In Proceedings of the 11th IEEE-RAS International Conference on Humanoid Robots (Humanoids), Bled, Slovenia, October 2011

Following Human Guidance to Cooperatively Carry a Large Object

Jörg Stückler and Sven Behnke
Autonomous Intelligent Systems Group
Computer Science Institute VI, University of Bonn, Germany

Abstract—**Carrying a large object like a table is a task that cannot be solved by a single robot or a single human, but that requires two workers, for example one human and one robot. For this human-robot cooperation, the robot must perceive the human and synchronize with its motion. It also must perceive the object to carry. In this paper, we present an approach that uses arm compliance to follow the human guidance on a fast time scale and moves the robot base to restore a nominal position for the arms. For perceiving the object, we acquire a model of it using an RGB-D camera and match this model with the current measurements. This real-time object pose estimate is suitable for approaching and grasping it, as well as for the detection of object lifting and lowering to the ground again. We evaluate our approach in lab experiments using a robot that has an anthropomorphic upper body and an omnidirectional base. We also report on the successful public demonstration of our approach in the @Home league at RoboCup 2011.**

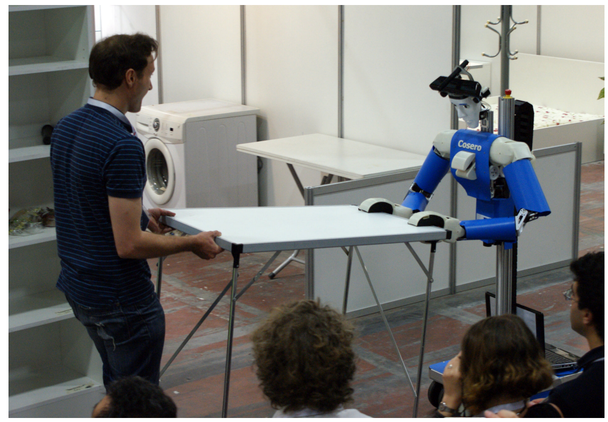

Fig. 1. Cosero cooperatively carries a table with a human. The human guides the robot through the appartment during the RoboCup@Home finale at RoboCup 2011, Istanbul, Turkey.

I. Introduction

Mobile manipulation in everyday environments has been subject to considerable attention recently. Many research groups work on issues like the recognition, grasping, transport, and placement of smaller objects, like bowls, plates, cups, and silverware, e. g., for setting a table.

Carrying large objects like the table itself is a task that cannot be solved by a single robot or a single human, but that requires two workers, for example one human and one robot (s. Fig. 1). It is a typical example for physical human-robot cooperation. For this task, the robot must perceive the human and synchronize with its motion. For grasping, it also must perceive the object to carry and estimate its pose.

In this paper, we present an approach that uses the back-drivability of the joints of our cognitive service robot Cosero to implement task-space compliance. This compliance is then utilized to follow the human guidance with the arms on a fast time scale. On a slower timescale, the robot base moves to restore a nominal position for the arms. For perceiving the object, we acquire a model of it beforehand using an RGB-D camera and match this model with the current measurements later on. This real-time object pose estimate is suitable for approaching and grasping it, as well as for the detection of object lifting and lowering to the ground again. We evaluate our approach using our robot Cosero, which has an anthropomorphic upper body and an omnidirectional base.

The remainder of this paper is organized as follows. After reviewing related work, we briefly present the robot Cosero used for the experiments. In Section IV, we describe how the backdrivability of this robot is used to implement compliant physical human-robot interaction. The perception of the table through an RGB-D camera is detailed in Section V. Section VI integrates both capabilities for cooperative transportation of large objects. We present the results of an experimental evaluation of the proposed method in Section VII.

II. Related Work

The task of carrying a table is an example for the cooperative transport of large objects, which has been investigated using mobile robots. Khatib et al. [1] investigated manipulation of large objects with multiple mobile manipulators. In their work, they extended the augmented object [2] and virtual linkage [3] models to a system of multiple mobile manipulators and proposed a decentralized control scheme. Their method combines the operational space control of the manipulators with a minimization of deviation from the midrange joint positions of the robot bases. This approach requires exact identification of the dynamics of the mobile manipulators.

A different approach proposed by Sugar and Kumar [4] divides the team of robots into leader and followers. The leader tracks a planned trajectory, while the other robots follow the motion of the object using compliant motion control. One or more actively controlled, compliant arms control the grasp forces in the formation allowing the robot platforms to be position controlled. Leader-follower type of control was also investigated by Kosuge et al. [5], [6] using two tracked robots transporting an elongated object. Each mobile robot was

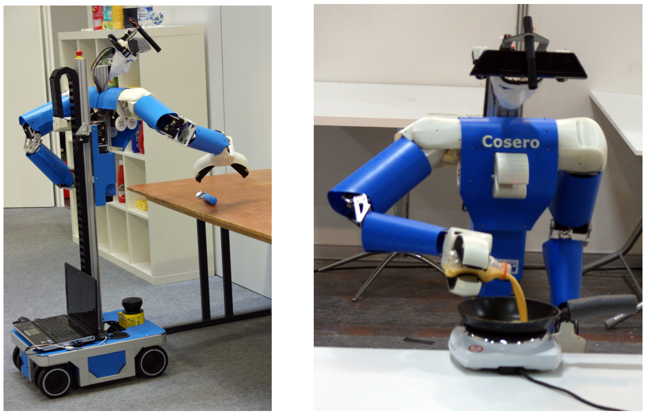

Fig. 2. Cosero grasps a spoon and bakes omelett.

equipped with a force sensor and held the object through a free joint. Given the desired trajectory to the leader, the follower was controlled using dual-caster dynamics. Cetina and Adli [7] investigated the pushing of a cart on a frictionless plane. For the cooperative positioning of the cart by a human and a manipulator, they applied impedance control with parameters obtained from human-human cooperation.

We propose to combine person awareness, real-time object perception, and compliant control to solve a human-robot cooperative task. We implemented the cooperative carrying of a table using the leader-follower principle. The human user can show his intention to carry the table by simply lifting or lowering it. While carrying the table, the robot complies to the human-induced motion of the table and realigns itself to follow it.

The closest related work to ours is work by Yokoyama et al. [8]. They use an HRP2 humanoid robot to carry a large panel together with a human. The robot finds the panel by stereo vision through a model-based recognition system. The walking direction of the robot is controlled by voice commands and by force-torque sensors on the robot wrist. In our approach, the robot recognizes the intention of the person through the motion of the table. It simply perceives when the table is lifted or lowered on the human side through pose tracking. The robot follows the guidance of the human to keep its arms aligned in reference to the table. Instead of specific force-torque sensing in the wrist, we apply compliance control to let the human move the end-effectors through the table.

III. Design of Cognitive Service Robot Cosero

Domestic environments are designed for the specific capabilities of the human body. It is therefore natural to endow the robot with an anthropomorphic upper body scheme for similar manipulation abilities. Furthermore, the actions of the robot become predictable and interpretable when they are performed human-like. In such environments, robots also have to interact closely with humans. By its lightweight design, our robot Cosero [9], shown in Fig. 2, is inherently less dangerous than a heavy-weight industrial-grade robot. The robot should also possess natural sensing capabilities, e. g., vision and audio, since humans design their environments salient and distinguishable in such perception channels. We focused the design of Cosero on such typical requirements for daily settings.

We equipped Cosero with an omnidirectional drive to maneuver in the narrow passages found in household environments. Its two anthropomorphic arms resemble average human body proportions and reaching capabilities. A yaw joint in the torso enlarges the workspace of the arms. In order to compensate for the missing torso pitch joint and legs, a linear actuator in the trunk can move the upper body vertically by approx. 0.9 m. This allows the robot to manipulate on similar heights like humans, which includes picking-up objects from the floor.

Cosero has been constructed from aluminum parts. All joints in the robot are driven by Robotis Dynamixel actuators. These design choices allow for a light-weight and inexpensive construction, compared to other domestic service robots. While each arm has a maximum payload of 1.5 kg and the drive has a maximum speed of 0.6 m/sec, the low weight (in total ca. 32 kg) requires only moderate actuator power.

Cosero perceives its environment with a variety of complementary sensors. A SICK S300 laser scanner measures the distance to objects in a height of approx. 24 cm within 30 m maximum range and with a 270° field-of-view. It is primarily used for 2D mapping and localization. In its vicinity, the robot senses the environment in 3D with a Microsoft Kinect RGB-D camera in its head that is attached to the torso with a pan-tilt unit in the neck.

IV. Mobile Manipulation in Close Interaction with Humans

We developed means for Cosero to solve a variety of mobile manipulation tasks in everyday environments. For this purpose, we combine safe navigation of the robot through the environment with motion control methods for the upper body.

A. Motion Control

We implemented omnidirectional driving for Cosero's eight-wheeled mobile base [10]. The linear and angular velocity of the drive can be set independently and can be changed continuously. We determine the steering direction and the individual wheel velocities of the four differential drives, which are located at the corners of the rectangular base, from an analytical solution to the drive's inverse kinematics.

For the anthropomorphic arms, we implemented differential inverse kinematics with redundancy resolution [10]. Cosero can perform a variety of parameterizable motions like grasping, placing objects, and pouring out containers.

B. Navigation

Cosero localizes and plans paths in a 2D occupancy grid map of the environment. The main sensor for localization is the SICK S300 laser scanner on the mobile base. For 3D collision avoidance, we integrate measurements from any 3D sensing device, such as the tilt laser in the robot's chest.

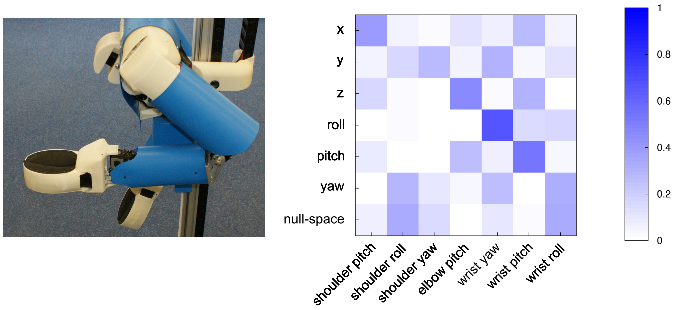

Fig. 3. Example activation matrix for a specific arm pose. The task-space dimensions correspond to the hand's forward/backward (x), lateral (y), and vertical (z) direction, and rotations around the x-axis (roll), y-axis (pitch), and z-axis (yaw).

C. Mobile Manipulation

We propose a coarse-to-fine strategy to align the robot to the objects involved in mobile manipulation. Since the robot is not statically mounted to the environment, it has to estimate its pose in reference to the walls, objects, and persons. For example, when the robot grasps an object from a table, it first approaches the table roughly within the reference frame of the static map. Then, it adjusts in height and distance to the table. Finally, it aligns itself to bring the object into the workspace of its arms.

Cosero grasps objects on horizontal surfaces like the floor, tables and shelves in a height range from ca. 0 m to 1 m [10]. It carries the object and hands it to human users. We also developed solutions to pour-out containers, to place objects on horizontal surfaces, to dispose objects in containers, to grasp objects from the floor, and to receive objects from users.

D. Compliance Control

Compliance in motion control opens up new application domains for manipulation robots. It allows to compensate for inaccurate models and measurement errors. In addition, compliant motion enables direct but safe physical interaction with humans, for example, when the human and the robot cooperatively manipulate an object. We therefore developed task-space compliance control for the arms [11]. For our method, we exploit that the servo actuators are back-drivable and that the torque which the servo applies for position-control can be limited.

From differential inverse kinematics, we derive a method to limit the torque of the joints depending on how much they contribute to the achievement of the motion in task-space. Our approach not only allows to adjust compliance in the null-space of the motion but also in the individual dimensions in task-space. This is very useful when only specific dimensions in task-space shall be controlled in a compliant way. For example, the end-effector can be kept loose in both lateral directions while it keeps the other directions at their targets to pull on a door handle.

We consider the differential mapping from task-space to joint-space

$$\dot{q} \approx J^{\dagger}\dot{x} + \alpha\,(I - J^{\dagger}J)\,\nabla g(q), \tag{1}$$

where $q \in \mathbb{R}^m$ and $x \in \mathbb{R}^n$ are states in joint- and task-space, respectively, J is the Jacobian of the forward mapping $x = f(q)$, g is a cost function of secondary objectives that is projected into the null-space of the motion, and α is a step-size parameter.

We measure the responsibility of each joint on the motion in task-space through the inverse of the Jacobian J with the responsibility matrix

$$R_{task}(t) := \text{abs}\left[J^{\dagger}(q(t)) \begin{pmatrix} \Delta x_1 & 0 & \cdots & 0 \\ 0 & \Delta x_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \Delta x_n \end{pmatrix} \right]. \tag{2}$$

Similarly, we obtain the responsibility of each joint for the null-space motion

$$R_0(t) := \text{abs}\left[\alpha\,(I - J^{\dagger}J)\,\nabla g(q(t))\right]. \tag{3}$$

We then find a torque limit $\tau^q = A(t)\,\tau^x$ for each joint by distributing virtual torque limits τ^x through each compliant task-space direction. The activation matrix $A(t)$ corresponds to the normalized responsibility matrix $R(t) := (R_{task}(t), R_0(t))$. We normalize the responsibilities of each joint to sum to one. Fig. 3 illustrates the activation matrix for an example pose.

In order to implement compliance, the distributed torque limit linearly depends on the displacement in each task direction. By adapting the target to an intermediate target towards the displacement, the speed of the pull-back motion can be adjusted. This also allows for setting the motion fully compliant to external forces.

V. Perception

For a cooperative manipulation task, the robot needs the ability to perceive objects and persons.

A. Perception of Human Interaction Partners

We combine complementary information from laser scanners and vision to continuously detect and keep track of people [12]. The laser scanner on the base detects legs, while the laser scanner in the torso detects trunks of people. In a multi-hypothesis tracker, we fuse both kinds of detections to tracks. In the color images of the Kinect, we can verify that a track belongs to a person by detecting more distinctive human features like faces and upper bodies on the track.

B. Object Perception

We developed real-time tracking and pose estimation of objects with the Kinect RGB-D camera. Our approach processes images in a resolution of 160×120 at a frame-rate of ca. 20 Hz.

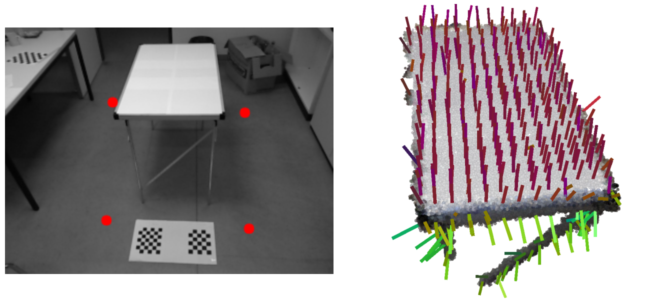

Fig. 4. Left: Example view used to train a model of a table. We select 3D points (red) to form a convex hull on the checkerboard plane. Right: Learned model of a table. The point cloud shows samples from the shape and color distribution modeled in our multi-resolution surfel map at a resolution of 0.05 m. Thick lines indicate surfel normals (color coded for orientation, best viewed in color).

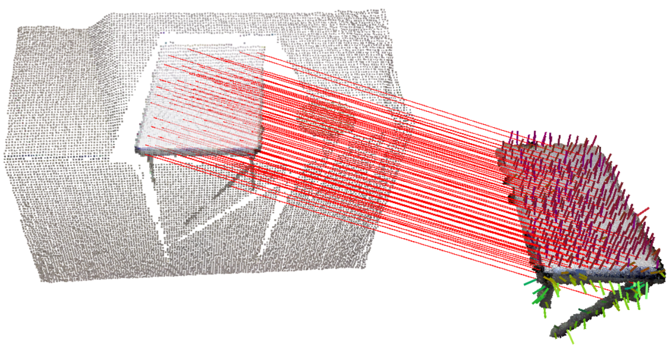

Fig. 5. In order to track the table, we incrementally register RGB-D images to the model at high frame-rates (ca. 20 Hz). Red lines indicate surfel associations between scene and model map.

1) Multi-Resolution Surfel Maps: We represent joint color and shape distributions at multiple resolutions in a probabilistic map. We use octrees as a natural data structure to represent spatial data at multiple resolutions. In each node of the tree, we assume color and shape samples to be normally distributed. We therefore store mean and covariance that can be incrementally updated with new points using the sufficient statistics of the normal distribution. By maintaining the joint distribution of 3D coordinates and color in a 6D normal distribution, we also model the spatial distribution of color. In order to separate chrominance from luminance information and to gain approximate illumination invariance, we represent color in the YUV color space.

RGB-D sensors such as the Microsoft Kinect have specific error characteristics that must be considered. For the Kinect, we add each pixel down to a maximum octree resolution that is proportional to the point's squared distance from the sensor. This captures noise and systematic errors that are caused by the discretization of disparity.

2) Model Learning: In the training phase, our method learns a model of the object from several views (see Fig. 4). We use visual markers, i. e., checkerboard patterns, to obtain the ground truth pose for mapping. This also allows to specify a reference frame in which we can select a bounding volume to segment the object from the background. In one or more images, we manually select 3D points in the reference frame to form a convex hull in the checkerboard plane around the object of interest. We attribute every point to the object that lies within a specific distance range from this plane and whose projection onto the plane lies within the selected convex hull. From these points, we build a surfel map for the object and attach SURF features [13] to the surfels.

3) Real-Time Tracking: We developed fast scan-matching to keep track of the object and to estimate its pose in real-time at frame-rates of ca. 20 Hz (see Fig. 5). For incremental registration of images to the model, we assume that we are already given an initial guess on the correct pose from previous iterations. By this assumption, we can use fast nearest-neighbor look-ups to find corresponding surfels between the maps. We measure the matching likelihood for the surfel correspondences and iteratively optimize this likelihood to find the most likely transformation between the maps.

Instead of comparing the image pixel-wise to the map, we compress the image content in a second (scene) map m_s. We then determine the observation likelihood by the matching likelihood between scene and model map,

$$p(m_s|x, m_m) = \prod_{(i,j)\in\mathcal{C}} p(s_{s,i}|x, s_{m,j}), \tag{4}$$

where $\mathcal{C}$ is the set of surfel correspondences between the maps, and $s_{s,i} = (\mu_{s,i}, \Sigma_{s,i})$ and $s_{m,j} = (\mu_{m,j}, \Sigma_{m,j})$ are corresponding surfels. The observation likelihood of a surfel match is the difference of the surfels under their normal distributions,

$$\begin{aligned} p(s_s|x, s_m) &= \mathcal{N}\left(d(s_m, s_s, x); 0, \Sigma(s_m, s_s, x)\right), \\ &= \mathcal{N}\left(\mu_m - T(x)\mu_s; 0, \Sigma_m + R(x)\Sigma_s R(x)^T\right), \end{aligned} \tag{5}$$

where $T(x)$ is a transformation matrix that rotates and translates the spatial dimensions according to the pose x and $R(x)$ is the corresponding rotation matrix.

We determine the most likely pose by gradient descent on the log likelihood of Eq. (4). In order to determine the correspondences between surfels in both maps, we apply an efficient multi-resolution strategy. Our coarse-to-fine strategy finds the finest resolution possible for each scene surfel by fast position queries in the octree. We only establish a correspondence, if the surfels also match in the color cues.

Our association strategy not only saves redundant comparisons on coarse resolution. It also allows to match surface elements at coarser scales if fine-grained shape and texture details cannot be matched on finer resolutions.

4) Global Registration: Incremental registration requires an initial guess of the pose that is already close to the actual pose. In order to initialize pose tracking, we therefore match SURF features with the current view and find the rigid transformation between the current view towards the object map using weighted SVD [14]. We increase the robustness of our method by means of RANSAC [15].

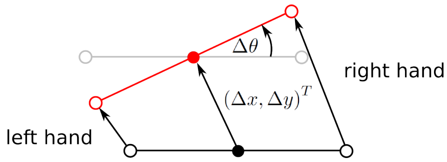

Fig. 6. The robot initially holds the table with two hands at its border (black). The user pulls and pushes the table into the desired direction. The robot complies with its arms to the motion of the table. From the displaced position of the hands (red), the robot steers in proportion to the translational $(\Delta x, \Delta y)^T$ and rotational $\Delta\theta$ displacement.

VI. Human-Robot Cooperative Mobile Manipulation

We demonstrate the use of our approaches to task-space arm compliance and real-time object perception in a cooperative mobile manipulation task. In this task, a human and a robot cooperatively carry a table to a specific location.

A. Task Description

The task starts with the robot approaching the table. It grasps the table and waits for the person to lift it. When the robot detects the lifting of the table, it also lifts the table and starts to follow the motion of the human. The human user may cease the carrying of the table by lowering the table.

B. Approach

We apply our object perception method to find the initial pose of the table towards the robot. The robot then keeps track of the object while it drives towards a predefined approach pose, relative to the table. It grasps the table and waits, until the person lifts the table which is indicated by a significant pitch rotation (0.02 rad) of the table.

As soon as the lifting is detected, the robot also lifts the table. It sets the motion of the grippers compliant in the sagittal and lateral direction, and in yaw orientation. By this, the robot complies when the human pulls and pushes the table.

The robot follows the motion of the human by controlling its omnidirectional base to realign the hands to the initial grasping pose with respect to the robot. We derive a steering command for the robot's base from the displacement of its hands towards the initial pose when the robot lifted the table. The robot measures the actual pose of its hands through forward kinematics. As illustrated in Fig. 6, the mean displacement $(\Delta x, \Delta y)^T$ of both hands directly indicates the translational direction of motion in which the robot is dragged. From the line between the hands, we determine the angular difference $\Delta\theta$ to the initial posture. We set the velocities of the omnidirectional drive proportional to these displacements to follow the motion of the user. In order to prevent from oscillatory behavior, we add a dead range to accept small displacements without compensation through motion of the base.

While the robot follows the motion of the human, it keeps track of the object. When the user puts the table down, it again detects a significant pitch of the table, stops, and also lowers the table.

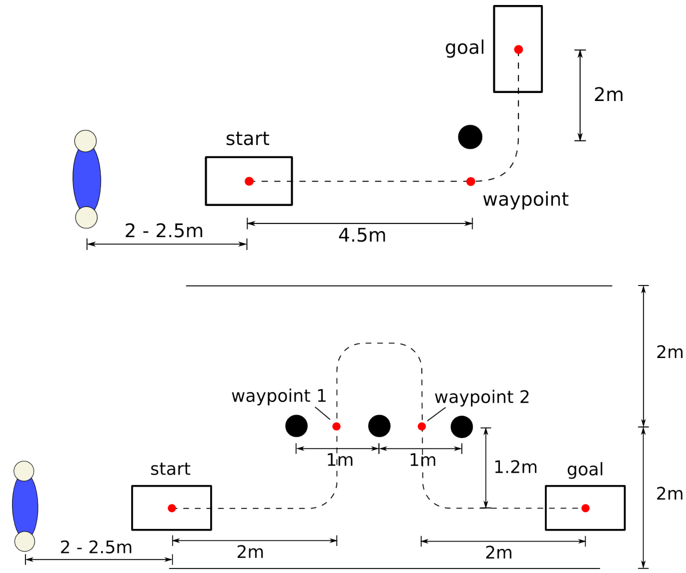

Fig. 7. Experiment setups (top: parcour 1, bottom: parcour 2). The robot (blue) starts behind the table. It approaches the table, grasps it, and waits for the human user. The human guides the robot around the obstacles (black disks) towards the goal.

TABLE I
DURATIONS IN SECONDS TO REACH THE WAYPOINTS AND THE GOALS.

	parcour 1		parcour 2		
trial	waypoint	goal	waypoint 1	waypoint 2	goal
1	25	36	34	74	90
2	23	34	25	68	85
3	23	35	17	69	82
4	23	32	19	55	78
5	22	32	28	62	78
6	22	34	30	78	93
7	20	33	20	73	85
8	22	33	16	75	88
9	20	33	19	67	79
10	22	37	15	59	75
mean	22.2	33.9	22.3	68	83.3
std	1.48	1.66	6.53	7.44	5.89

VII. Experiments

We evaluated our approach in lab experiments and report on a public demonstration at RoboCup 2011.

A. Lab Experiments

1) Experiment Setup: We set up two parcours, shown in Fig. 7, through which the human user should guide the robot while they together carry the table (size: $1\,\mathrm{m} \times 0.6\,\mathrm{m}$). In the first parcour, the table has to be carried on two straight lines and on a turn of 90° around an obstacle. The second parcour consists of two obstacle gates with a width of 1 m. The motion involves turning by 90° into the first gate and by 180° into the second gate in narrow space. In both parcours, the robot starts 2 m to 2.5 m behind the table. It finds the table and aligns itself for grasping. Then the human user lifts the table, and they both begin to carry the table towards the goal, which is unknown to the robot.

We measure the duration for reaching the waypoints and the goals in both parcours. We begin to measure the time when the human starts to move. When the table reached its goal, the human puts the table down. The robot has to detect this event and also put the table down. We accept the table to have reached its goal when its legs are placed within a distance of 0.2 m to their goal positions.

2) Results: Table I shows the durations in ten runs for each parcour and the average timings. The tasks have been successfully achieved in all runs in reasonable time. The timings not only depend on the maximum driving speed of the robot. They are also influenced by the maneuverability of the robot by the human. With our approach, it is intuitive and easy to maneuver the robot. In the first parcour, the human user can achieve the task mainly by pulling in the sagittal direction and turning the table. In order to complete the second parcour, the human has to maneuver the table and the robot in each direction, especially when turning by 180° to pass through the second waypoint. Since the omnidirectional drive is limited in acceleration and the robot exerts force with its hands towards the initial grasping pose, the direction of motion cannot be changed instantaneously by the human user. These properties mainly limit the achievable duration in the tasks.

One problem we encountered was that the human eventually held the table too low shortly above the ground such that the robot detected the lowering of the table too early. However, the robot can simply wait in such a case such that the human may lift the table again and both can resume the task. It is also possible that the human drags too strongly on the table such that the robot looses grip. In such a case, the robot should detect this failure, realign to the table, and regrasp it.

B. Public Demonstration

We demonstrated our approach publicly in the RoboCup@Home league at RoboCup 2011[1]. In the finales, we started our demonstration with the task to rearrange the appartment. The robot waited behind the table until the human user appeared in front of the robot. After the robot detected the person, it approached the table and grasped it. The user lifted the table, which was detected by the robot. Together, they carried the table to its goal location, where the human lowered the table. The robot detected this event, also lowered the table, and resumed with preparing breakfast for the human user. Our finale demonstration was well received by the jury and was awarded the highest score. We could successfully defend the first rank that we achieved during the tests in advance to the finale and won the RoboCup@Home competition in 2011.

VIII. Conclusion

In this paper, we proposed means for cooperative transportation of a table by a human and a robot. In our approach, the human guides the robot by pulling and pushing the table into the desired direction. We apply compliance control to let the robot adapt its arm posture to the human-induced motion of the table. A simple steering method then realigns the robot base towards the table.

The robot perceives the relative pose of the table in real-time with an RGB-D camera. This not only allows to precisely align towards the table for grasping. The robot can also judge when the table is lifted or put down by the human user through measuring its pitch rotation. In lab experiments and at RoboCup 2011, we demonstrated that the human user and the robot could move the table reliably and fast through a parcour. We found that the robot is well maneuverable with our approach.

In future work, we will extend our approach to allow for objects with varying shape. For this purpose, a versatile grasp planning method will be required that derives feasible grasps from the object model. Then, also approach poses for the grasps need to be planned. We could also consider that the robot is informed about the intended goal and steers appropriately to achieve it, instead of purely following the motion of the human. Similar to the work of Cetina and Adli [7], we could learn such a controller from examples how humans carry objects cooperatively.

[1] A video of our performance at RoboCup 2011, including the finale, can be found at http://www.youtube.com/watch?v=nG0mJiODrYw

References

[1] O. Khatib, K. Yokoi, K. Chang, D. Ruspini, R. Holmberg, A. Casal, and A. Baader, "Force strategies for cooperative tasks in multiple mobile manipulation systems," in *Proceedings of International Symposium of Robotics Research (ISRR)*, vol. 7, 1995, pp. 333–342.

[2] O. Khatib, "Object manipulation in a multi-effector robot system," *International Journal of Robotic Research*, p. 137144, 1988.

[3] D. Williams and O. Khatib, "The virtual linkage: A model for internal forces in multi-grasp manipulation," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 1993.

[4] T. G. Sugar and V. Kumar, "Control of cooperating mobile manipulators," *IEEE Trans. on Robotics and Automation*, 2002.

[5] K. Kosuge, T. Oosumi, M. Satou, K. Chiba, and K. Takeo, "Transportation of a single object by two decentralized-controlled nonholonomic mobile robots," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 1998.

[6] T. O. K. Kosuge and H. Seki, "Decentralized control of multiple manipulators handling an object," in *Proc. of IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 1996.

[7] A. E. Cetina and M. A. Adli, "Cooperative control of a human and a robot manipulator for positioning a cart on a frictionless plane," *Mechatronics*, vol. 16, pp. 461–469, 2006.

[8] K. Yokoyama, H. Handa, T. Isozumi, Y. Fukase, K. Kaneko, F. Kanehiro, Y. Kawai, F. Tomita, and H. Hirukawa, "Cooperative works by a human and a humanoid robot," in *Proc. of IEEE Int. Conf. on Robotics and Automation (ICRA)*, vol. 3, 2003, pp. 2985–2991.

[9] J. Stückler, R. Steffens, D. Holz, and S. Behnke, "Real-time 3D perception and efficient grasp planning for everyday manipulation tasks," in *Proc. of the 5th European Conf. on Mobile Robots (ECMR)*, 2011.

[10] J. Stückler and S. Behnke, "Integrating indoor mobility, object manipulation, and intuitive interaction for domestic service tasks," in *Proc. of the 9th IEEE-RAS Int. Conf. on Humanoid Robots (Humanoids)*, 2009.

[11] ——, "Compliant task-space control with back-drivable servo actuators," in *Proc. of the 15th RoboCup International Symposium*, 2011.

[12] ——, "Improving people awareness of service robots by semantic scene knowledge," in *Proc. of the International RoboCup Symposium*, 2010.

[13] H. Bay, T. Tuytelaars, and L. V. Gool, "SURF: speeded up robust features," in *9th European Conference on Computer Vision*, 2006.

[14] K. S. Arun, T. S. Huang, and S. D. Blostein, "Least-squares fitting of two 3-d point sets," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1987.

[15] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Commun. of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.