



**HAL**  
open science

## Collaborative Human-Humanoid Carrying Using Vision and Haptic Sensing

Don Joven Agravante, Andrea Cherubini, Antoine Bussy, Pierre Gergondet, Abderrahmane Kheddar

► **To cite this version:**

Don Joven Agravante, Andrea Cherubini, Antoine Bussy, Pierre Gergondet, Abderrahmane Kheddar. Collaborative Human-Humanoid Carrying Using Vision and Haptic Sensing. ICRA: International Conference on Robotics and Automation, May 2014, Hong Kong, China. pp.607-612, 10.1109/ICRA.2014.6906917 . lirmm-00950727

**HAL Id: lirmm-00950727**

**<https://hal-lirmm.ccsd.cnrs.fr/lirmm-00950727>**

Submitted on 21 May 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Collaborative Human-Humanoid Carrying Using Vision and Haptic Sensing

Don Joven Agravante<sup>1</sup>, Andrea Cherubini<sup>1</sup>, Antoine Bussy<sup>1,2</sup>, Pierre Gergondet<sup>2</sup> and Abderrahmane Kheddar<sup>1,2</sup>

**Abstract**— We propose a framework for combining vision and haptic information in human-robot joint actions. It consists of a hybrid controller that uses both visual servoing and impedance controllers. This can be applied to tasks that cannot be done with vision or haptic information alone. In this framework, the state of the task can be obtained from visual information while haptic information is crucial for safe physical interaction with the human partner. The approach is validated on the task of jointly carrying a flat surface (e.g. a table) and then preventing an object (e.g. a ball) on top from falling off. The results show that this task can be successfully achieved. Furthermore, the framework presented allows for a more collaborative setup, by imparting task knowledge to the robot as opposed to a passive follower.

**Index Terms**— Physical Human-Robot Interaction

## I. INTRODUCTION

Humanoid robots provide many advantages when working together with humans to perform various tasks. This is because humans have an extensive experience in physically collaborating with each other. Hence, humanoids can interact with humans because of their human-like range of motion and sensing capabilities. This reduces the need to learn how to interact with the robot. However, many challenges are still present in the various research areas that study physical human-robot collaboration. Here, the area of interest is using vision and force information together to enable human-robot joint actions, which are collaborative tasks requiring both parties to physically interact with each other (e.g. carrying a large object together). In such tasks, the robot:

- 1) must move safely and regulate interaction forces,
- 2) shares control with a human-in-the-loop,
- 3) can only use its on-board sensors.

The first two items are the main aspect of all physical human-robot collaborative tasks. The last constraint is important for true autonomy. For example if vision has a limited field-of-view, external room cameras should not be used.

Physical human-robot collaboration has largely relied on the use of haptic data (force/torque) for control. This is because the main priority is the regulation of the interaction forces between the human and robot. For example, previous works [1]–[4] have demonstrated that using only haptic

information, a humanoid robot can help a human carry large objects (e.g. a table, a beam or a panel). A possible future application of this is in construction sites [1]. The same scenario can also be applied to the household, such as moving furniture (e.g. table). While doing this task, one can imagine the need to prevent an object on top from falling off. For example in moving a table a short distance, it might be necessary to move it carefully with the objects on top rather than removing the objects, transporting the table and then placing the objects back on top. In this scenario, haptic information alone is not rich enough to give the robot knowledge about the state of the objects on top of the table. But vision can provide such information, being largely complementary to haptics (analogous to human sight and touch). Using both information sources may enable a humanoid to perform more complicated tasks, similar to a human. Although the benefits are great, there are not many established methods integrating vision and force control.

In [5], three general categories for combining vision and force control are identified: traded, hybrid and shared. Traded control is the simplest, and switches between a pure visual servoing controller and a pure force control method given a certain threshold of the task error. In hybrid methods, a prior specification of a “task-frame” [6], [7] is required to decouple vision and force into orthogonal spaces. With this, the controllers can be designed separately. Finally, shared control methods aim at utilizing both vision and force information together in the same space, such that all available information is used [5]. For example, in [8], a force feedback is used to correct the visual servo control trajectory.

In this paper, the impedance control framework [9], is used. This allows a manipulator to be compliant by defining a virtual impedance. In this framework, vision can be used to provide a reference trajectory that is tracked in the absence of external forces [10]–[12]. When contact does occur, it has the properties of shared control methods where vision and force determine the control of the same degree of freedom (DOF) simultaneously. This approach is preferred over the others since it can allow for compliance in all DOF. Previous works using this methodology [10], [11] presented experiments of a robot interacting with objects. Here and in our previous work [12], we use this approach for physical human-humanoid collaboration experiments. Having a human as a physical collaborator means that some issues of this framework need to be revisited: the implication of impedance parameters and how vision and haptics are combined in the

<sup>1</sup>CNRS-UM2 LIRMM UMR 5506, Interactive Digital Human group, 161 Rue Ada, 34392 Montpellier, France {firstnames.lastname}@lirmm.fr.

<sup>2</sup>CNRS-AIST, JRL (Joint Robotics Laboratory), UMI 3218/CRT, Intelligent Systems Research Institute, AIST Central 2, Umezono 1-1-1, Tsukuba, 305-8568, Japan.

context of physical collaboration.

Our work is partly based on the joint object transportation framework which was introduced in [3] and utilized in [12]. The overall system is hybrid - using the 3 DOF that are controlled with only haptic information in [3] and designing shared controllers for the remaining DOF. These new controllers use both vision and force to affect the DOF at the same time. This complicates the semantics to describe the whole system. In [12], vision was used together with haptic data to stabilize the height of the table, providing the reference trajectory of the impedance controller in 1DOF. As a continuation to this work, we use the same general framework of building a visual servoing controller for providing a reference trajectory to the impedance controller. However, we relax the constraint on having a static object in [12]: the object on the beam is free to move so control in 2 DOF is needed. This is a more difficult task and vision is indispensable for acquiring the state of the task. Another contribution of this work is the analysis of collaboration in joint actions when each agent has his/her own notion of how to do the joint task, as opposed to the common leader-follower strategy.

To continue, the general framework is described in Section II. The example task of human-humanoid table carrying while keeping an object on top from falling is then described in Section III. The details on implementing the general framework to this specific task are given in Section IV. Results from experiments are then presented in Section V. The novelty of equal collaboration and the challenges it poses are discussed in Section VI. Finally, Section VII concludes and outlines some future works to be done.

## II. GENERAL CONTROL FRAMEWORK

Our general approach to combining vision and haptic cues is coupling a visual servoing controller to an impedance controller. This simplifies the design by decoupling the vision and force controllers in a systematic way. An overview of the complete control framework is shown in Fig. 1.

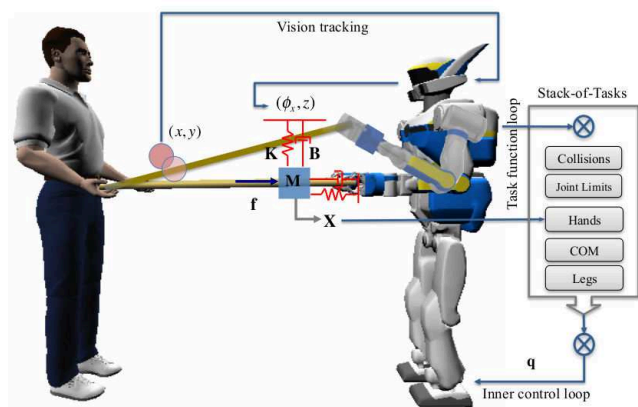


Fig. 1. The general control framework applied to the task of balancing an object on the table

Fig. 1 also shows the task example used in this paper - balancing an object on the table. The following subsections

explain this general framework in a bottom-up approach starting from the lower level controllers and abstracting it higher to the cognitive level. The lowest level of control is the inner joint-level control. This is represented by  $q$  in Fig. 1. To abstract from the joint level to the “task level”, the Stack-of-Tasks framework is used [13]. It is a generalized inverse kinematics abstraction layer that creates a hierarchical organization of different tasks to be executed giving higher priority to critical tasks [13]. It allows for easier integration with sub-tasks. For example, our experiments make use of the walking algorithm in [14] as a sub-task.

### A. Impedance Control

The other sub-task concerns the grippers. In Fig. 1 the humanoid uses its grippers to co-manipulate an object with a human. To do this, it needs to be safe and intuitive to use. Here, impedance control [9] is used to regulate the contact interaction (for safety) between the robot and its environment. It is based on a simple physical analogy to the control - a virtual mass-spring-damper system [9]. This system is governed by the general equation:

$$f = M(\ddot{x}_d - \ddot{x}) + B(\dot{x}_d - \dot{x}) + K(x_d - x). \quad (1)$$

The contact interaction is measured by the force-torque sensors in the robot grippers and is represented as  $f$ . The vectors  $x_d, \dot{x}_d$  and  $\ddot{x}_d$  are a desired pose and its first and second derivative. Correspondingly, vectors  $x, \dot{x}$  and  $\ddot{x}$  represent an actual pose and its first and second derivative. Finally, matrices  $M, B$  and  $K$  are the inertia, damping and stiffness parameters that define the desired virtual mass-spring-damper system [9]. Strictly following the terminology and causality from [9], our implementation on the HRP-2 humanoid, is an “admittance controller” since the robot is position-controlled by the Stack-of-Tasks, which uses the output of  $x, \dot{x}$  and  $\ddot{x}$  from the impedance controller. These are obtained by solving the differential equation of Eq. (1) given the other variables. The parameters  $M, B$ , and  $K$  are determined empirically to provide comfort for the human collaborator. Finally,  $x_d, \dot{x}_d$  and  $\ddot{x}_d$  are the desired pose and trajectory of the mass-spring-damper’s reference position. These are detailed in the next subsection.

### B. Proactive Behavior and Visual Servoing

For the general impedance controller of Eq. (1) a “passive” behavior is defined by setting the desired pose  $x_d$  as constant. This case is illustrated in Fig. 2(a) where only the human knows about the task to be done. This is the “classical” case in human-robot collaboration. In such a case (and considering constant impedance parameters  $M, B, K$ ), the robot’s motion ( $x, \dot{x}, \ddot{x}$ ) can only be initiated by an external force  $f$  due to Eq. (1). Recent research aims to make the robot a proactive follower to make the system more comfortable for the human. A way to achieve this is by creating a suitable desired pose and trajectory ( $x_d, \dot{x}_d, \ddot{x}_d$ ) such that the human effort is minimized [3], [12], [15]. These works differ in the approach taken to produce the desired pose and trajectory. In [15], human motion is predicted by a minimum jerk model

to give the desired pose. In [3], a human pair doing a joint transportation task was studied and it was observed from the data that the pair moves in constant velocity phases during this task. A finite state machine (FSM) is then designed by using the constant velocity assumption, giving the desired pose and trajectory. Haptic cues are used to determine the switching of states in the FSM [3]. Our latest work [12] takes the same approach as the one in this paper and is illustrated by Fig. 2(b). Here, the humanoid is given knowledge of the task. This is done by designing a visual servoing controller specific to the task and using the output as the desired trajectory ( $\mathbf{x}_d, \dot{\mathbf{x}}_d, \ddot{\mathbf{x}}_d$ ) of the impedance controller. This also means that the robot has some autonomy in doing the task driven by its own knowledge of the state of the task. With the reasonable assumption that during the collaborative task human motion is task-driven, the source (human intention to do the task) is taken into account rather than the result (human motion). This differentiates our approach from those that aim to model/predict human motion such as [15].

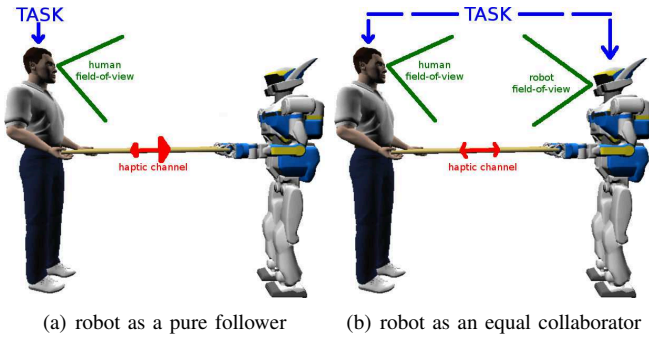


Fig. 2. Human-humanoid collaboration. (a) shows the passive case with the robot as a pure follower guided only by haptic information. (b) illustrates an equal collaboration approach where both human and robot have a complete knowledge of the task (represented by the blue arrows). Furthermore, each uses both vision (green) and haptic (red) information to achieve this task.

Visual servoing consists in controlling robots using visual information [16]. To create the visual servoing portion of the framework, two important components are needed: visual feature tracking and a controller based on this feature [16]. However, in the current state-of-the-art for both modules there is no “best” approach that fits all tasks and problems. Existing methods have important tradeoffs to consider for the whole system [16]. In our works, we take an analytical approach to building the visual servoing portion.

### III. TASK DESCRIPTION

As a test for the general framework described, the task of jointly transporting a surface while keeping a mobile object on top from falling off is used. Fig. 3 illustrates the task with the reference frames and naming convention used in the rest of this paper. The vectors composing the Cartesian frames are color coded: Red-Green-Blue correspond to  $(\vec{x}, \vec{y}, \vec{z})$  respectively.

Fig. 3 shows that the robot can control the table through its hands  $\{rh\}$  and  $\{lh\}$ . The control design consists in driving a reference “control frame”  $\{cf\}$ , rigidly linked to the table,

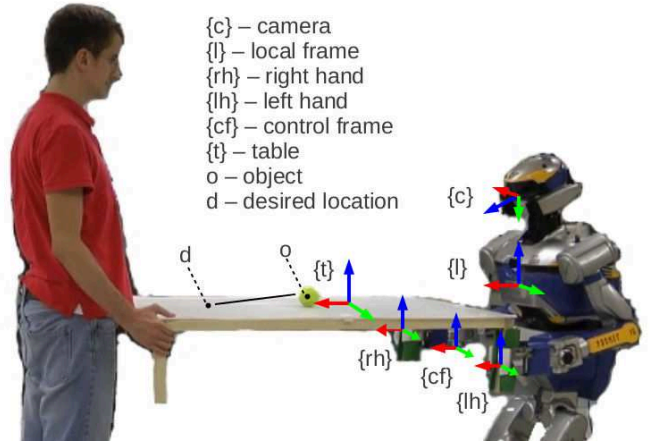


Fig. 3. Human-humanoid table carrying task, with reference frames.

to a desired pose with respect to a local frame  $\{l\}$ , rigidly linked to the robot torso. This pose is represented by the homogeneous transformation matrix  ${}^l\mathbf{T}_{cf}$ . To achieve this, the hand poses  $\{rh\}$  and  $\{lh\}$  are controlled in the local frame according to:

$${}^l\mathbf{T}_h = {}^l\mathbf{T}_{cf} {}^{cf}\mathbf{T}_h \quad h = \{rh, lh\}.$$

Assuming a rigid grasp of the table, the homogeneous transformation matrices  ${}^{cf}\mathbf{T}_{rh}$  and  ${}^{cf}\mathbf{T}_{lh}$  are constant and known once  $\{cf\}$  has been defined. For the implementation of the impedance controller,  ${}^l\mathbf{T}_{cf}$  is converted into the 6-dimensional vector  $\mathbf{x} = {}^l[x, y, z, \phi_x, \phi_y, \phi_z]_{cf}^T$  made up of the Cartesian coordinates and Euler angles (the ZYX convention is used which conveniently places the singularity at  $\phi_y = \pm 90^\circ$ , an impossible case of the joint transportation task).

An intuitive description of the task is to “keep the object on the table from falling off”. The control design can be defined to attract the object ( $o$ ) towards an appropriate desired goal point ( $d$ ) (refer to Fig. 3). To realize this task, direct force application on the object is not possible, since the priority is table transportation. Hence, only an indirect action can be applied by tilting the table to contrast or exploit gravity. To do this, a decoupled approach can be used, where the control of  ${}^t y_o$  is done through  $\phi_x$ , and that of  ${}^t x_o$  through  $z$ . It is chosen to regulate  ${}^t x_o$  with  $z$  and not with  $\phi_y$  because the latter option would make the task uncomfortable for the human. In fact, controlling  $\phi_y$  forces the human to actively move his/her  $z$  position. Instead, by controlling  $z$  and leaving  $\phi_y$  compliant, the human at the other end just needs to be compliant in his/her  $\phi_y$ , which is more comfortable.

To integrate the whole system, the important part is defining the admittance controller’s desired trajectory ( $\mathbf{x}_d, \dot{\mathbf{x}}_d, \ddot{\mathbf{x}}_d$ ) for all 6 DOF. The vision-based control takes care of two DOF ( $z, \phi_x$ ). Three DOF ( $x, y, \phi_z$ ) of the pose are defined from the FSM of our group’s earlier work [3], [4] (described briefly in Section II-B). Finally, the remaining DOF ( $\phi_y$ ) is made compliant by setting  $\phi_{y,d} = 0$ .

#### IV. IMPLEMENTATION DETAILS

As explained in Section II-B, the approach is to design a visual servoing controller to make the robot proactive throughout the task. Two main components are needed: visual tracking and the control design.

##### A. Vision Algorithm

In the HRP-2 robot, RGB-D (color + depth) data is obtained from an embedded ASUS Xtion device located in the head. Fig. 4 shows typical data of the task. The aim of the vision algorithm is to process this raw data into visual features that can be used for control. An error signal can be defined by  ${}^t x_o - {}^t x_d$  and  ${}^t y_o - {}^t y_d$ . For the example task here,  $z$  is irrelevant, since  ${}^t z_d \equiv {}^t z_o$ . Since the desired location  ${}^t(x, y)_d$  is arbitrarily defined, the vision algorithm only needs to obtain  ${}^t(x, y)_o$ . A variety of vision algorithms that can do this may be used, with speed as another consideration. For example, given the object model and the table model, it is possible to use a model based tracker. Designing a novel vision algorithm is not the focus of this work, so we use well-known methods [17]–[19]. Nevertheless, the methods used here are briefly described for completeness.

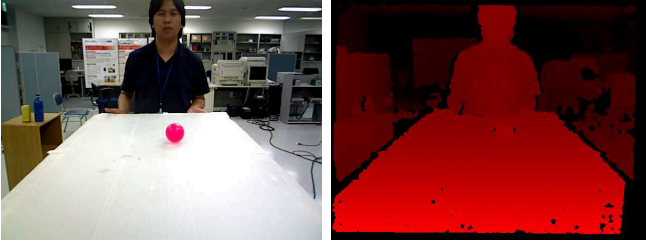


Fig. 4. Typical raw data (RGB + Depth images) during the task. Left: RGB image. Right: Depth image, where dark red→bright red corresponds to “far”→“near” and black pixels are regions without data.

The features used here are the centroids of the object and of the table. The first step is to segment these from the image. Color segmentation is used in our system. For example the pink object in Fig. 4 and yellow object in Fig. 3 can be easily characterized and thresholded by a specific hue range and a high saturation (from the HSV color space). To add robustness, morphological operations (opening and closing) are used to remove outliers. After this, sliding window detection (sped up using the image pyramids concept) finds the most probable location. The centroid of the detected blob is  $(u, v)$  in pixel coordinates. This is then converted into  ${}^c x_o$  and  ${}^c y_o$  by using the intrinsic camera calibration parameters  $(f_x, f_y, c_x, c_y)$  and the depth  ${}^c z_o$  in the following equations:

$${}^c x_o = \frac{{}^c z_o(u - c_x)}{f_x}, \quad {}^c y_o = \frac{{}^c z_o(v - c_y)}{f_y}. \quad (2)$$

The next step is to segment the table in the image. A flood fill algorithm [19] is run in saturation-value-depth space. This algorithm starts with a “seed” point and grows the region based on a connectivity criterion between neighboring pixels. Here, the seed point is the bottom pixel of the ball. A low

saturation and high value characterize well the “white” color of the table. The addition of depth ensures connectivity in Cartesian space, simplifying for example the segmentation between table and floor pixels. Finally, some morphological operations (opening and closing) are done to remove outliers. From these segmented points, the Cartesian centroid is used as  ${}^c \mathbf{t}_t$  (a translation vector). The Cartesian coordinates of the object in the table frame are then obtained by:

$${}^t \mathbf{t}_o = {}^c \mathbf{T}_t^{-1} {}^c \mathbf{t}_o. \quad (3)$$

The homogeneous transformation matrix  ${}^c \mathbf{T}_t$  is composed of the table centroid position  ${}^c \mathbf{t}_t$  and the rotation matrix  ${}^c \mathbf{R}_t$ . A simple approximation consists in setting  ${}^c \mathbf{R}_t$  equal to  ${}^c \mathbf{R}_{cf}$ , which is obtained from proprioception.

##### B. Vision-Based Control

The control design needs to drive  ${}^t \mathbf{t}_o$  to  ${}^t \mathbf{t}_d$ . Several existing methods can be used. Here, a simple PD controller is used such that:

$$C_i(s) = K_{p,i} + K_{d,i}s \quad i = \{x, y\}. \quad (4)$$

This choice is justified by analyzing the task using a simple sliding model (i.e., neglecting friction and angular momentum). Fig. 5 illustrates the necessary variables for this analysis. Since a control with  $z$  rather than  $\phi_y$  is desired, the trigonometric identity  $z_r = l_t \sin \phi_y$  is used, where  $l_t$  is the length of the table and  $z_r$  is the differential height.  $z_r$  can be converted to  $z$  by a trivial change of frame.

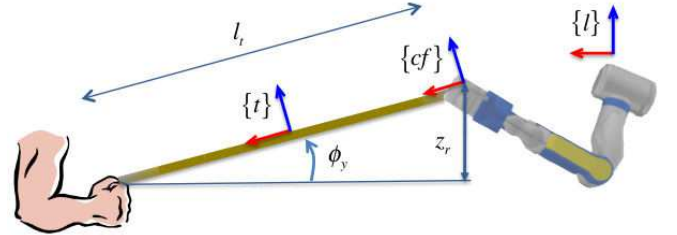


Fig. 5. A simplified “thin beam” model used to control the table height

The Lagrangian equation of motion along  ${}^t \vec{x}$  is:

$$m\ddot{x} = mg \sin \phi_y = mg z_r / l_t. \quad (5)$$

Along  $y$ , linearization of the Lagrangian equation about  $\phi_x = 0$  leads to:

$$m\ddot{y} = -mg\phi_x. \quad (6)$$

Taking the Laplace transforms of these two equations yields:

$$\begin{cases} s^2 X(s) = g Z_r(s) / l_t \\ s^2 Y(s) = -g \Phi(s). \end{cases} \quad (7)$$

Rearranging, the transfer functions describing the dynamics on the 2 DOF can be derived:

$$\begin{cases} P_x(s) = \frac{X(s)}{Z_r(s)} = \frac{g}{l_t s^2} \\ P_y(s) = \frac{Y(s)}{\Phi(s)} = -\frac{g}{s^2}. \end{cases} \quad (8)$$

It should be noted that both are double integrators. As such, they are only marginally stable when feedback controlled with a Proportional gain. But the Proportional Derivative controller (PD) chosen can be used. The denominator of the closed loop system transfer function in the two cases is:

$$\begin{cases} D_x(s) = l_t s^2 + gK_{d,x}s + gK_{p,x} \\ D_y(s) = s^2 - gK_{d,y}s - gK_{p,y}. \end{cases} \quad (9)$$

The two systems are asymptotically stable if all the roots of these two polynomials have non-multiple negative real parts. This condition is verified, for a second order polynomial, if all the coefficients are strictly positive. In the case of the characteristic polynomials in (9), this is equivalent to:

$$K_{p,x} > 0 \quad K_{d,x} > 0 \quad K_{p,y} < 0 \quad K_{d,y} < 0. \quad (10)$$

Finally, the applied controllers are:

$$\begin{cases} z = K_{p,x}(x_d - x) - K_{d,x}\dot{x} \\ \phi_x = K_{p,y}(y_d - y) - K_{d,y}\dot{y}. \end{cases} \quad (11)$$

By numerical differentiation  $\dot{x}$  (and  $\dot{y}$ ) is obtained as:

$$\dot{x}(t) = \frac{x(t) - x(t - \Delta t)}{\Delta t},$$

with  $\Delta t$  the sampling step. Tuning the gains in (11) according to (10) guarantees stability of the closed loop system, as long as the linear approximation is valid. This implies that  ${}^t\mathbf{t}_o$  will converge to  ${}^t\mathbf{t}_d$ , as desired. The outputs of (11) are fed to the admittance controller (1) as desired values  $z_d$  and  $\phi_{x,d}$ . Numerical differentiation is used to obtain  $\dot{z}$ ,  $\dot{\phi}_x$  in  $\dot{\mathbf{x}}_d$ . However, for  $\dot{\mathbf{x}}_d$  a piece-wise constant velocity is assumed such that  $\ddot{z} = \dot{\phi}_x = 0$ . This also prevents too much noise introduced by a second numerical differentiation.

## V. RESULTS

For the experiments, we chose a ball to be the moving object. This makes it similar to a well-studied problem/example in control theory: the “ball-and-plate” system, which is a 2-DOF generalization of the “textbook example” ball-on-beam system (used to study advanced control methods [20]). Although similar, significant differences exist - notably that collaboration is the main issue here.

Several experiments were performed and with 2 different balls - a yellow tennis ball which tends to move slower and a pink ball which moves quite fast. A few different users also tested this early system, but as the described experience was similar this is not discussed here. Some experiments are shown in the accompanying video and in Fig. 6. The video also shows some results of the vision algorithm detecting the ball and the table. In the initial experiments, both human and humanoid stand stationary and balance the ball on the table. Some disturbance is then introduced (e.g. the ball is pushed by another person) and the gains of the PD controller are tuned according to (10) in order to be able to handle such a disturbance. After “light” gain tuning of the vision-based controller with such experiments, we test the complete system where the human-humanoid dyad transport the table with the ball on top. Here, walking introduces a significant disturbance that can move the ball. The experiments show

that although the ball moves a lot, it doesn’t fall off the table during this transportation task.

From the recorded data of the force/torque sensors in the HRP-2 wrists, we found that during this task  $\tau_x$  (the total torque about the x-axis of {cf}) averages to about 0 Nm, which means that this interaction torque with the human is regulated well. Furthermore,  $f_z$  (again with {cf} as the reference) averages to about 12 N. This means that the robot carries part of the weight of the table and thus lightens the burden on the human. Finally, we notice that in both signals a noticeable oscillation occurs which correlates to the frequency of the walking gait and the disturbance that it causes.

## VI. DISCUSSION ON EQUAL COLLABORATION

The results show that the complete system (Fig. 2b) can do the job well: the vision-based controller tries to keep the ball on the table while the impedance controller regulates interaction forces. A simple analysis of Fig. 2 shows that a disadvantage of the pure follower (Fig. 2a) is that the success/failure of the vision task depends solely on the human partner. Specifically, the human needs to use his/her vision to observe the state of the task and then apply a sufficient force to haptically communicate to the robot what s/he wants to do. Instead, in Fig. 2b the cognitive load of the task is shared in some capacity - both human and robot are able to observe the state of the task and act accordingly. However, this sharing can become a disadvantage when the human and robot disagree on the state of the task and the action to take [21]. Experimentally, this is handled in our system by making the robot more compliant and less stiff (impedance parameter tuning). This ensures that the human can always safely impose his/her intention through the haptic channel. This also shows a possible extension of the system which is to dynamically change the impedance parameters: making it more stiff when the robot is more certain of his observations and more compliant when there is more uncertainty. In effect, this makes the impedance parameters a method to weigh the importance between visual (task knowledge) and haptic (human intention) information channels. But, it is important to note that this disadvantage of equal collaboration also applies to human-human pairs and more generally in teams - “teamwork” (or the lack of it). Some preliminary experiments have been made with both the passive follower and the approach of equal collaboration and the advantages/disadvantages briefly described here can be observed by the human collaborator. One difficulty in presenting these results is in the use of proper evaluation methods since the most important aspect - the comfort of the human collaborator - is very subjective. Another difficulty is to separate the contribution of the human and robot. Although in the results presented here the human is told to be more “passive” (does not try “too much” to keep the ball on the table) he also does not try to make the ball fall off, since teamwork is a factor in the overall result. The resolution of these issues is left for future work.

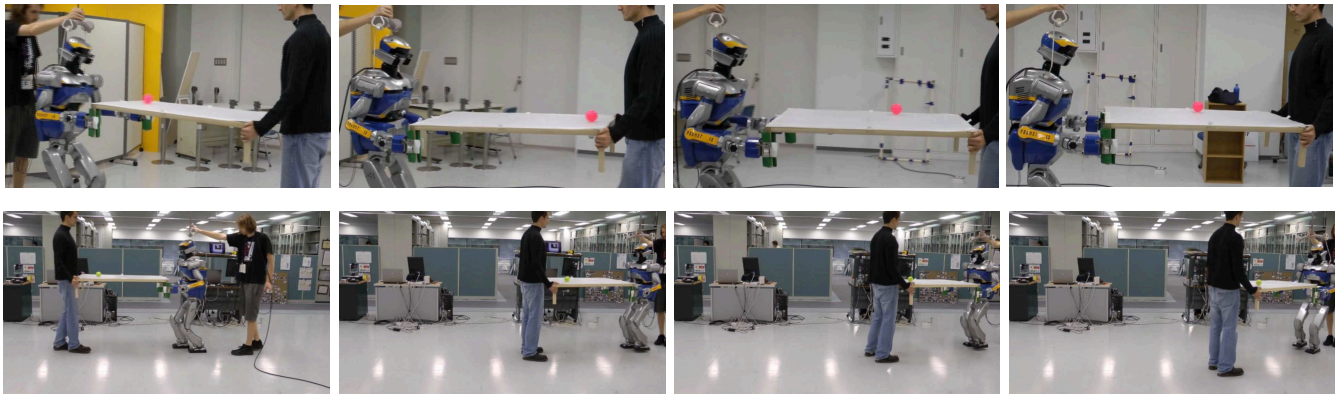


Fig. 6. Snapshots of two experiments where the human-humanoid dyad transports a table with a ball (fast in the top sequence, slow in the bottom one).

## VII. CONCLUSION AND FUTURE WORK

In this paper, a general framework for human-robot joint collaborative tasks was presented. It uses a visual servoing controller to realize the task and a haptic channel to recognize human intention. Both vision and force control are combined in the impedance control framework. This is implemented and tested on a joint transportation task where a human and humanoid robot carry a table with a freely moving ball on top. The objective is to transport the table while keeping the ball from falling off. This task is used to explore some important issues in robotics: the combination of vision and force information and the issues concerning collaboration - safety and effective human-robot collaboration strategies.

To continue the work here, it is planned to further investigate the combination of vision and force information. Another major area for continued study is in collaboration, such as the idea of dynamically changing the impedance parameters described in Section VI. Further works are to utilize good statistical methodology and experiments with different users (usability study) to better analyze the qualitative results.

## VIII. ACKNOWLEDGEMENT

This work is supported in part by the FP7 IP RoboHow.Cog project (www.robohow.eu). FP7-ICT-2011-7 Contract No 288533. The authors would like to thank François Keith, Damien Petit, Hervé Audren, Kenji Kaneko and Eiichi Yoshida for their help with the experiments.

## REFERENCES

- [1] 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 *IEEE International Conference on Robotics and Automation*, vol. 3, pp. 2985–2991, IEEE, 2003.
- [2] P. Evrard and A. Kheddar, "Homotopy switching model for dyad haptic interaction in physical collaborative tasks," in *EuroHaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, pp. 45–50, IEEE, 2009.
- [3] A. Bussy, A. Kheddar, A. Crosnier, and F. Keith, "Human-humanoid haptic joint object transportation case study," in *IEEE/RSJ International Conference on Robots and Intelligent Systems*, pp. 3633–3638, IEEE, 2012.
- [4] A. Bussy, P. Gergondet, A. Kheddar, F. Keith, and A. Crosnier, "Proactive behavior of a humanoid robot in a haptic transportation task with a human partner," in *IEEE International Symposium on Robot and Human Interactive Communication*, pp. 962–967, IEEE, 2012.
- [5] B. J. Nelson, J. D. Morrow, and P. K. Khosla, "Improved force control through visual servoing," in *Proc. of the American Control Conference*, vol. 1, pp. 380–386, IEEE, 1995.
- [6] M. T. Mason, "Compliance and force control for computer controlled manipulators," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 11, no. 6, pp. 418–432, 1981.
- [7] J. Baeten, H. Bruyninckx, and J. De Schutter, "Integrated vision/force robotic servoing in the task frame formalism," *The International Journal of Robotics Research*, vol. 22, no. 10-11, pp. 941–954, 2003.
- [8] M. Prats, P. Martinet, A. P. del Pobil, and S. Lee, "Vision force control in task-oriented grasping and manipulation," in *IEEE/RSJ International Conference on Robots and Intelligent Systems*, pp. 1320–1325, IEEE, 2007.
- [9] N. Hogan, "Impedance control - An approach to manipulation. I - Theory. II - Implementation. III - Applications," *ASME Transactions Journal of Dynamic Systems and Measurement Control B*, vol. 107, pp. 1–24, Mar. 1985.
- [10] A. De Santis, V. Lippiello, B. Siciliano, and L. Villani, "Human-robot interaction control using force and vision," *Advances in Control Theory and Applications*, pp. 51–70, 2007.
- [11] G. Morel, E. Malis, and S. Boudet, "Impedance based combination of visual and force control," in *IEEE International Conference on Robotics and Automation*, vol. 2, pp. 1743–1748, IEEE, 1998.
- [12] D. J. Agravante, A. Cherubini, A. Bussy, and A. Kheddar, "Human-humanoid joint haptic table carrying task with height stabilization using vision," in *IEEE/RSJ International Conference on Robots and Intelligent Systems*, IEEE, 2013.
- [13] N. Mansard, O. Stasse, P. Evrard, and A. Kheddar, "A versatile generalized inverted kinematics implementation for collaborative working humanoid robots: The stack of tasks," in *International Conference on Advanced Robotics*, pp. 1–6, IEEE, 2009.
- [14] A. Herdt, N. Perrin, P.-B. Wieber, et al., "Walking without thinking about it," in *IEEE/RSJ International Conference on Robots and Intelligent Systems*, pp. 190–195, 2010.
- [15] Y. Maeda, T. Hara, and T. Arai, "Human-robot cooperative manipulation with motion estimation," in *IEEE/RSJ International Conference on Robots and Intelligent Systems*, vol. 4, pp. 2240–2245, IEEE, 2001.
- [16] F. Chaumette and S. Hutchinson, "Visual servo control. i. basic approaches," *Robotics & Automation Magazine, IEEE*, vol. 13, no. 4, pp. 82–90, 2006.
- [17] R. C. Gonzalez, R. E. Woods, and S. L. Eddins, *Digital image processing using MATLAB*, vol. 2. Gatesmark Publishing Tennessee, 2009.
- [18] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second ed., 2004.
- [19] G. Bradski and A. Kaehler, *Learning OpenCV: Computer vision with the OpenCV library*. O'Reilly Media, Incorporated, 2008.
- [20] J. Hauser, S. Sastry, and P. Kokotovic, "Nonlinear control via approximate input-output linearization: The ball and beam example," *IEEE Transactions on Automatic Control*, vol. 37, no. 3, pp. 392–398, 1992.
- [21] A. Kheddar, "Human-robot haptic joint actions is an equal control-sharing approach possible?," in *4th International Conference on Human System Interactions (HSI)*, pp. 268–273, IEEE, 2011.