Towards the Robotic Co-Worker

Sami Haddadin, Michael Suppa, Stefan Fuchs, Tim Bodenmüller,
Alin Albu-Schäffer, and Gerd Hirzinger

Abstract. Recently, robots have gained capabilities in both sensing and actuation, which enable operation in the proximity of humans. Even direct physical interaction has become possible without suffering the decrease in speed and payload. The DLR Lightweight Robot III (LWR-III), whose technology is currently being transferred to the robot manufacturer KUKA Roboter GmbH, is such a device capable of realizing various features crucial for direct interaction with humans. Impedance control and collision detection with adequate reaction are key components for enabling "soft and safe" robotics. The implementation of a sensor based robotic co-worker that brings robots closer to humans in industrial settings and achieve close cooperation is an important goal in robotics. Despite being a common vision in robotics it has not become reality yet, as there are various open questions still to be answered. In this paper a sound concept and a prototype implementation of a co-worker scenario are developed in order to demonstrate that state-of-the-art technology is now mature enough to reach this aspiring aim. We support our ideas by addressing the industrially relevant bin-picking problem with the LWR-III, which is equipped with a Time-of-Flight camera for object recognition and the DLR 3D-Modeller for generating accurate environment models. The paper describes the sophisticated control schemes of the robot in combination with robust computer vision algorithms, which lead to a reliable solution for the addressed problem. Strategies are devised for safe interaction with the human during task execution, state depending robot behavior, and the appropriate mechanisms, to realize robustness in partially unstructured environments.

Sami Haddadin · Michael Suppa · Stefan Fuchs · Tim Bodenmüller ·
Alin Albu-Schäffer · Gerd Hirzinger
Institute of Robotics and Mechatronics
DLR e.V. - German Aerospace Center
P.O. Box 1116, D-82230 Wessling, Germany
e-mail: `sami.haddadin@dlr.de`

1   Motivation and Introduction

The idea of human and robot working together is of major interest for both
the academic community and industrial robot manufacturers. Pioneering ex-
amples of intimate collaboration between human and robot, which origin can
be found in [1], are Intelligent Assist Devices (IADs), as the skill assist de-
scribed in [2]. In 1983 a method was proposed at DLR for allowing immediate
"programming by touch" of a robot through a force-torque-sensor-ball [3], see
Fig. 1 (left).

In this paper, we propose an approach to effectively combine human and
robot capabilities for performing tasks in a partially unknown workcell, i.e.
a semi-structured environment. We elaborate the theoretical basis, prerequi-
sites regarding task execution and safe interaction mainly relying on sensor
based reaction strategies. The concept requires flexibility from the robot in
terms of sensor integration and programming. This flexibility is currently not
available in the state-of-the-art first generation industrial robots, designed
mainly to position objects or tools in six degrees of freedom (DoF).

However, for the second generation industrial robot, a fundamental
paradigm shift is required to enable the implementation of the robotic co-
worker. This concept is derived from a meaningful fusion of robots with in-
novative and robust control concepts, so called "soft robotics" features, and
exteroceptive sensing as e.g. 3D vision sensing modalities for safely perceiving
the environment of the robot. Together with additional sensing capabilities
for surveillance such technology will open entirely new application fields and
manufacturing approaches. In order to develop and evaluate the proposed
concept, the DLR Co-Worker was constructed as a demonstration platform,
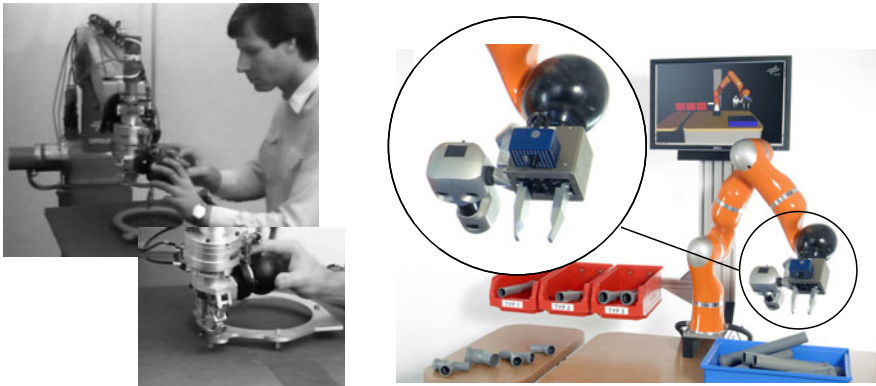see Fig. 1 (right).

Fig. 1 The concept of sensor programming was developed at DLR in 1983 for
teaching robot paths and forces/torques simultaneously (left). The DLR Co-Worker
consisting of the DLR Lightweight Robot III, the DLR 3D-Modeller (DLR-3DMo),
and a Time-of-Flight Camera (ToF-camera) (right).

Complementary sensor fusion[1] plays a key role in achieving our desired performance by the combination of complementary input information. As demonstrated in [4] a prioritized and sequential use of vision and force sensor based control leads to robust, fast, and efficient task completion using the appropriate sensor information depending on the particular situation. Currently, we believe that parallel use of both is needed mainly for very specific problems, which are usually irrelevant for industrial settings.

Presently, industrial robot applications require complete knowledge of the process and environment. This approach is prone to errors due to model inaccuracies. Our central approach is to use intelligent sensor-based reaction strategies to overcome the weaknesses of purely model-based techniques. Thus, we can deal with sensor noise and limited robot positioning accuracy. The robot task is described in high-level functions encapsulated in the states of hybrid automata, where transitions base on decisions made using sensor inputs. This enables the robot to react to "unexpected" events not foreseen by the programmer. These events are induced by the human behavior, which cannot be completely modeled analytically. Apart from robust behavior, safety is of fundamental concern [5, 6, 7] if human-robot cooperation shall ever be realized beyond the prototype phase. In terms of mechanical design it is not effective to attempt to use large robots and try to make them sufficiently safe [8, 9]. We showed in recent work that in physical human-robot interaction (pHRI) slight collisions with the robot are not fatal, if robots with suitable mechanical design are used and the proper sensor-based reaction strategies are implemented [10]. Furthermore, the human is encouraged in our setup to physically interact with the robot as a modality to "communicate" with it and provide task-relevant information. This also improves the fault tolerance level of the task since only absolutely worst-case contacts are solved by a complete emergency stop in contrast to approaches for current robots.

Apart from the described approach, the presented concept for the robotic co-worker is fundamentally different from classical industrial ones. None of the components are supposed to be intrinsically fail safe, but the appropriate combination of all components makes the system more safe, robust, and reliable. We use multiple sensor information of the robot and external sensing for increasing the error tolerance and fault recovery rate. The work we present is an attempt to merge our results for safe Human-Robot Interaction with robust exteroceptive and external sensing to achieve the robotic co-worker. We will discuss how to extend our schemes and the available solutions for particular problems and finally reach the stage of a highly flexible state-based programming concept for various applications. This task description allows for novel switching strategies between control modes, sensory reaction strategies, and error handling.

In this overview paper we discuss mostly the general concept. For further readings and details on the methodologies the interested reader is referred to the cited literature.

---

[1] Please note the difference of complementary from competitive sensor fusion.

The remainder of this paper is organized as follows. First, the general functional modes required for a robotic co-worker are described. Then, the interaction concept is outlined in detail. Furthermore, the task description performed autonomously by the robot is elaborated. Finally, the developed concepts are applied to a robotic bin picking scenario with user interaction as a case study in order to present their practical relevance and implementation.

2   Functional Modes

Currently, industrial settings incorporate, in most cases, simple sequences of tasks whose execution orders are static, allowing sometimes some binary branching. Fault tolerance during task execution is, apart from certain counterexamples[2], usually not an issue due to the well designed environment. Furthermore, human-robot interaction is not yet safely and effectively implemented and the legal foundation for it is, to a large extent, non existent at the current stage. In industrial settings a fault immediately leads to a complete stop of the manufacturing process, i.e. robust behavior in an unstructured environment has not been addressed until now. We propose an integrative and flexible approach to carry out the desired task in a very robust yet efficient way. At the same time, this approach is able to distinguish between different fault stages, which stop the entire process and lower the efficiency only in the worst case. Flexible jumps within execution steps are part of the concept and do not require special treatment. In order to optimally combine human and robot capabilities, the robot must be able to quickly adapt to the human intention during task execution for both safe interaction and high productivity. Thus, the measured human state is the dominant transition between the proposed functional modes. Estimating the human state is a broad topic of research and has been addressed in recent work [11]. The focus is often on estimating the affective state of humans, which is of secondary interest during an industrial process. The more relevant information is the physical state that the human currently occupies, and the estimation of the human attention, so that a clear set of sufficient behaviors can be selected and activated, which leads to robust and reliable overall performance. This paper does not consider attention estimation, instead focusing on the human state.

We compiled the following selection of physical states to provide sufficient coverage for cases relevant to our study, also shown in Fig. 2 (left).

- **oP:** out of perception
- **iP:** in perception
- **iCM:** in collaborative mode
- **iHF:** in human-friendly zone

oP denotes that the human is out of the perceptional ranges of the robot and therefore not part of the running application. **iP** indicates that the human

---

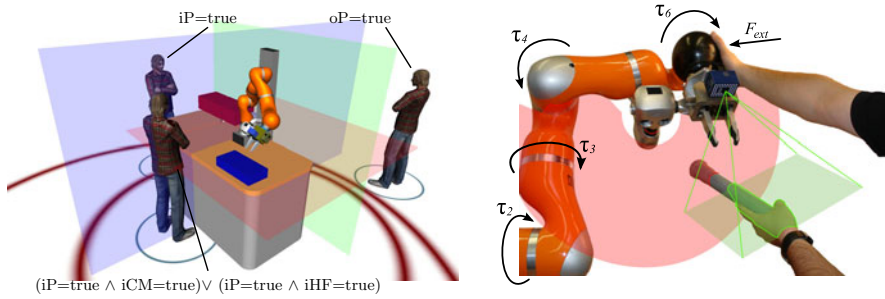[2] Checking for a successful grasp is e.g. commonly used.

Fig. 2 Proximity and task partition (left) and modalities for multi sensor human robot interaction in the DLR Co-Worker (right).

is in the measurement range of the robot, and thus its presence has to be part of the robot control. **iCM** and **iHF** indicate whether a collaborative or human-friendly behavior must be ensured. Each physical state is subdivided, depending on the task. However, only when **iCM** $= true$, the collaborative intention should be taken into account: This leads to a complex physical interaction task. In this paper we will use the "hand-over and receive" process as an example, see Fig. 2 (right).
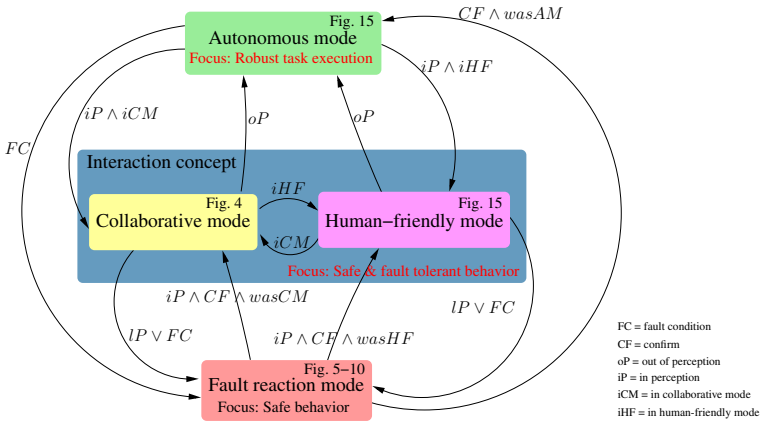

Fig. 3 Functional modes for the DLR Co-Worker.

The human state is primarily used to switch between different functional modes of the robot which in turn are associated with fault behavior. As shown in Fig. 3 we distinguish between four major functional modes of the robot in a co-worker scenario:

1. **Autonomous task execution:** autonomous mode in human absence
2. **Human-friendly behavior:** autonomous mode in human presence

3. **Co-Worker behavior:** cooperation with human in the loop
4. **Fault reaction behavior:** safe fault behavior with and without human in the loop

Their interrelation and transition conditions provide high flexibility in the application design. In the **first** functional mode the robot is autonomously fulfilling its given task without considering the human presence. The task is carried out under certain optimality criteria, such as cycle time, in order to maximize the productivity. In the **second** and **third** modes, we need a meaningful partition of the task space which subdivides the given workspace of the robot into regions of interaction. These incorporate the "hand over" schemes as described in Sec. 3.2 and human-friendly behavior, whose core elements are reactive collision avoidance and self-collision avoidance schemes. In the **third** mode interaction tasks are carried out that have to be specified or generated for fulfilling a common desired goal, involving a synergy of human and robot capabilities in an efficient manner. These two modes form an integrative interaction concept, allowing seamless switching between each other. The **fourth** mode defines the fault reaction behavior, addressing the appropriate and safe state dependent fault reaction of the robot. It incorporates both the robustness concepts during autonomous reaction, as well as human-safe behavior. Since each mode possesses an underlying safety concept, it will be described later in more detail.

3 Interaction Concept

In this section we describe the developed interaction schemes. First, the proposed task space partition is outlined, followed by the interaction layer, different collision avoidance techniques, as well as physical collision detection and reaction for safe pHRI. Finally, the resulting safety architecture, which unifies the different schemes, is presented.

3.1 *Proximity and Task Partition*

In case humans are in close proximity to robots in current industrial installations, the robots reside inside safety cages in order to prevent any physical contact and thus minimize the risk for humans. However, when humans and robots shall collaborate, such a plant design is no longer an option. The human location has to be taken into account in the control scheme and in higher level control of the robot as an integral part of the system design. The previously introduced physical human states have to be mapped into a meaningful topology shown in Fig. 2 (left), where the four distinct classes are indicated. They should be established with respect to the task and the robot workspace for assessing, whether the human does not have to be taken into account and therefore, the robot still behaves autonomously regardless of the iP state. In

case the human does not enter the robot workspace, it is not necessary to degrade the productivity of the robot. In this sense the functional mode of the robot changes only, if the human clearly enters the workspace of the robot (indicated by the inner circle). If the human has entered the robot workspace a distinction between human friendly behavior (on the right side of the table in Fig. 2 (left)) and the cooperative mode (and their respective submodes) is required (on the left side of the table in Fig. 2 (left)). If perception is lost while $iP = true$, the robot assumes a severe error condition, stops and waits for further instructions. If the presence of the human was not detected at all, i.e. a worst case from a safety point of view, various safe control schemes ensure the safety of the human during possible unforeseen collisions.

Defining these regions is part of the application design and definition phase. Furthermore, we introduce switching zones, which are boundary volumina of pre-defined thickness between task partitions (see Sec. 3.3 for details).

3.2   Interaction Layer

Interaction between robot and human is a delicate task, which needs multi-sensor information. Furthermore, robust as well as safe control schemes are called for to enable intuitive behavior. The main physical collaboration schemes are "joint manipulation" and "hand over and receive". "Parallel execution" may be part of a task, but usually without physical interaction. Some work has been carried out on exchanging objects between human and robot based on reaching gestures [12]. In [13] the concept of interaction history was used to achieve cooperative assembly.

Figure 2 (right) shows the "hand-over" and "receive" for the DLR Co-Worker Central entity is the LWR-III with its soft robotics features. As a default we utilize its high-performance impedance control, and only switch to other schemes, such as position control, if necessary. The robot is equipped with joint torque sensors in every joint. It is well suited for realizing various important features such as load loss detection and online load identification without additional force sensing in the wrist. Collision detection and reaction, depending on the potential physical severity of the impact and current state, is a central feature used for detection and isolating contacts of different intensity along the entire robot structure. By being able to distinguish different contact types, fault tolerant and situation suited behavior is possible.

We utilize virtual walls for avoiding collisions with the environment through control schemes. In order to realize an effective reactive behavior, it is important to change stiffness, velocity, disturbance residuals (see Sec. 3.3), trajectory generators, collision severity reaction strategies, and robot control parameters on the fly within the lower level control rates (here 1 ms), holding also during motion or state execution. With the combination of exteroceptive sensing, capabilities of object recognition, tool surveillance, and human proximity detection (shown in Fig. 2 (right)), we can achieve such aforementioned
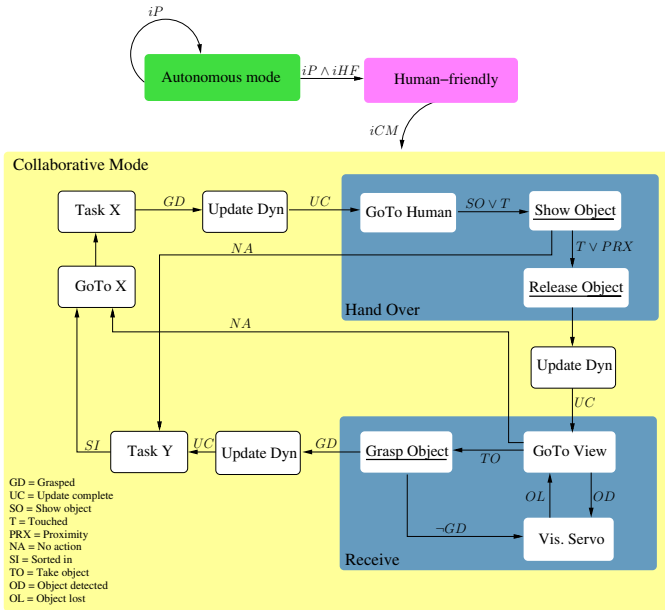
Fig. 4 Example for "hand-over" and "receive". Underlined states incorporate explicit physical interaction.

complex processes as "hand-over" and "receive", shown in Fig. 4. "Receiving" or "handing over" the object is simply triggered by touching the robot at any location along its entire structure or by using the proximity information from the mounted exteroceptive sensors.

3.3 Absolute Task Preserving Reaction: Time Scaling

While the robot is in human-friendly mode, its intention is to fulfill the desired task time efficiently, despite human presence. In order to accomplish this, it is necessary to equip the robot the robot with reactive motion generators that take into account the human proximity and thus prevent inefficient task abortion.

Trajectory scaling preserves the original motion path and at the same time provides compliant behavior by influencing the time generator of the desired trajectory, see [10]. This scheme can even be used to enable a position controlled robot to react compliantly in such a way that it remains on the nominal path, albeit with limited maximum forces in case of physical external disturbances.

A desired trajectory is usually parameterized with respect to time. If the discrete sampling time Δt is modified in such a way that it is used to respond to such external forces, it can be used to step back and forth along the
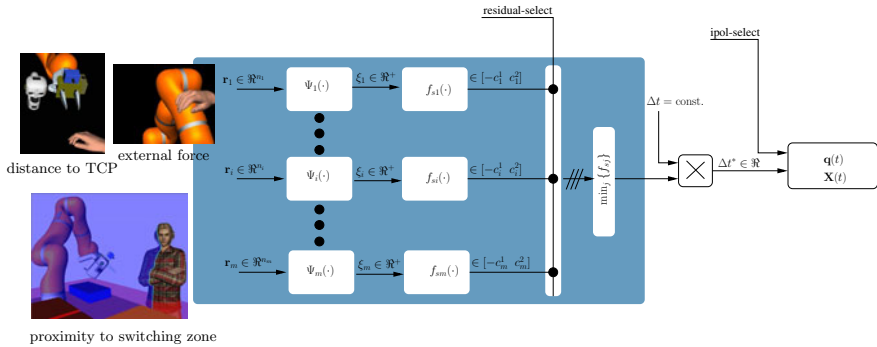
Fig. 5 Residual fusion for integrated trajectory scaling. Ψ_i is a normalization function and f_{si} a sigmoid function for time scaling, [10, 14].

desired path, by "scaling the trajectory in time". In our approach we use physical contact residuals such as the estimated external joint torque, or the external contact wrench, together with proximity based residual signals such as the human-robot proximity, the human-switching zones proximity, and the human-workspace proximity, see Fig. 5. The usefulness of the approach becomes also apparent when considering cases where humans are moving close to switching zones. If the robot would simply use binary switching information about the current state of the human, undesired oscillating behavior would occur due to the imprecise motions and decisions of the human. By using the human proximity to this border as a residual the robot always slows down and stops until the human clearly decides his next action. This way, the user receives intuitive visual feedback, indicating that the robot is aware of his presence and waits for further action.

The fusion of the different residuals is shown in Fig. 5 for several aforementioned signals. This concept allows us to bring quantities of different physical interpretation together and use them in a unified way for trajectory scaling. Each residual is normalized[3] and then nonlinearly shaped to be an intuitive time scale. Depending on the current state, the user can choose suitable residuals accordingly during application design.

3.4 Task Relaxing Reaction: Reactive Path Deformation

Apart form task preserving reaction as described in the previous subsection, reactive real-time reaction with task relaxation is an important element for dealing with dynamic environments as well. A well known technique in this respect is the elastic strips framework [15].

---

[3] Please note that we refer to an appropriate handling as e.g. projecting external forces to the velocity direction of the robot or similar transformations.
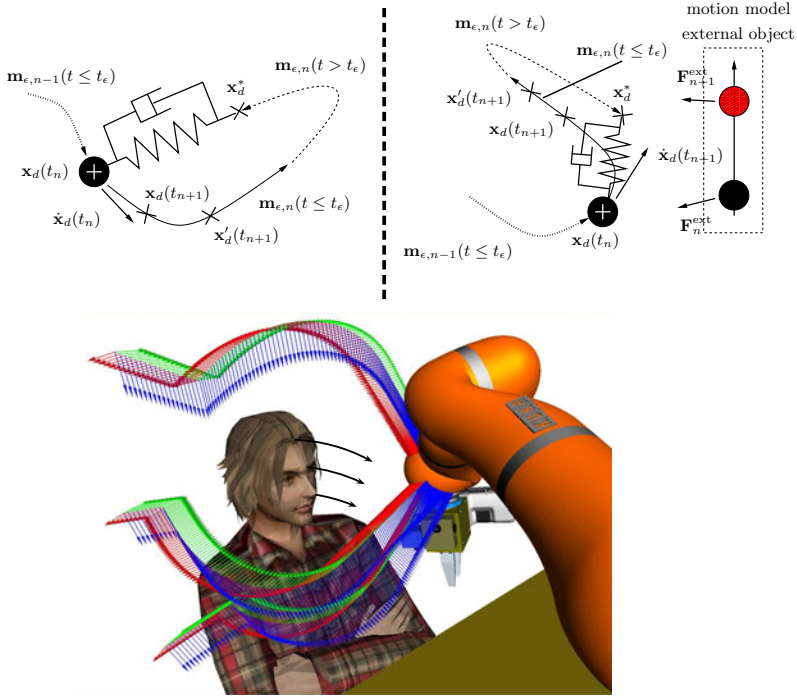
Fig. 6 Schematic views of the collision avoidance for two consecutive iteration steps. The upper left figure denotes free motion, whereas the upper right one takes into account a motion model of an external virtual object. The lower figure shows the collision avoidance for a full dynamic simulation.

The subsequent method uses a decoupled second order impedance system (mass-spring-damper) as a starting point and describes the (dynamic) environment by means of virtual forces, which are generated with a motion model of the according object, see Fig. 6 (upper). The virtual mass associated with the robot is placed at the starting configuration $\mathbf{x}_d(t_n), \dot{\mathbf{x}}_d(t_n)$ of the robot, whereas the equilibrium of the system is the desired goal configuration $\mathbf{x}_d^*$. The resulting trajectory of the excited system leads to (assuming no local minima) a valid trajectory towards the goal. However, such a simple solution by means of velocity and acceleration of the trajectory leads for most cases to very undesired properties of the generated path. In order to overcome this deficit we calculate the traverse path of the system $\mathbf{m}_{\epsilon,n}(t \leq t_\epsilon)$ every time step (incorporating the dynamic behavior of the environment) within a certain reasonable time interval t_ϵ, but dismiss the time information associated with it. In order to match a desired velocity $\dot{\mathbf{x}}_d'(t_{n+1})$, we search for the configuration $\mathbf{x}_d'(t_{n+1})$ along the path that ensures this velocity. Thus, we keep the smooth properties of the generated local path but the velocity of the robot can be commanded independently. During (virtual) contact the velocity can be additionally scaled similarly to the previously described trajectory scaling method. Thus, due to

the collision avoidance, the robot would continously reduce speed, or even retract, and at the same time actively avoid the upcoming collision. Figure 6 (lower) shows the result for different starting points and the common goal configuration of the robot. The nominal trajectory is a straight line from different starting points to a common end point. The avoidance takes place for a dynamic motion of the human towards the robot.

3.5 Dealing with Physical Collisions

In our recent work we concentrated on evaluating the injury severity for the human during worst-case robot-human impacts [8]. Furthermore, we developed various control schemes to provide a high level of safety to the human during desired physical interaction and *unexpected collisions*. Numerous solutions against human injury are available. Crucial features for them are an effective physical collision detection and reaction. Furthermore, after a collision is detected and isolated, an appropriate reaction has to be triggered [10]. One possible solution is to stop the robot as soon as a collision is detected. Another one is to switch from position control to zero-gravity torque control [16] and let the robot react in a convenient compliant manner. These approaches provide the possibility to divide the impact severity into several stages, using a disturbance observer. This method for detecting contacts is also able to give an accurate estimation of the external joint torques τ_{ext}, which in turn can be used to classify collisions with the environment according to their "severity level". This allows us to react variably to different *collision severity stages* [14], leading to a *collision severity based behavior*. Apart from this nominal contact detection, our algorithms are also able to detect malfunction of the joint torque sensors, based on model inconsistencies interpreted as a collision.

3.6 Safety Architecture

Apart from gaining insight into mechanisms behind safe pHRI and isolated tools, it is critical to determine how to apply the knowledge and methodologies in a consistent and appropriate manner. We have developed schemes to utilize these features appropriately in order to maximize task performance under the constraint of achieving sufficient human-friendly behavior, see Fig. 7. Each feature is shown at the according hierarchical level where it is introduced and made available in the appropriate layer of the process.

Figure 8 outlines how the fault management and emergency components are embedded as underlying components for each task. Every task has the appropriate low-severity-fault tolerant components to make it robust against external disturbances in general and prevent unnecessary task abortion. Each of them activates their distinct safety set S_j which is compatible with the particular goal (see Fig. 10 for details).
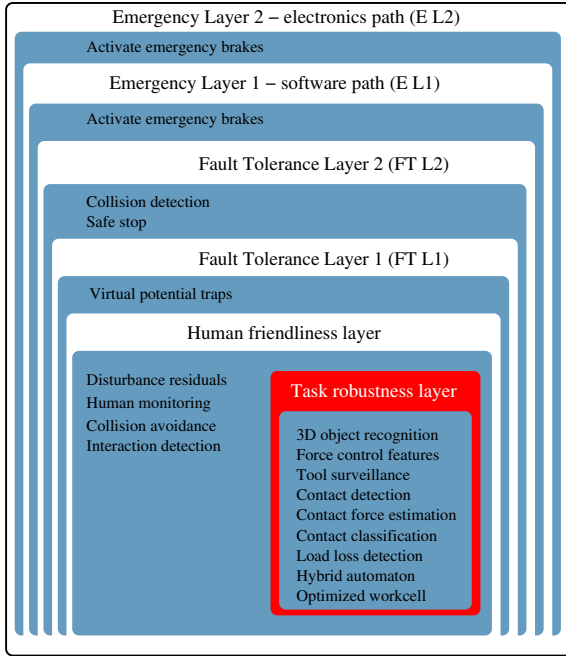
Fig. 7 Safety architecture of the DLR Co-Worker. Only the first two stages are user specific.
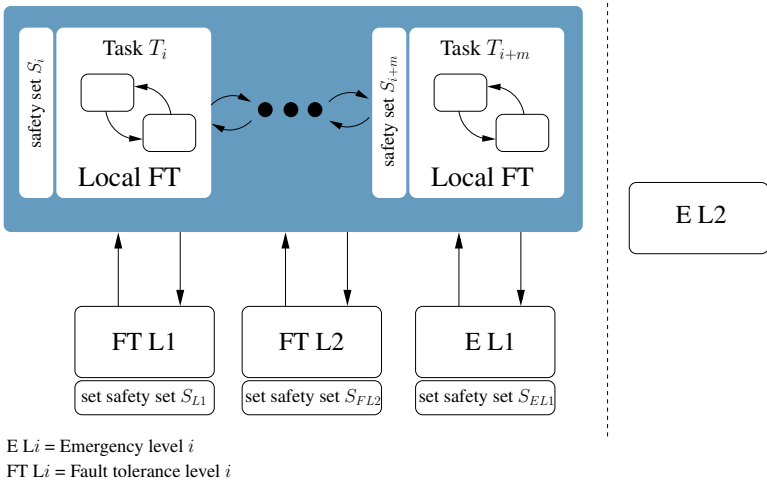

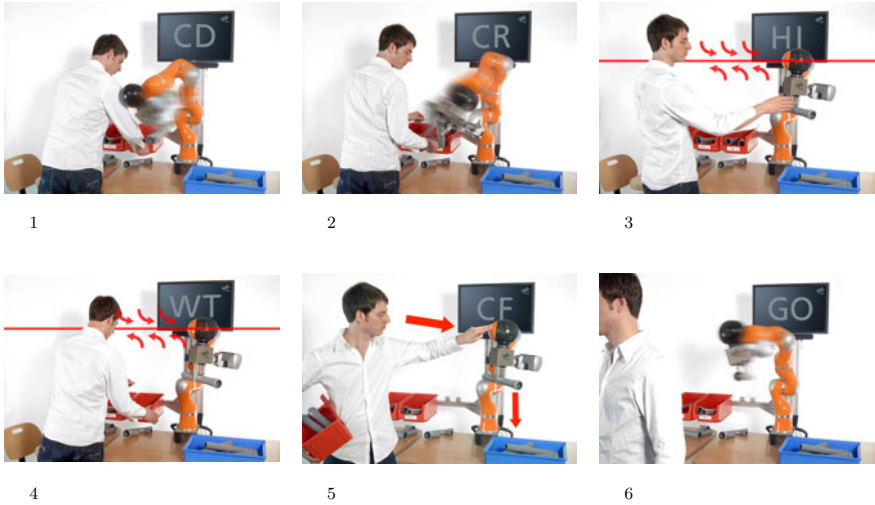**Fig. 8** Safety background of the DLR Co-Worker.

Fig. 9 Safe physical human-robot interaction. Detecting and recovering from a collision in FT L1. It was assumed that the human was not perceived to have entered the workspace.

Figure 9 shows an example of an *unexpected collision* between a worker and a human ①, leading to a collision in layer FT L1. The robot switches to a compliant behavior ② after the collision is detected (CD). Due to the collision reaction (CR), the robot is able to be freely moved in space. This could lead to secondary collisions with the environment. Therefore, we have designed nonlinear virtual walls (Fig. 2) with rigid properties to prevent physical collisions of the robot and secure the sensitive parts as the ToF-camera and the 3DMo. As a result, the human can simply grab the robot anywhere along the structure and hang it like a tool into a predefined arbitrarily shaped virtual potential trap[4] (HI) ③, which smoothly drags it in and keeps it trapped. The human can then complete his task, which he intends to fulfill ④, while the robot waits (WT) for further action. After completion is confirmed (CF) in ⑤ the robot continues ⑥ with the interrupted task (GO). If this was not the case, the robot stays in his constrained passive behavior until either a confirmation for continuation occurs, or a human would drag him out of the hang-in field, depending on a predefined direction of disturbance. Figure 10 shows how such behavior is triggered in a hybrid automaton and the safety sets involved in this process.

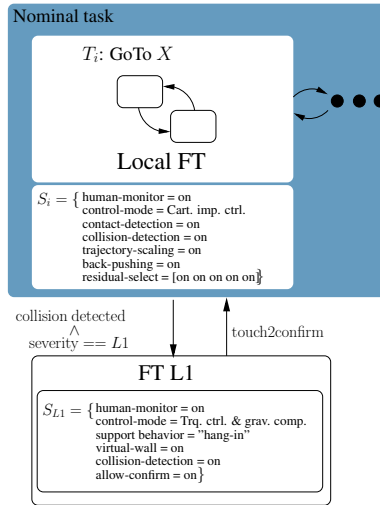[4] This feature "feels" like a magnet.

Fig. 10 Safe reaction to a collision in FT L1 under the assumption that the human was not perceived to have entered the workspace. A simple and convenient behavior is triggered, which can be realized by intuitive use of well designed state dependent control scheme selection.

4   A Sensor Based Approach for Interactive Bin Picking

In this section we focus on describing our solutions to solve an industrially relevant autonomous task by combining computer vision techniques with soft-robotics features and embed it into an interaction scenario with the human. To demonstrate the performance of our system during autonomous task execution, we address the classical bin picking problem, which is well known since the mid-1980s. However, such problems have remained difficult to be solved effectively. This sentiment can be found in different literature, as exemplified below:

> "Even though an abundance of approaches has been presented a cost-effective standard solution has not been established yet."

> *Handbook of Robotics 2008* [17]

We have combined environmental modeling, robust and fast object recognition, as well as quick and robust grasping strategies in order to solve the given task. The setup depicted in Fig. 1 (right) serves as our demonstration platform. It is further used for realizing a scenario where the human assembles parts, which are supplied by the robot and, after a "hand over" and "receive" cycle, sorted into a depot by the robot, see Fig. 16. This fully sensor-based concept is entirely embedded in the proposed safe interaction concepts. The intention of this application is to augment human capabilities

with the assistance of the robot and achieve seamless cooperation between each other.

4.1 Vision Concept

The LWR-III is equipped with two exteroceptive sensors: the DLR 3D-Modeller and a time-of-flight camera so that different proximity sensors with complementary features can be used within this scenario.

DLR 3D-Modeller

Fig. 11 Generated 3D model from a series of sweep scans over the filled bin.

System: The DLR-3DMo is a multi-purpose vision platform [18], which is equipped with two digital cameras, a miniaturized rotating laser scanner and two line laser modules, see Fig. 1 (right). The DLR-3DMo implements three range sensing techniques:

1. laser-range scanning [19]
2. laser-stripe profiling [20]
3. stereo vision

These techniques are applicable to a number of vision tasks, such as the generation of photo realistic 3D models, object tracking, collision detection, and autonomous exploration [21].

Implementation: The laser-range scanner, used for securely determining obstacles and free regions, provides range data enriched with a confidence value. The proposed application employs the rotating laser range scanner for two tasks. First, the wide scan angle of 270 degrees enables nearly complete

surveillance of the working range around the gripper. Secondly, the measured distance data provides information about occupation of the space between the jaws of the gripper and indicates whether a target object is located there.

The laser-stripe profiler is used for modeling the environment and can be used for the localization of the bin or accurate modeling of the entire workcell, see Fig. 11. The shown model was generated with a series of sweep motions of the LWR-III across the scenario. The main purpose of the laser-stripe profiler is to acquire accurate data for model generation, in contrast to the safety functionality of the laser-range scanner.
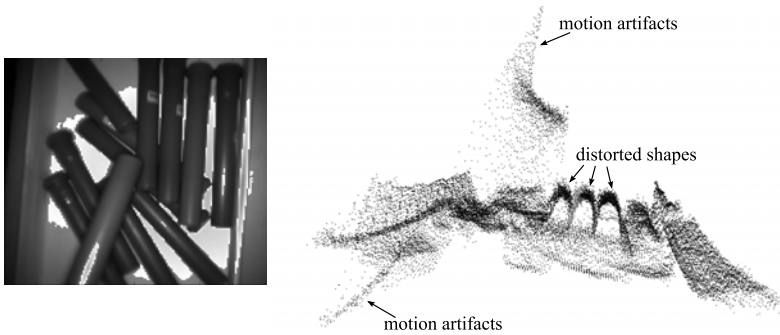
Time-of-Flight camera

Fig. 12 Amplitude and depth data from view into the bin (left) showing large signal noise (right).

System: The ToF camera Swissranger SR 3000, mounted on the robot, has a resolution of 176×144 pixels. An important feature of this device, beneficial for this application, is the ability to capture $2\frac{1}{2}$D depth images at ≈ 25 Hz. Unlike stereo sensors, ToF-cameras can measure untextured surfaces because the measurement principle does not depend on corresponding features. Furthermore, due to the active illumination, ToF-cameras are robust against ambient illuminations changes. These properties enable the recently established use in the robotics domain for tracking, object detection, pose estimation, and collision avoidance. Nonetheless, the performance of distance measurements with ToF-cameras is still limited by a number of systematic and non-systematic error sources, which turn out to be a challenge for further processing.

Figure 12 highlights the non-systematic errors such as noise, artifacts from moving objects, and distorted shapes due to multiple reflections. While noise can be handled by appropriate filtering, the other errors mentioned here are system inherent. The systematic distance-related error can be corrected by a calibration step down to 3 mm, see [22].

Implementation: Generally, the high sampling rate of the ToF-camera guarantees fast object localization and robust object tracking performance based
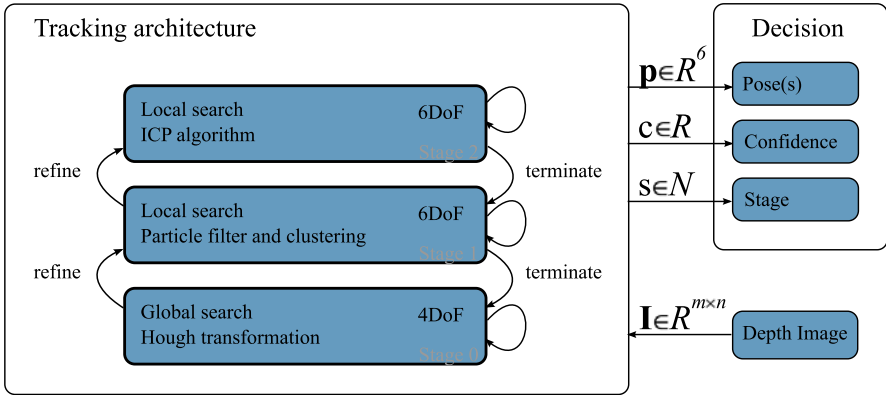
Fig. 13 Multi-stage tracking architecture based on [23].

on a three staged tracking architecture, see Figure 13. In each stage a different algorithm processes an incoming depth image to provide a list of pose hypotheses for the potential object, which is additionally tagged with a confidence value. The stages are continously monitored and executed according to suitable termination criteria or reentered for refinement.

The first stage is a global search, consisting of edge filtering and a Hough transformation for identifying lines as initial hypotheses for the tube location. In the second stage these hypotheses are locally consolidated and clustered by a particle filter. Third, an Iterative Closest Point algorithm (ICP) provides an accurate pose estimation of the target object at a frame rate of ≈ 25 Hz. Both ICP, and particle filter directly process 3D data, and a 3D model of the target. The 3D model is represented by a point set with corresponding normals. This can be either generated from CAD models or surface reconstruction. The object target can be localized and tracked with an accuracy of ≈ 7 mm.

4.2 Soft Robotics Control for Grasping

The soft-robotics features of the LWR-III greatly provide powerful tools to realize such a complex task as bin picking. Cartesian impedance control [16] is used as a key element for robust grasping despite the aforementioned recognition uncertainties. The impedance behavior of the robot is adjusted according to the current situation in order to achieve maximal robustness. Furthermore, the previously introduced strategies for fault detection are used to recognize impossible grasps or unexpected collisions with the environment based on force estimation. Furthermore, there are virtual walls preventing collisions with the static environment. The robustness of grasping against errors in object localization and errors in positioning due to the used impedance control is of great importance for this application. The grasping strategy shown in
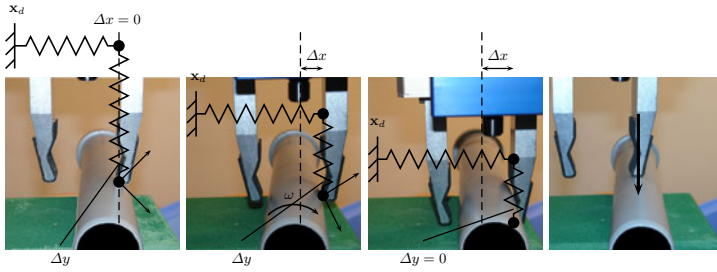
Fig. 14 Compliant grasping strategy.

Fig. 14 successfully copes with possible translational deviations in the range of 55 mm before the grasp fails. Due to the compliant behavior of the robot and gripper-object and object-ground friction, the object is rotated into the firm grasp. The last image shows a case expected to be a failure. However, due to the rotational stiffness we implemented along the axis perpendicular to the image plane grasping can still be achieved.

4.3 Autonomous Task Execution

Figure 15 depicts the autonomous bin picking task automaton, which merges the presented concepts into a high-level view of task description. The application is comprised of object recognition, grasping, and sort-in phases[5]. If the bin is depleted, the robot waits for further supply. Fault tolerant behavior is realized by introducing various branching possibilities for each state
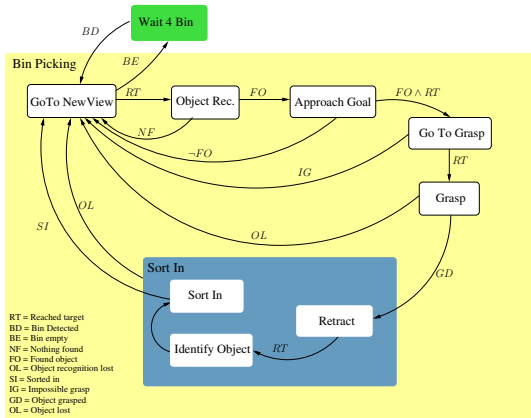

Fig. 15 Automaton for autonomous bin picking.

---

[5] The initial view and sort-in frames are taught in torque control with gravity compensation. This enables the user to freely move the robot to a desired configuration and save it in the application session.

execution. In case of failure, the robot recovers by monitoring conditions like object recognition dropouts, load losses, or impossibility of grasps.

4.4 Evaluation of Grasping Success

The efficiency and robustness of our approach was tested in a series of autonomous grasps. For this evaluation we replenished the bin (Fig. 11) after each successful grasp in order to have a filled bin and independent trials. On average, the robot needed 6.4 s for one grasping process, which comprises object detection from an arbitrary viewing position, approaching and grasping, unbagging, and moving back to the initial viewing position. The robot was able to grasp an object in every cycle for 80 trials, i.e. the overall cycle success rate was 100%. This result was only achievable due to the fault tolerance capabilities of the system along the entire process, such as the detection of a physical impossibility of a planned grasp, the non-successful grasp (overall 3 times), loosing an object in tracking, or localization without any result. The last fault was mainly caused by the fact that searched objects are often only partially in the field of view, so that the robot had to move to a new view position. All of these failure modes where detected or realized by the system and induced a restart of the grasping process. Consequently, the number of average views to recognize an object was $N_{\text{view}} = 2.2$.

4.5 Extension to Interactive Bin-Picking

Figure 16 describes our implementation of an interactive bin-picking demonstrator, merging the concepts for interaction and the autonomous capabilities of the robot. The initial entrance into the scene by the human is not shown, but is part of the demonstrator, i.e. it is assumed that the human has entered the scene, the "way into interaction" is completed, and the human is part of the process. ① shows the view into the bin and the corresponding object recognition (OR). Then, the robot grasps an object out of the bin ② and identifies it according to its weight, followed by a motion towards the human (GH) in ③. The "hand over" ④ then takes place, after which the robot waits (WT) for the human to complete his process ⑤. As soon as the human has finished, the robot receives the object in a visual servoing loop (VS) in ⑥. Now, the classified object is sorted into (SI) one of the trays ⑦ and the robot goes back to ①. ⑧ and ⑨ show how human-friendly (HF) behavior is an integrative part even in the presence of multiple humans. In ⑧ and ⑨ the tool surveillance and the physical contact during task execution are shown, respectively.

In summary, the system described here presents a versatile and robust solution with standard components for achieving safe and effective human-robot collaboration and a solution for the bin picking problem. Various explicitly non-trained test subjects were able to intuitively use the system.
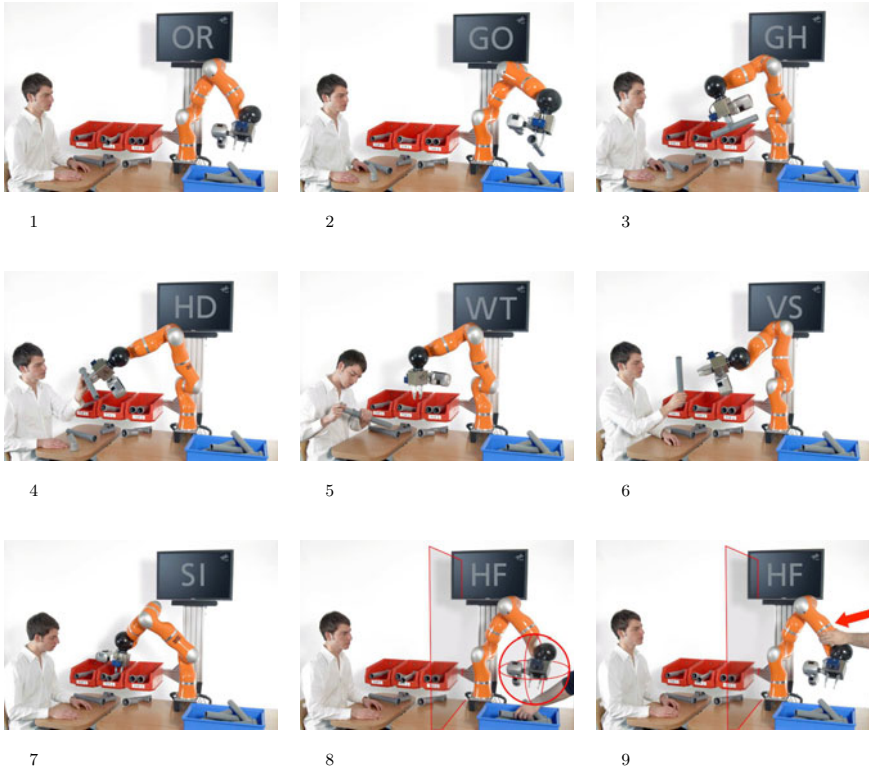
Fig. 16 Interactive bin picking.

5   Conclusion and Outlook

In this paper we proposed a general concept for the robotic co-worker and developed a prototype demonstration for validation based on commercially available technology. We outlined an integrative concept for combining soft-robotics concepts with multi-sensor vision schemes. Flexible hybrid automata can robustly and safely control the modalities of the co-worker in a partially known environment and especially handle the complexity as well as the necessary branching factor during the execution of the tasks. Based on our results in safe physical Human Robot Interaction we were able to effectively combine various control and motion schemes with vision sensing capabilities for the robot to effectively accomplish the task with sufficient safety. Furthermore, exteroceptive sensing is used in combination with compliance control for implementing industrially relevant autonomous tasks. The fusion of these concepts leads to high fault tolerance, supported by the results of the presented bin picking application. The thorough use of multi-sensor information

enabled us to combine the proposed interaction and robust autonomy concepts needed for the **robotic co-worker**.

Future work will focus on broadening the interaction scenarios and incorporate cooperative tasks with joint assembly. For this purpose human surveillance and modeling will become a major focus in order to solve complex and dynamic joint processes. Furthermore, the extensive use of sufficiently fast motion planning is contemplated to reduce the "teaching by demonstration" amount.

Videos are provided at www.robotic.dlr.de/Sami.Haddadin/isrr2009.

Acknowledgment

References

1. Moshner, R.: From handiman to hardiman. Trans. Soc. Autom. Eng. 16, 588–597 (1967)
2. Yamada, Y., Konosu, H., Morizono, T., Umetani, Y.: Proposal of Skill-Assist: a system of assisting human workers by reflecting their skills in positioning tasks. In: IEEE International Conference on Systems, Man, and Cybernetics (SMC 1999), Tokyo, Japan, pp. 11–16 (1999)
3. Hirzinger, G., Heindl, J.: Sensor programming - a new way for teaching a robot paths and forces. In: International Conference on Robot Vision and Sensory Controls (RoViSeC3), Cambridge, Massachusetts, USA (1993)
4. Stemmer, A., Albu-Schäffer, A., Hirzinger, G.: An Analytical Method for the Planning of Robust Assembly Tasks of Complex Shaped Planar Parts. In: Int. Conf. on Robotics and Automation (ICRA 2007), Rome, Italy, pp. 317–323 (2007)
5. Yamada, Y., Hirasawa, Y., Huand, S., Umetani, Y.: Fail-Safe Human/Robot Contact in the Safety Space. In: IEEE Int. Workshop on Robot and Human Communication, pp. 59–64 (1996)
6. Zinn, M., Khatib, O., Roth, B.: A New Actuation Approach for Human Friendly Robot Design. Int. J. of Robotics Research 23, 379–398 (2004)
7. Bicchi, A., Tonietti, G.: Fast and Soft Arm Tactics: Dealing with the Safety-Performance Trade-Off in Robot Arms Design and Control. IEEE Robotics & Automation Mag. 11, 22–33 (2004)
8. Haddadin, S., Albu-Schäffer, A., Hirzinger, G.: Safe Physical Human-Robot Interaction: Measurements, Analysis & New Insights. In: International Symposium on Robotics Research (ISRR 2007), Hiroshima, Japan, pp. 439–450 (2007)

9. Haddadin, S., Albu-Schäffer, A., Hirzinger, G.: Safety Evaluation of Physical Human-Robot Interaction via Crash-Testing. In: Robotics: Science and Systems Conference (RSS 2007), Atlanta, USA, pp. 217–224 (2007)

10. Haddadin, S., Albu-Schäffer, A., De Luca, A., Hirzinger, G.: Collision Detection & Reaction: A Contribution to Safe Physical Human-Robot Interaction. In: IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS 2008), Nice, France, pp. 3356–3363 (2008)

11. Kulic, D., Croft, E.A.: Affective State Estimation for Human-Robot Interaction. IEEE Transactions on Robotics 23(5), 991–1000 (2007)

12. Edsinger, A., Kemp, C.C.: Human-Robot Interaction for Cooperative Manipulation: Handing Objects to One Another. In: IEEE International Symposium on Robot & Human Interactive Communication (RO-MAN 2007), Jeju Island, Korea, pp. 1167–1172 (2007)

13. Dominey, P.F., Metta, G., Natale, L., Nori, F.: Anticipation and Initiative in Dialog and Behavior During Cooperative Human-Humanoid Interaction. In: IEEE-RAS International Conference on Humanoid Robots (HUMANOIDS 2008), Daejeon, Korea, pp. 693–699 (2008)

14. Haddadin, S.: Towards the Human-Friendly Robotic Co-Worker. Master's thesis, Technical University of Munich (TUM) & German Aerospace Center (DLR) (May 2009)

15. Brock, O., Khatib, O.: Elastic Strips: A Framework for Motion Generation in Human Environments. Int. J. Robotics Research 21(12), 1031–1052 (2002)

16. Albu-Schäffer, A., Ott, C., Hirzinger, G.: A Unified Passivity-based Control Framework for Position, Torque and Impedance Control of Flexible Joint Robots. Int. J. of Robotics Research 26, 23–39 (2007)

17. Siciliano, B., Khatib, O. (eds.): Springer Handbook of Robotics. Springer, Heidelberg (2008)

18. Suppa, M., Kielhoefer, S., Langwald, J., Hacker, F., Strobl, K.H., Hirzinger, G.: The 3D-Modeller: A Multi-Purpose Vision Platform. In: Int. Conf. on Robotics and Automation (ICRA), Rome, Italy, pp. 781–787 (2007)

19. Hacker, F., Dietrich, J., Hirzinger, G.: A Laser-Triangulation Based Miniaturized 2-D Range-Scanner as Integral Part of a Multisensory Robot-Gripper. In: EOS Topical Meeting on Optoelectronic Distance/Displacement Measurements and Applications, Nantes, France (1997)

20. Strobl, K.H., Wahl, E., Sepp, W., Bodenmueller, T., Seara, J., Suppa, M., Hirzinger, G.: The DLR Hand-guided Device: The Laser-Stripe Profiler. In: Int. Conf. on Robotics and Automation (ICRA 2004), New Orleans, USA, pp. 1927–1932 (2004)

21. Suppa, M.: Autonomous robot work cell exploration using multisensory eye-in-hand systems. Ph.D. dissertation, Gottfried Wilhelm Leibniz Universität Hannover (2007)

22. Fuchs, S., Hirzinger, G.: Extrinsic and Depth Calibration of ToF-Cameras. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2008), Anchorage, USA, pp. 1–6 (2008)

23. Sepp, W., Fuchs, S., Hirzinger, G.: Hierarchical Featureless Tracking for Position-Based 6-DoF Visual Servoing. In: IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS 2006), Beijing, China, pp. 4310–4315 (2006)