are actuated by one motor respectively. The motor torque is distributed to the fingers using an adaptive underactuation mechanism based on the TUAT/Karlsruhe mechanism [21]. This mechanism allows the fingers to wrap around objects and to adapt automatically to their shapes.

As the hands are used in maintenance tasks involving heavy objects, they have to be robust and light at the same time to reduce the load on the arm. Hence, most of the structural hand parts are 3D printed from durable polyamide using selective laser sintering (SLS). All cables are hidden in the interior of the hand to avoid electric failures. The hand itself is covered by a protective work glove.

D. Sensor Head

The sensor head equips the robot with a range of visual sensors. It includes five cameras, divided into three independent systems: The first system is a Roboception rc_visard 160, a $16\,\mathrm{cm}$ baseline stereo camera system with on-board depth

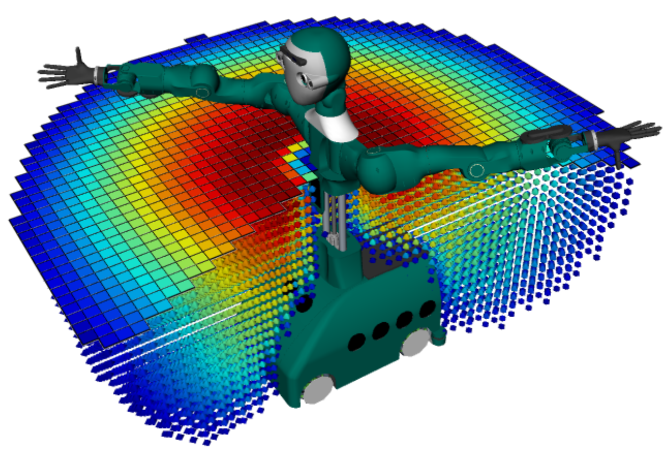

Figure 4. Sectional view through the $10.7\,\mathrm{m}^3$, nearly ellipsoidal workspace of the dual arm system incorporating the prismatic joint of the torso. The color of the voxels correspond to the local manipulability as described by Yoshikawa [22] with the extension that the manipulability maps for both arms are merged into one manipulability map. The warmer the color, the higher the manipulability.

image processing. It is used for peripheral vision with an optimal depth measurement range from $0.5\,\mathrm{m}$ to $3\,\mathrm{m}$. The second stereo camera system, a pair of Point Grey Flea3 cameras, has a baseline of $27\,\mathrm{cm}$ and is used for foveal vision. These two cameras are placed on both sides of the Roboception sensor with parallel image planes. The third system is a PrimeSense active RGB-D sensor, which is particularly advantageous for uni-colored, featureless surfaces.

The camera systems are mounted on a common aluminum frame, whose direction is controlled by a pan-tilt unit in the neck. Its two joints are driven by custom-built, modular sensor-actuator units, a new miniaturized version of the SAC units. Hence, the modules combine the same highly precise absolute position encoder with a backlash-free Harmonic Drive reduction gearbox and stiff structure to achieve a high positioning accuracy of the camera systems. In combination with the speed of up to $510\,°/\mathrm{s}$ and the high control frequency enabled by EtherCAT, the sensor head is precise and fast, enabling effective pose tracking and gaze stabilization. Similar to the hands, the head and neck are protected by 3D printed cover parts.

E. Mobile Platform and Torso

To enable the robot to support technicians in different areas of the warehouse, mobility is a key requirement. We designed an omnidirectional, wheeled mobile platform which meets all requirements of an industrial workplace. The platform offers a space for all necessary components to allow the robot to operate autonomously and without external cables. It contains four computers and a GPU for on-board computing, a sound system, network peripherals, the power management system and a $1.8\,\mathrm{kWh}$ battery pack that enables up to $4\,\mathrm{h}$ of power-autonomous operation under nominal operating conditions. All components are covered and protected by a fiberglass housing.

To increase the robot's workspace it has an extensible torso, realized as a single prismatic joint, driven by a DC motor with

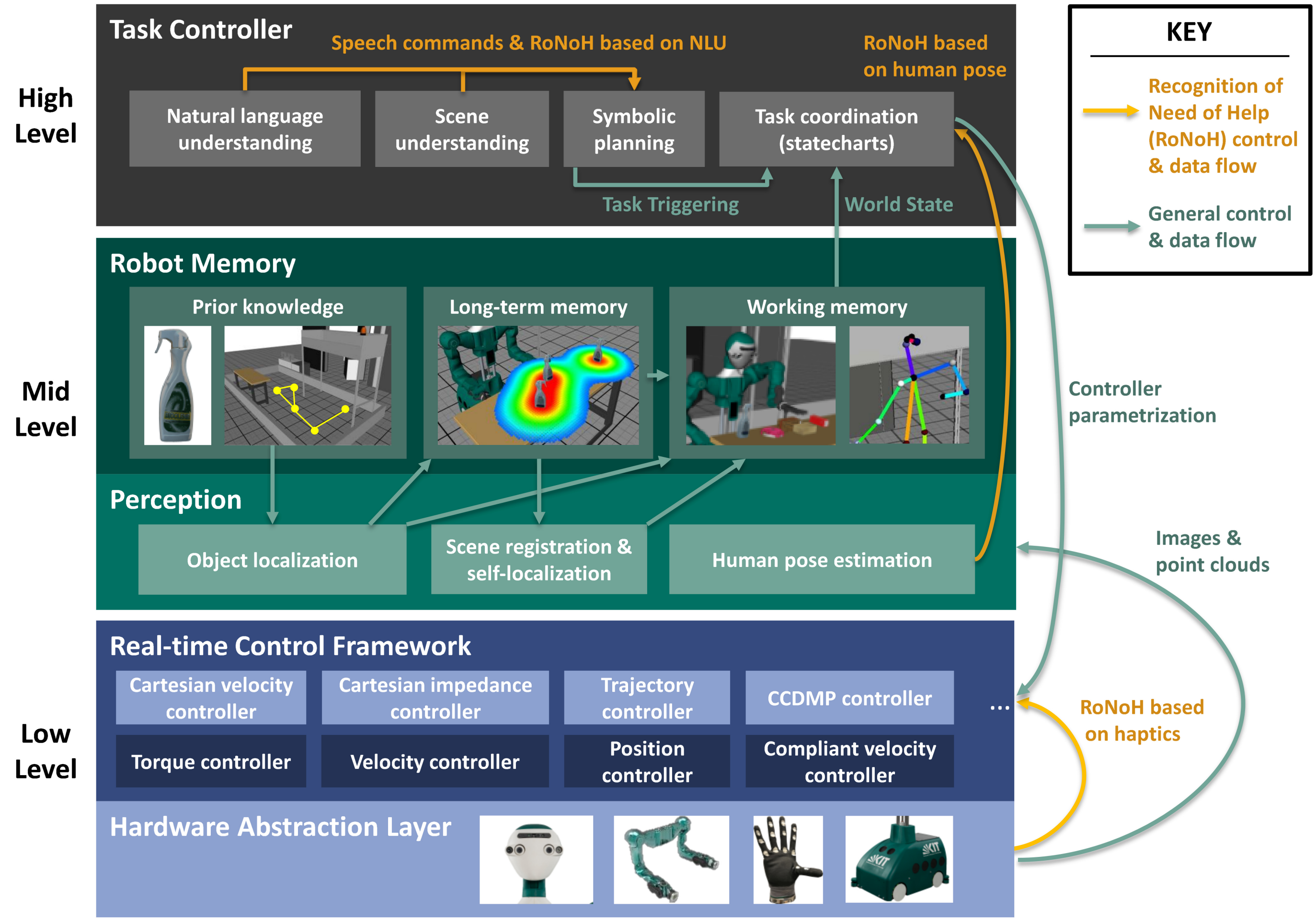

Figure 5. Overview of ArmarX as the functional software architecture of ARMAR-6. The depiction is divided into its three main conceptual levels and their principal interconnections. The information is increasingly abstracted from the low level (e. g. raw sensor values) to the high level (e. g. symbolic predicates).

a spindle drive and guided by two parallel linear bearings. The torso joint allows the robot to change its height by 40 cm. This increases the workspace of the dual arm system from 8.4 m^3 to 10.7 m^3 (Figure 4). As a result, ARMAR-6 is able to pick up tools from the ground as well as to hand over a tool to a technician on a ladder at a height of 2.4 m. A brake keeps the height of the shoulders constant when desired and thereby ensures energy-efficient actuation. The absolute height of the torso is determined by a draw wire sensor. All cables between the upper body and the platform are routed through an energy chain which protects them and avoids mechanical stress.

For safe human-robot interaction, in addition to the comprehensive sensor setup of the dual arm system (in particular the torque and force/torque sensors) and the brakes in the torso and the mobile platform, ARMAR-6 also features a wireless and a wired remote emergency stop. The platform further contains two planar laser scanners that provide 360° distance information, which is used for safety, obstacle detection and map-based navigation.

F. Functional Software Architecture

The software architecture of ARMAR-6 satisfies similar requirements as the hardware, i. e. modularity, re-usability and scalability. All our recent humanoid robots [2], [23] share the same cognitive and functional software architecture implemented in ArmarX[3] [24]. For a complex humanoid robot system, ArmarX is not only a robot middleware, but also a complete functional cognitive architecture offering ready-to-use modules on multiple levels of abstraction. Figure 5 shows the key modules of the architecture, which are divided into three levels of abstraction: 1) Real-time control and hardware abstraction, 2) perception and the robot memory system and 3) the task planning. On the lowest level, almost all data is subsymbolic. The hardware abstraction layer connects with the robot-agnostic real-time control framework. This abstraction level provides its data to the upper levels and receives commands from the highest level. On the mid-level, perception modules for sensor processing such as object localization, self-localization and human pose estimation are located. Their results are stored in the ArmarX memory system MemoryX, which includes modules such as the working memory for a consistent world state, long-term memory for self-experienced data and the prior knowledge, containing data provided by robot programmers (such as 3D object and world models) as well as pre-programmed trajectories. The result of the

[3]ArmarX: Publicly available at https://armarx.humanoids.kit.edu

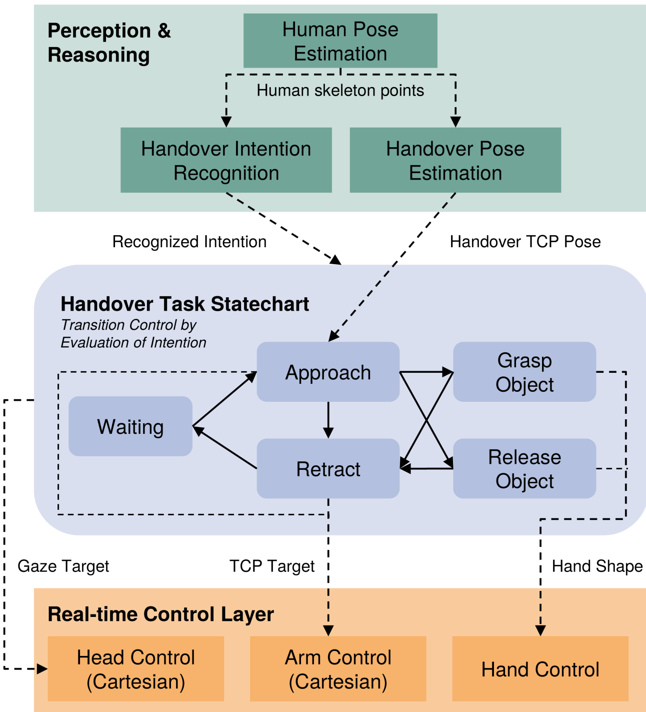

Figure 6. Flowchart of the handover task showing the coordinating statechart as well as the involved perception and control modules needed for an interactive and adaptive handover.

sensor processing is usually represented in (but not restricted to) easy-to-process formats such as 6D poses or symbolic relations. On the highest abstraction level, reasoning modules, such as natural language understanding, symbolic planning and statecharts, coordinate the robot subsystems dominantly on a symbolic level.

Figure 6 shows an abstract depiction of an exemplary statechart (in this case for the handover task described in subsection III-C) and the interconnections between respective perception and control modules. All modules run in parallel and provide their data asynchronously to the statecharts, which generate targets for the controllers that are running synchronously in the real-time layer.

One key principle of the robot architecture is the modularity of the software based on common interfaces. This allows to easily exchange or insert new software modules and use them on multiple robot platforms, which is especially important for the purpose of research. Due to this modularity, failures of individual modules do not affect the complete system. The complexity of the robot's software system results in a high number of interconnected modules, which create a complex network of dependencies. To be able to assess the system state, ArmarX offers various monitoring tools on sensor and module level. This enables advanced safety features: If any software module that is configured as *system-relevant* stops working, all robot actuators go immediately to an emergency stop, where they hold the current position with reduced torque.

To allow the reuse of modules on different robot hardware platforms, hardware abstraction is realized on two levels: The first level is a generic high-performance real-time control framework in ArmarX, which offers implementations of general real-time motion controllers, which can be directly used for different robots, in a synchronized manner. They are accessible to other modules via the middleware. On the second level, simple and unified interfaces offer the basic control functionality and make all sensor values available for modules in the robot network.

Although the real-time control framework is designed for a general purpose, it was co-designed with the specific hardware of ARMAR-6 to ensure that the framework fulfills the constraints of the real hardware. Since ARMAR-6 uses the high-performance real-time bus EtherCAT, the software framework needs to be able to fulfill the real-time constraints while still being extensible and modular. This is mainly achieved by template programming and inheritance throughout the whole framework as well as by pre-allocating all resources needed for the current hardware setup.

With this real-time control framework, ARMAR-6 offers state-free, basic controllers for individual joints such as position, velocity or torque controllers, but also multi-joint Cartesian controllers as well as controllers for trajectory execution, impedance control and dynamic movement primitives. Further, the control framework offers unrestricted access to all sensor values and low-level controllers in a synchronous control-loop at 1 kHz. More complex or non-real-time controllers run their own control loop in a separate thread and synchronize their control targets in a lock-free manner with the real-time control loop.

For scalability, ArmarX relies on a multi-threaded and distributed design concept of modules. Every module can run transparently on any machine in the robot network. Additionally, there is no central processing loop, which allows to run every module at individual frequencies. This ensures that the large number of components (i. e. sensors and actuators) of ARMAR-6 does not affect the overall performance of the robot.

III. Autonomous Abilities for Physical Human-Robot Collaboration

The hardware and software components described in the previous section endow ARMAR-6 with the abilities required in the context of safe, physical human-robot collaboration. This section will highlight a few of them in isolation, before we describe an integrated validation scenario leveraging a variety of these abilities in the next section. All abilities are implemented as software modules, statecharts or real-time controllers within ArmarX.

A. Task Space Impedance Control and Bimanual Control

Compliant control is a crucial part of a collaborative robot's capabilities to allow safe human-robot interaction. We developed a task space impedance control scheme using joint level torque control described in [25] that allows compliant arm motions. We model the robot hand motion as driven by a damped spring system guided by an attractor to a moving target. These targets are generated by a dynamic movement primitive (DMP) learned in either joint or task space. Figure 7

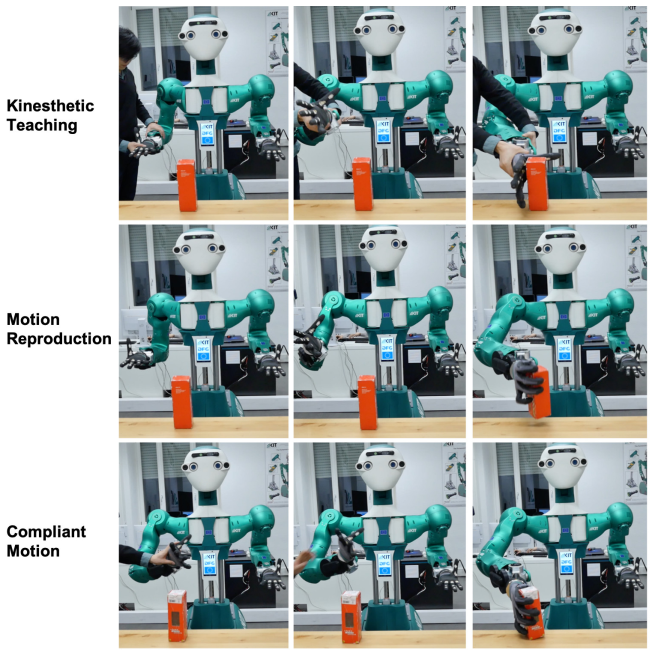

Figure 7. Kinesthetic teaching and compliant motion reproduction based on task space impedance control. During motion execution, the robot is able to compliantly adapt its action to external force perturbations and to safely interact with humans.

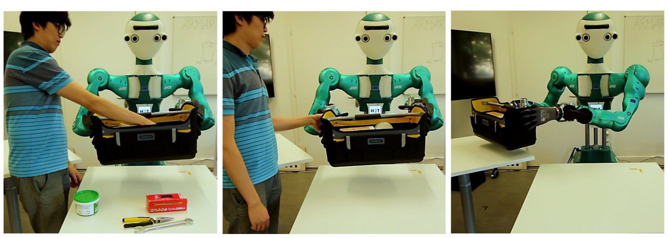

Figure 8. The compliant bimanual control system allows manipulating large and heavy objects. **Left:** Holding a toolbox filled with objects with a total weight of 6kg. **Middle:** The coupling of both arms is guaranteed by the bimanual controller during perturbations. **Right:** The motion is encoded with CC-DMP and the forces exerted on the toolbox are generated by the bimanual controller.

shows the use of the task space impedance control during a grasping motion learned from kinesthetic teaching. In particular, we investigated two methods for bimanual manipulation tasks such as carrying large and heavy objects. The first method, Coordinate Change Dynamic Movement Primitive (CC-DMP [26]), combines two DMPs describing the motion of each arm, coupled in a leader-follower manner. The second method is based on force admittance control as described in [27]. An example task with the bimanual controller is shown in Figure 8, where the robot holds a large and heavy object and puts it to the side. The motion is encoded by CC-DMP while the bimanual controller guarantees persistent coupling, even during external perturbations.

B. Grasping of Known and Unknown Objects

ARMAR-6 has the ability to grasp known, familiar and unknown objects. For grasping known objects, grasps are

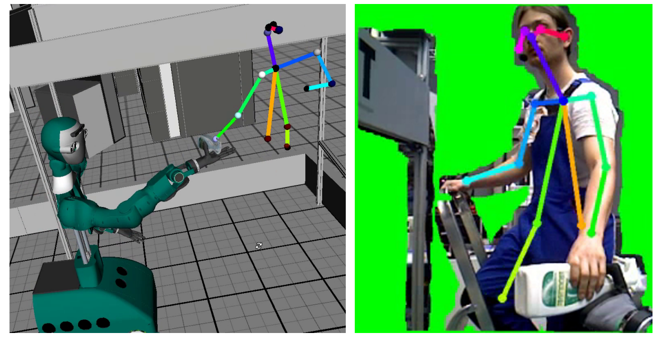

Figure 9. Active handover of a tool between ARMAR-6 and the human technician. The robot uses marker-free 3D human pose estimation and force feedback to recognize the human's intentions to provide help when needed.

determined offline using a grasp planner that exploits topological object information, represented as an object skeleton, to identify suitable grasping regions. It selects the grasp type and aligns the hand according to local skeleton and surface information [28]. The subsequent motion planning includes planning of time-optimal collision free trajectories for fast grasping as well as a human-like approach direction and hand orientation.

For grasping familiar objects, part-based grasping methods are used to generate grasps based on the object shape using machine learning methods, which are trained on a set of various object classes (see [29] and [30]).

In the case of unknown objects, ARMAR-6 combines a deep learning grasping approach with reactive grasp execution [31]. Based on fused tactile and visual data, a neural network generates several grasp candidates with associated grasp success probability ratings. To cope with model uncertainty while executing the best-rated grasp candidate, a reactive grasping scheme is applied. To this end, the 6D force/torque sensor in the wrist is used to infer the contact point between the fingers and the object. Based on the force sensor data, the grasp pose is iteratively refined, adjusted and executed.

C. Human Activity Recognition and Active Handover

One central aspect of human-robot interaction is to recognize and predict the human actions. To estimate the 3D human pose, we use state-of-the-art marker-free methods (based on [32] and [33]), which use RGB-D images as input and run on the on-board GPU at $\sim 15\,\mathrm{Hz}$. A neural-network classifier uses the human pose to estimate the current activity of the human. Furthermore, the human pose provides the input for the active handover skill of the robot, which is implemented using the handover controller described in [34]. This controller receives the human hand pose to adapt the approach motion during the handover task. Force feedback from the 6D force/torque sensor in the wrist is used to trigger hand opening and closing actions in robot-human and human-robot handover tasks. Figure 9 shows the 3D visualization of the working memory of the robot and an overlay skeleton on the input image during the handover execution. Figure 6 visualizes the software modules involved in the handover task.

D. Recognition of the Need of Help

One of the key aspects of a robot assistant is the ability to recognize that a human co-worker needs help, and to provide such help pro-actively.

In our work, we investigate how recognizing the need of help can be implemented based on different sensory modalities provided by the robot: visual information, haptic and force information, as well as speech and dialog. Using natural language provides an intuitive way for communicating the need of help either using speech commands to directly request help, or based on the task-specific interpretation of utterances provided by the natural language component. *Vision-based* human pose tracking and activity recognition provides a second way of recognizing the need of help during task execution. Given a description and all required objects and actions of a collaborative task, usually performed by two humans, the robot can infer that a second agent is missing by interpreting the current scene. This allows the robot to pro-actively offer help by executing the task of the missing agent. The third way for recognizing the need of help is the use of force and haptic information in collaborative human-robot manipulation tasks such as jointly carrying a large object. Based on irregular force patterns and/or sudden changes of these patterns during the task, the robot is able to infer that the human collaboration partner is struggling and a regrasping action may provide help as it will reduce the load on the human. We implemented and integrated the above strategies and demonstrated the performance in several complex warehouse maintenance tasks.

E. Natural Language Understanding

The natural language dialog system of ARMAR-6 enables intuitive human-robot communication. The speech recognition and dialog system is a deep learning architecture trained in an end-to-end fashion. It is implemented as an attention-based encoder-decoder network [35] while the natural language understanding is refined using a multi-task learning approach [36]. ARMAR-6 can react to a large variety of direct commands (e. g. ”Bring me the spray bottle”), but can also infer instructions from spoken language utterances to suggest providing help in a task or to ask for missing information. When the technician tells the robot that the conveyor drive-train needs to be cleaned for example, ARMAR-6 will infer that the cleaning fluid is needed and will assist the technician by bringing it and handing it over.

IV. Assisting Humans in Maintenance Tasks

To validate the capabilities of ARMAR-6 against a realistic benchmark, we devised a challenging demonstration scenario that requires advanced scene understanding, human intention recognition, mobile manipulation and physical human-robot interaction.

A. Demonstration Scenario

Derived from the goals of the SecondHands project, the scenario represents a routine inspection and maintenance of an overhead conveyor system in an automated customer fulfilment center. This work is typically carried out by two human technicians, one of which is the expert while the other assumes the role of a subworker. This latter role is physically demanding, and the cognitive tasks that are involved make it challenging for humanoid robots. In our scenario, ARMAR-6 is the subworker that is tasked to seamlessly and effectively support the expert technician. A graphical overview over the scenario as well as the robot's tasks, actions, primary sensor modalities and ways of human-robot interaction are depicted in Figure 10. An uncut video of this scenario is available online[4]. The scenario is split into two main parts. These two parts, i. e. removing the cover panel and fetching a spray bottle, are of very different nature. While the first part requires very close, seamless and compliant physical human-robot collaboration, the second part requires mostly autonomous mobile manipulation abilities, including robust and comprehensive scene perception and interpretation.

1) Compliant Collaborative Manipulation: The technician realizes that the conveyor system is defective and informs ARMAR-6, using plain English. With the continuously active automatic speech recognition (ASR) system [37], ARMAR-6 understands this information and infers the imminence of a maintenance task that it is able to assist in. This pro-active, multimodal *recognition of the need of help* is one of the key cognitive abilities we are investigating in the SecondHands project, as it represents a crucial enabling technology for seamless human-robot interaction. Having realized that help is needed, the robot follows the technician, who proceeds without waiting to inspect the defective conveyor system.

When the technician starts to *remove the cover panel* (depicted in Figure 10, top) of the conveyor to gain access to the drive train, the robot recognizes this action and concludes that its help, in the form of supporting the cover panel on the opposite end, is needed. This recognition is based on visual 3D human pose tracking (based on [32], [33]) rather than natural language understanding. Using its laser-based self-localization system, the robot finds the right position under the conveyor and raises both hands to support the panel. ARMAR-6 relies on its tactile sensing (enabled by its wrist-mounted 6D force-torque sensors) to gently make contact with the cover panel and grasp it. Since the geometry of the cover panel is known to the robot, suitable grasps for both hands and the motion sequences for both arms are defined as prior knowledge. Once in place, ARMAR-6 verbally confirms that it is ready to lower the panel. As soon as the technician has fully unmounted the panel, which the robot again realizes through haptic feedback, the human-robot team jointly lowers it. Once lowered, the robot confirms in natural language that it is ready for the next step. To carry the panel to a temporary location, the technician has to step around the posts of the conveyor setup, which requires him to grasp the cover panel on one of its corners. ARMAR-6 realizes that the panel is out of balance and infers through tactile sensing that it can help the technician by regrasping it on the opposite corner. It does so and again verbally confirms its readiness for the next step.

[4]https://youtu.be/6AFkrGkKI7g

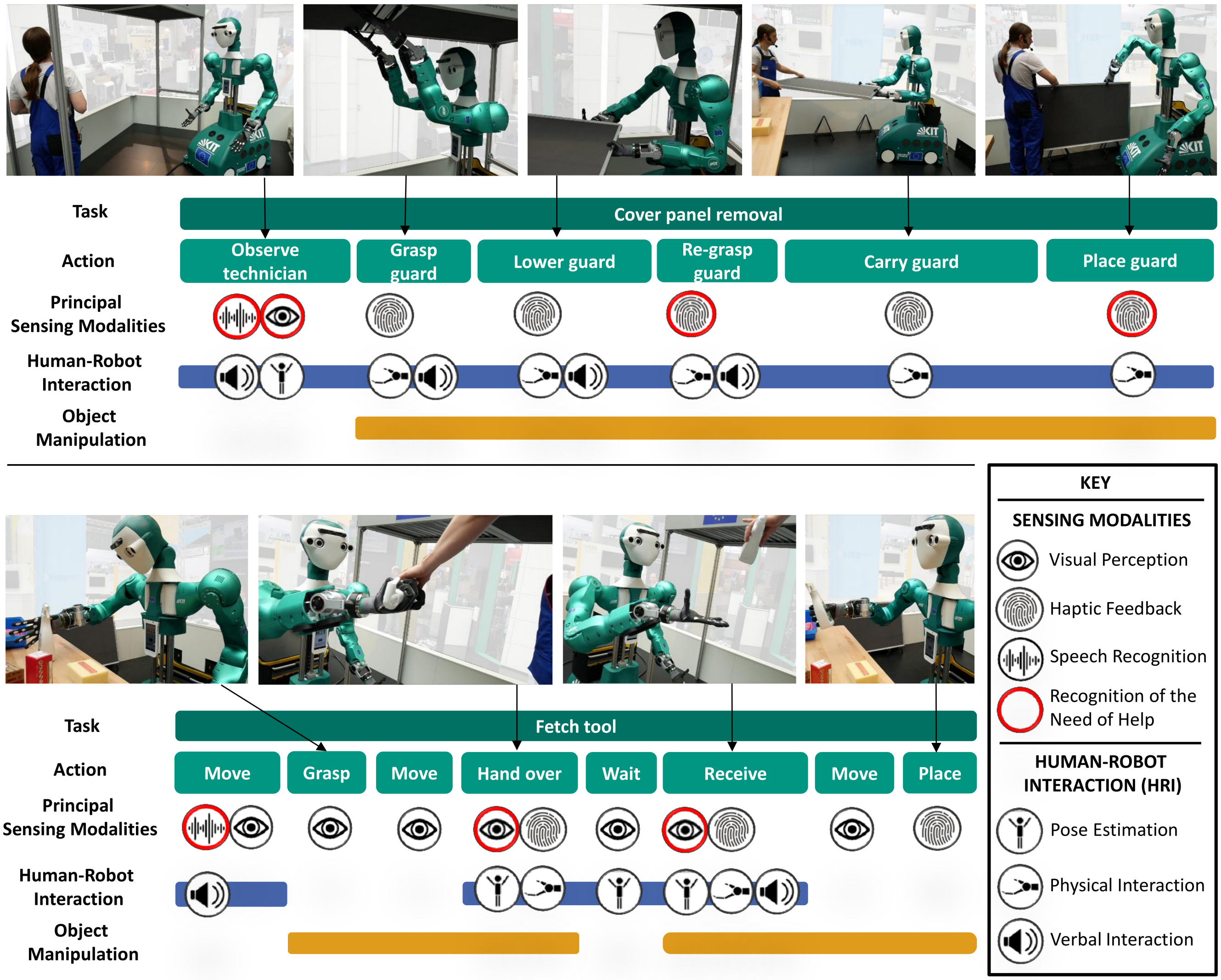

Figure 10. The warehouse demonstration scenario broken down into tasks, actions, principal sensor modalities, human-robot interaction modalities and object manipulation. On the top, the compliant collaborative manipulation is shown in a *cover panel removal* task. On the bottom, autonomous mobile manipulation is demonstrated with *fetching a tool* for the technician.

The technician then starts to walk towards the intended placing location of the cover panel. ARMAR-6 has no a-priori notion of this location. It is instead guided by the movement of the technician that it senses through the displacement of its own hands. This force-based method for guiding is enabled by joint-level compliance in every joint of the robot's arms, which is one of the principal joint control modes of ARMAR-6 and has proven to be invaluable for physical human-robot interaction.

Once the cover panel is at the intended location, the technician starts to initiate a placing movement. The robot senses the onset of this motion with its wrist-mounted 6D force-torque sensors, recognizes that it can assist the technician with placing the heavy object and in turn starts the placing motion. As the robot's arms are still compliant, the technician can determine the exact placing position while the robot simply supports the weight of the panel.

2) Autonomous Mobile Manipulation: After the cover panel is removed, the technician inspects the exposed inner workings of the conveyor and decides that the drive train needs to be cleaned. He expresses using natural language what needs to be done by saying *the drive train needs to be cleaned* (depicted in Figure 10, bottom). ARMAR-6 understands this message using its ASR system and recognizes that its help with the cleaning task is needed. Concretely, ARMAR-6 infers that cleaning requires a cleaning agent, a spray bottle on the nearby table, which needs to be localized and grasped by the robot in order to hand it over to the technician, who has already mounted a ladder to access the drive train.

To hand over the spray bottle to the human, ARMAR-6 observes the technician, again using 3D human pose tracking. This action is shown in Figure 9 as 3D visualization of the working memory of the robot and an overlay skeleton on the input image of the RGB-D sensor.

Once the technician extends the hand towards the robot, ARMAR-6 recognizes the activity and initiates the handover. If the technician retreats the hand, ARMAR-6 will also retreat from the handover and keeps observing the human. However, if the technician takes the bottle from the robot's hand, it senses this with its 6D force-torque sensors and opens the hand to complete the handover action. Once the cleaning task is finished, the technician will handover the spray bottle back to the robot. Again, the handover intention is recognized and the robot's hand is closed as soon as contact is detected. While the technician stows the ladder, the robot moves back to table and returns the spray bottle to its initial location.

B. Key Aspects

Figure 10 only indicates the primary sensor modality used in each action. We want to emphasize that most of the perception systems, from laser-based navigation to ASR and 3D visual perception, are constantly active, providing the robot with rich information for situational awareness and scene understanding. Throughout the entire scenario, the robot recognizes on multiple occasions that the technician needs its help (red circles in Figure 10). The automatic recognition of the need of help enables the human and the robot to work together naturally.

We have executed system validation studies using the described scenario at different locations and with more than 20 different technicians. Most notably at our lab at KIT, at the international CEBIT 2018 trade fair in front of large audiences, and at an actual automated fulfilment center in the UK. The usability and impact of ARMAR-6 for the maintenance tasks in this fulfilment center with a *System Usability Scale* questionnaire has been evaluated in a user study conducted in collaboration with SecondHands project team members of EPFL. In this study, the technicians were also asked for their subjective perception of ARMAR-6 as coworker using a *Godspeed Questionnaire Series*. Details of the user study are described in [38]. The result of the evaluation helps us to continuously improve on-board algorithms for more intuitive human robot interaction.

A few of ARMAR-6's additional abilities, which are not showcased in the described demonstration scenario, are motion planning in unknown and dynamic environments, grasping of unknown objects, and the effortless teach-in of complex motions using the gravity-compensated zero torque control of the arms.

V. Conclusion

In this paper, we provide a comprehensive overview of the current capabilities of ARMAR-6, as well as the vision we are pursuing in our research. We presented the requirements-driven design and implementation of ARMAR-6 as a platform for collaborative robotics, and the design choices we made towards this goal. The joint development of the hardware and software with the common goal of creating a robot that pushes the envelope for seamless physical human-robot interaction has led to a highly integrated, versatile and robust humanoid robot system. The frequent demonstration of ARMAR-6 in its intended scenario, where it assists a human technician in repairing a conveyor system of an automated warehouse, not only in the laboratory but also at exhibitions and at an actual industrial warehouse, highlights its capabilities, robustness and adaptability.

We have only just begun to explore the possibilities for research on human-robot collaboration that ARMAR-6 as an integrated robot platform facilitates. Among others, we are currently developing methods to further improve the robot's situational awareness and human intention recognition by means of integrating multi-modal sensory input, prior knowledge, and an episodic memory that allows recalling previous experience. The ability to learn from human observation and experience is essential and thus we are transferring and extending our previous work in this area to allow intuitive robot programming. Despite the robust execution of the complex tasks described in this paper, failure will occur in different situations and thus we are working on methods for failure prediction and recovery on the different abstraction layers of the underlying functional architecture. We are continuously upgrading the robot's hardware itself (e. g. improved hands, improved neck actuators) and the low-level controllers, as we gain new insights into the requirements through extensive use of the robot. Finally, we are working on improving the robot's dexterous manipulation skills, with a focus on grasping and manipulation of unknown objects, e. g. by leveraging simulation-based learning.

Acknowledgment

The research leading to these results has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 643950 (SecondHands). The authors would like to acknowledge the help and support of all other members and students of the H^{2}T lab as well as all members of the SecondHands project team, especially Aude Billard and Mahdi Khoramshahi from EPFL, Stefan Constantin, Sebastian Stüker and Alex Waibel from KIT, Fiora Pirri from Sapienza University of Rome, Lourdes Agapito from University College London, and Graham Deacon from Ocado Technology. Special thanks to Michael Suppa of Roboception for providing one of their earliest sensors to ARMAR-6.

References

[1] A. Albu-Schäffer, S. Haddadin, C. Ott, A. Stemmer, T. Wimböck, and G. Hirzinger, "The dlr lightweight robot: design and control concepts for robots in human environments," *Industrial Robot: an international journal*, vol. 34, no. 5, pp. 376–385, 2007.

[2] T. Asfour, K. Regenstein, P. Azad, J. Schröder, N. Vahrenkamp, and R. Dillmann, "ARMAR-III: An Integrated Humanoid Platform for Sensory-Motor Control," in *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, Genova, Italy, December 2006, pp. 169–175.

[3] C. Ott, O. Eiberger, W. Friedl, B. Bauml, U. Hillenbrand, C. Borst, A. Albu-Schaffer, B. Brunner, H. Hirschmuller, S. Kielhofer *et al.*, "A humanoid two-arm system for dexterous manipulation," in *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, 2006, pp. 276–283.

[4] N. Kashiri, J. Malzahn, and N. G. Tsagarakis, "On the sensor design of torque controlled actuators: A comparison study of strain gauge and encoder-based principles." *IEEE Robotics and Automation Letters*, vol. 2, no. 2, pp. 1186–1194, 2017.

[5] O. Stasse, T. Flayols, R. Budhiraja, K. Giraud-Esclasse, J. Carpentier, J. Mirabel, A. Del Prete, P. Souères, N. Mansard, F. Lamiraux *et al.*, "TALOS: A new humanoid research platform targeted for industrial applications," in *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, 2017, pp. 689–695.

[6] A. Bolotnikova, K. Chappellet, A. Paolillo, A. Escande, G. Anbarjafari, A. Suarez-Roos, P. Rabaté, and A. Kheddar, "A circuit-breaker use-case operated by a humanoid in aircraft manufacturing," in *IEEE Conference on Automation Science and Engineering (CASE)*, 2017, pp. 15–22.

[7] K. Kaneko, H. Kaminaga, T. Sakaguchi, S. Kajita, M. Morisawa, I. Kumagai, and F. Kanehiro, "Humanoid robot hrp-5p: an electrically actuated humanoid robot with high power and wide range joints," *IEEE Robotics and Automation Letters (RA-L)*, pp. 1431–1438, 2019.

[8] M. Fuchs, C. Borst, P. R. Giordano, A. Baumann, E. Kraemer, J. Langwald, R. Gruber, N. Seitz, G. Plank, K. Kunze *et al.*, "Rollin'Justin - design considerations and realization of a mobile platform for a humanoid upper body," in *2009 IEEE International Conference on Robotics and Automation*. IEEE, 2009, pp. 4131–4137.

[9] J. Englsberger, A. Werner, C. Ott, B. Henze, M. A. Roa, G. Garofalo, R. Burger, A. Beyer, O. Eiberger, K. Schmid *et al.*, "Overview of the torque-controlled humanoid robot toro," in *2014 IEEE-RAS International Conference on Humanoid Robots*. IEEE, 2014, pp. 916–923.

[10] N. G. Tsagarakis, D. G. Caldwell, F. Negrello, W. Choi, L. Baccelliere, V.-G. Loc, J. Noorden, L. Muratore, A. Margan, A. Cardellino *et al.*, "Walk-man: A high-performance humanoid platform for realistic environments," *Journal of Field Robotics*, vol. 34, no. 7, pp. 1225–1259, 2017.

[11] T. Yoshiike, M. Kuroda, R. Ujino, H. Kaneko, H. Higuchi, S. Iwasaki, Y. Kanemoto, M. Asatani, and T. Koshiishi, "Development of experimental legged robot for inspection and disaster response in plants," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017, pp. 4869–4876.

[12] N. Kashiri, L. Baccelliere, L. Muratore, A. Laurenzi, Z. Ren, E. Mingo Hoffman, M. Kamedula, G. F. Rigano, J. Malzahn, S. Cordasco, P. Guria, A. Margan, and N. Tsagarakis, "Centauro: A hybrid locomotion and high power resilient manipulation platform," *IEEE Robotics and Automation Letters (RA-L)*, pp. 1595–1602, 2019.

[13] O. Khatib, X. Yeh, G. Brantner, B. Soe, B. Kim, S. Ganguly, H. Stuart, S. Wang, M. Cutkosky, A. Edsinger *et al.*, "Ocean one: A robotic avatar for oceanic discovery," *IEEE Robotics & Automation Magazine*, vol. 23, no. 4, pp. 20–29, 2016.

[14] N. A. Radford, P. Strawser, K. Hambuchen, J. S. Mehling, W. K. Verdeyen, A. S. Donnan, J. Holley, J. Sanchez, V. Nguyen, L. Bridgwater *et al.*, "Valkyrie: Nasa's first bipedal humanoid robot," *Journal of Field Robotics*, vol. 32, no. 3, pp. 397–419, 2015.

[15] T. Asfour, R. Dillmann, N. Vahrenkamp, M. Do, M. Wächter, C. Mandery, P. Kaiser, M. Kröhnert, and M. Grotz, "The karlsruhe armar humanoid robot family," A. Goswami and P. Vadakkepat, Eds. Dordrecht: In: Goswami A., Vadakkepat P. (eds) Humanoid Robotics: A Reference, Springer Netherlands, 2017, pp. 1–32.

[16] N. Vahrenkamp, M. Wächter, M. Kröhnert, K. Welke, and T. Asfour, "The robot software framework armarx," *it-Information Technology*, vol. 57, no. 2, pp. 99–111, 2015.

[17] T. Asfour, L. Kaul, M. Wächter, S. Ottenhaus, P. Weiner, S. Rader, R. Grimm, Y. Zhou, M. Grotz, F. Paus, D. Shingarey, and H. Haubert, "ARMAR-6: A collaborative humanoid robot for industrial environments," in *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, 2018, pp. 447–454.

[18] S. Rader, L. Kaul, P. Weiner, and T. Asfour, "Highly Integrated Sensor-Actuator-Controller Units for Modular Robot Design," in *IEEE International Conference on Advanced Intelligent Mechatronics (AIM)*, Munich, Germany, 2017, pp. 1160–1166.

[19] T. Asfour, *Sensomotorische Bewegungskoordination zur Handlungsausführung eines humanoiden Roboters*. Ph.D. thesis, Department of Informatics, University of Karlsruhe, 2003.

[20] S. Rader, L. Kaul, H. Fischbach, N. Vahrenkamp, and T. Asfour, "Design of a High-Performance Humanoid Dual Arm System with Inner Shoulder Joints," in *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, Cancún, Mexico, 2016, pp. 523–529.

[21] N. Fukaya, S. Toyama, T. Asfour, and R. Dillmann, "Design of the TUAT/Karlsruhe Humanoid Hand." in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Takamatsu, Japan, October 30 - November 5 2000, pp. 1754–1759.

[22] T. Yoshikawa, "Manipulability of Robotic Mechanisms," *The International Journal of Robotics Research*, vol. 4, no. 2, pp. 3–9, 1985.

[23] T. Asfour, J. Schill, H. Peters, C. Klas, J. Bücker, C. Sander, S. Schulz, A. Kargov, T. Werner, and V. Bartenbach, "ARMAR-4: A 63 DOF Torque Controlled Humanoid Robot," in *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, Atlanta, USA, October 2013, pp. 390–396.

[24] N. Vahrenkamp, M. Wächter, M. Kröhnert, K. Welke, and T. Asfour, "The Robot Software Framework ArmarX," *Information Technology*, vol. 57, no. 2, pp. 99–111, 2015.

[25] D. Shingarey, L. Kaul, and T. Asfour, "Torque-based velocity control for safe human-humanoid interaction," in *2019 28th International Conference on Robotics in Alpe-Adria-Danube Region (RAAD)*, 2019, pp. 61–68.

[26] Y. Zhou, M. Do, and T. Asfour, "Coordinate change dynamic movement primitives - a leader-follower approach," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Daejeon, Korea, October 2016, p. 5481–5488.

[27] J. Gao, Y. Zhou, and T. Asfour, "Projected force-admittance control for compliant bimanual tasks," in *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, Beijing, China, November 2018, pp. 607–613.

[28] N. Vahrenkamp, E. Koch, M. Wächter, and T. Asfour, "Planning High-Quality Grasps using Mean Curvature Object Skeletons," *IEEE Robotics and Automation Letters (RA-L)*, vol. 3, no. 2, pp. 911–918, 2018.

[29] P. Schmidt, N. Vahrenkamp, M. Wächter, and T. Asfour, "Grasping of Unknown Objects using Deep Convolutional Neural Networks based on Depth Images," in *IEEE International Conference on Robotics and Automation (ICRA)*, Brisbane, Australia, May 2018, pp. 6831–6838.

[30] N. Vahrenkamp, L. Westkamp, N. Yamanobe, E. E. Aksoy, and T. Asfour, "Part-based Grasp Planning for Familiar Objects," in *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, Cancun, Mexico, Nov 2016, pp. 919–925.

[31] S. Ottenhaus, D. Renninghoff, R. Grimm, F. Ferreira, and T. Asfour, "Visuo-haptic grasping of unknown objects based on gaussian process implicit surfaces and deep learning," in *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, Toronto, Canada, 2019.

[32] Z. Cao, G. Hidalgo, T. Simon, S. Wei, and Y. Sheikh, "OpenPose: Realtime multi-person 2D pose estimation using part affinity fields," *CoRR*, vol. abs/1812.08008, 2018.

[33] D. Tome, C. Russell, and L. Agapito, "Lifting from the deep: Convolutional 3d pose estimation from a single image," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 2500–2509.

[34] J. R. Medina, F. Duvallet, M. Karnam, and A. Billard, "A human-inspired controller for fluid human-robot handovers," in *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*. IEEE, 2016, pp. 324–331.

[35] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," in *Proceedings of the Third International Conference on Learning Representations (ICLR)*, 2015.

[36] S. Constantin, J. Niehues, and A. Waibel, "Multi-task learning to improve natural language understanding," in *Proceedings of the Tenth International Workshop on Spoken Dialogue Systems (IWSDS)*, 2019.

[37] T. S. Nguyen, M. Sperber, S. Stüker, and A. Waibel, "Building realtime speech recognition without CMVN," in *International Conference on Speech and Computer*. Springer, 2018, pp. 451–460.

[38] B. Busch, G. Cotugno, M. Khoramshahi, G. Skaltsas, D. Turchi, L. Urbano, M. Wächter, Y. Zhou, T. Asfour, G. Deacon, D. Russell, and A. Billard, "Evaluation of an industrial robotic assistant in an ecological environment," in *IEEE International Conference on Robot and Human Interactive Communication*, 2019.