



HAL
open science

Compliant Physical Interaction based on External Vision-Force Control and Tactile-Force Combination

Mario Prats, Philippe Martinet, Sukhan Lee, Pedro J Sanz

► **To cite this version:**

Mario Prats, Philippe Martinet, Sukhan Lee, Pedro J Sanz. Compliant Physical Interaction based on External Vision-Force Control and Tactile-Force Combination. MFI 2008 - IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems, Aug 2008, Seoul, South Korea. pp.405-410, <10.1109/MFI.2008.4648029>. <hal-02468548>

HAL Id: hal-02468548

<https://inria.hal.science/hal-02468548v1>

Submitted on 5 Feb 2020

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.



HAL Authorization

Compliant Physical Interaction based on External Vision-Force Control and Tactile-Force Combination

Mario Prats, Philippe Martinet, Sukhan Lee and Pedro J. Sanz

Abstract—This paper presents external vision-force control and force-tactile integration in three different examples of multisensor integration for robotic manipulation and execution of everyday tasks, based on a general framework, presented in [1], that enables sensor-based compliant physical interaction of the robot with the environment. The first experiment is a door opening task where a mobile manipulator has to pull open the handle with a parallel jaw gripper by using vision and force sensors in a novel external vision-force coupling approach [2] where the combination is done at the control level; the second one is another vision-force door opening task, but including a sliding mechanism and a different robot, endowed with a three-fingered hand; finally, the third task is to grasp a book from a bookshelf by means of tactile and force integration. The purpose of this paper is twofold: first, to show how vision and force modalities can be combined at the control level by means of an external force loop. And, second, to show how the sensor-based manipulation framework proposed in [1] can be easily applied to very different physical interaction tasks in the real world, allowing for dependable and versatile manipulation

I. INTRODUCTION

Management of uncertainty is one of the big challenges in the design of robot applications able to operate autonomously in unstructured environments like human or outdoor scenarios. Robot grasping and manipulation of objects is not an exception, but one of the fields in robotics more affected by the uncertainties of the real world. The use of data coming from multiple sensors is a valuable tool to overcome these difficulties.

In particular, vision and force are the most important sensors for task execution. Whereas vision can guide the hand towards the object and supervise the task, force feedback can locally adapt the hand trajectory according to task forces. When dealing with disparate sensors, a fundamental question stands: how to effectively combine the measurements provided by these sensors? One approach is to combine the measurements using multi-sensor fusion techniques [3]. However, such method is not well adapted to vision and force sensors since the data they provide measure fundamentally different physical phenomena, while multi-sensory fusion is aimed in extracting a single information from disparate sensor data. Another approach is to combine visual and force data at the control level, as we propose in this paper, where a novel vision-force control law [2], based on the concept of

external control, does the coupling in sensor-space, which allows to control vision and force on all the degrees of freedom, whereas only the vision control law is directly connected to the robot.

In the literature we can find several applications of robots performing physical interaction tasks in real life environments, as for example [4], [5] and [6]. However, very few approaches consider multiple sensors in a general framework. Instead, ad-hoc applications are usually implemented, leading in specialized robots unable to perform many different manipulation tasks.

In [1], we have presented a general framework for enabling compliant physical interaction based on multiple sensor information. The purpose of this paper is twofold: first, to show how vision and force modalities can be combined at the control level by means of external vision-force control. And, second, to show how this framework can be used for the fast implementation of sensor-based physical interaction tasks in very different robotic systems, as well as its versatility, allowing to perform very different tasks in household environments, without having specific models of them, and without being specifically programmed for a particular task. For this, three different applications are described in different scenarios, and involving different robots and sensors: the first one is a door opening through vision-force control; the second one is another vision-force door opening task, but including a sliding mechanism and a different robot, endowed with a three-fingered hand; finally, the third task is to grasp a book from a bookshelf by means of tactile and force integration.

The paper is organized as follows: the sensor-based physical interaction framework detailed in [1] is outlined in section II. Sections III, IV and V describe three different examples of the framework application, involving different robots and sensor combinations. The novel vision-force coupling approach is introduced in the first experiment. Conclusions and future lines are given in section VI.

II. A FRAMEWORK FOR SENSOR-BASED COMPLIANT PHYSICAL INTERACTION

A framework for describing physical interaction tasks, based on multisensor integration is presented in detail in another paper of this conference [1]. Our approach is based on the *Task Frame Formalism* [7], where a *task frame* is defined as a cartesian coordinate system, given in object coordinates, where the task is defined in terms of velocity and force references, according to the natural constraints imposed by the environment.

M. Prats and P.J. Sanz are with Robotic Intelligence Lab, Jaume-I University, Castellon, Spain. (mprats@uji.es)

P. Martinet is with LASMEA, Blaise Pascal University, Clermont-Ferrand, France. (Philippe.Martinet@lasmea.univ-bpclermont.fr)

S. Lee is with ISRC, Sungkyunkwan University, Suwon, South Korea. (lsh@ece.skku.ac.kr)

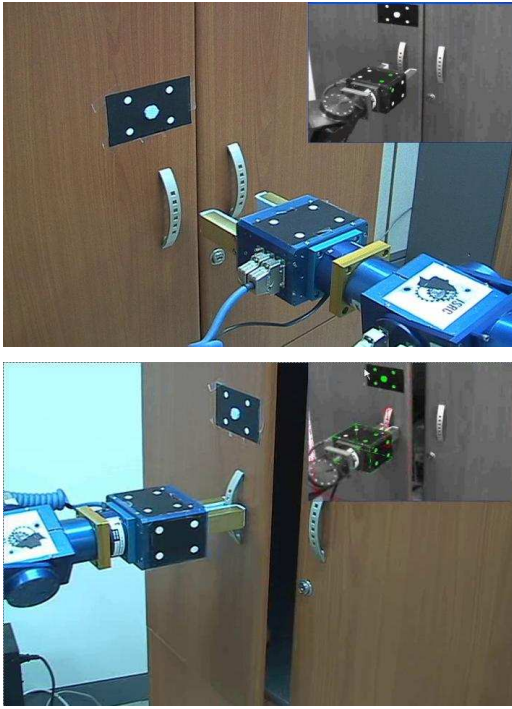


Fig. 2. The mobile manipulator at ISRC opening a door by means of force and vision combination

B. Task execution

For this task, a position-based vision-force servoing closed-loop approach has been adopted. A robot head observes both the gripper and the object and tries to achieve a relative pose between both.

1) *Estimating hand-handle relative pose:* Virtual visual servoing [9] is used to estimate the pose of the hand and the handle, using a set of point features drawn on a pattern whose model and position is known. One pattern is attached to the gripper, in a known position ${}^E\mathbf{M}_{GP}$. Another pattern is attached to the object, also in a known position with respect to the object reference frame: ${}^O\mathbf{M}_{OP}$. As future research we would like to implement a feature extraction algorithm in order to use natural features of the object instead of the markers. Figure 1 shows the different frames involved in the relative pose estimation process and the task.

The matrix ${}^H\widehat{\mathbf{M}}_G$, which relates hand and handle, is computed directly from the pose estimation of the gripper and the object, according to the following expression:

$$({}^C\mathbf{M}_{GP} \cdot {}^E\mathbf{M}_{GP}^{-1} \cdot {}^E\mathbf{M}_H)^{-1} \cdot {}^C\mathbf{M}_{OP} \cdot {}^O\mathbf{M}_{OP}^{-1} \cdot {}^O\mathbf{M}_G \quad (2)$$

where ${}^C\mathbf{M}_{GP}$ is an estimation of the pose of gripper pattern, expressed in the camera frame, and ${}^C\mathbf{M}_{OP}$ is an estimation of the object pattern pose, also in the camera frame. ${}^E\mathbf{M}_H$ and ${}^O\mathbf{M}_G$ are the hand and grasp frame positions with respect to the end-effector and the object reference frame respectively, as set in the previous points.

2) *Improving the grasp:* After pose estimation, a measure of the error between the desired (${}^H\mathbf{M}_G$) and current (${}^H\widehat{\mathbf{M}}_G$)

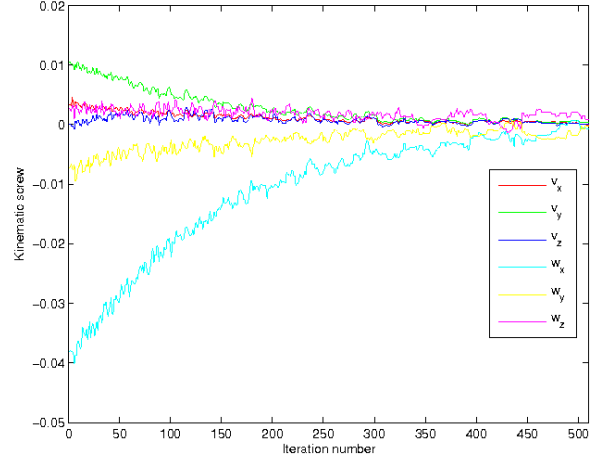


Fig. 3. Kinematic screw during hand-to-handle alignment, following an exponential decrease, which is the classical behaviour on visual servoing tasks. The kinematic screw converges to zero, when ${}^H\widehat{\mathbf{M}}_G = {}^H\mathbf{M}_G$.

hand-grasp relative pose is obtained. It is desirable to design a control strategy so that the grasp is continuously improving during task execution. With a vision-based approach, any misalignment between the gripper and the handle (due to sliding, model errors, etc.) can be detected and corrected through a position-based visual servoing control law [10]. We set the vector \mathbf{s} of visual features to be $\mathbf{s} = (\mathbf{t} \quad \mathbf{u}\theta)^T$, where \mathbf{t} is the translational part of the homogeneous matrix ${}^H\widehat{\mathbf{M}}_G$, and $\mathbf{u}\theta$ is the axis/angle representation of the rotational part of ${}^H\widehat{\mathbf{M}}_G$. The velocity in the hand frame τ_H is computed using a classical visual servoing control law:

$$\tau_H = -\lambda \mathbf{e} + \frac{\partial \mathbf{e}}{\partial \mathbf{t}} \quad (3)$$

where $\mathbf{e}(\mathbf{s}, \mathbf{s}^d) = \widehat{\mathbf{L}}_s^+ (\mathbf{s} - \mathbf{s}^d)$ (in our case, $\mathbf{s}^d = \mathbf{0}$, as ${}^H\mathbf{M}_G = \mathbf{I}_{4 \times 4}$). The interaction matrix $\widehat{\mathbf{L}}_s$ is set for the particular case of position-based visual servoing:

$$\widehat{\mathbf{L}}_s = \begin{pmatrix} -\mathbf{I}_{3 \times 3} & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 3} & -\mathbf{L}_w \end{pmatrix}$$

$$\mathbf{L}_w = \mathbf{I}_{3 \times 3} - \frac{\theta}{2} [\mathbf{u}]_{\times} + \left(1 - \frac{\text{sinc}(\theta)}{\text{sinc}^2(\frac{\theta}{2})} \right) [\mathbf{u}]_{\times}^2$$

where $[\mathbf{u}]_{\times}$ is the skew anti-symmetric matrix for the rotation axis \mathbf{u} . Finally, the end-effector motion is computed as $\tau_E = {}^E\mathbf{W}_H \cdot \tau_H$.

3) *Task motion and coping with uncertainties:* The end-effector velocity that the robot has to achieve in order to perform the task motion, is computed by transforming the task velocity, from the task frame to the end-effector frame, according to equation 1.

Even if the relative pose between the hand and the handle, ${}^H\widehat{\mathbf{W}}_G$, is estimated and corrected continuously, this estimation can be subject to important errors, considering that it is based on vision algorithms, that can be strongly

affected by illumination, camera calibration errors, etc. Due to this fact, the robot motion is also subject to errors, and cannot match exactly the desired motion for the task. As the hand is in contact with the environment, any deviation of the hand motion regarding the task trajectory will generate important forces on the robot hand that must be taken into account.

We adopt a novel external vision/force control law (see [2] for details) for integrating vision and force and coping with uncertainties. With this approach, the force vector, with current external forces, is used to create a new vision reference according to:

$$\mathbf{s}^* = \mathbf{s}^d + \widehat{\mathbf{L}}_s \cdot \widehat{\mathbf{L}}_x^{-1} \cdot \mathbf{K}^{-1}(\mathbf{f}^* - \mathbf{f}) \quad (4)$$

where \mathbf{f}^* is the desired wrench, added as input to the control loop (null in this particular case), \mathbf{K} is the environment stiffness matrix, and \mathbf{s}^* is the modified reference for visual features. $\widehat{\mathbf{L}}_x$ relates τ_E and $\dot{\mathbf{X}}_E$ according to $\dot{\mathbf{X}}_E = \widehat{\mathbf{L}}_x \cdot \tau_E$ [10]. Then, the visual servoing control law, described in the previous point, takes as visual reference the new computed reference, \mathbf{s}^* . Unlike most of existing approaches, our approach for vision-force coupling does the coupling in sensor-space, which allows to control vision and force on all the degrees of freedom, whereas only the vision control law is directly connected to the robot, thus avoiding local minima [2].

In conclusion, there are two simultaneous end-effector motions: one, computed by equation 1, which is in charge of performing the task motion, and another one, computed by equation 3, in charge of continuously aligning the hand with the handle by external vision-force control. Figure 3 shows the kinematic screw, computed by equation 3 from the error between the desired (${}^H\mathbf{M}_G$) and the current (${}^H\mathbf{M}_G$) relative pose of the hand and the grasp frame during reaching the handle. For more experimental results of the vision/force-guided door opening task, along with a detailed analysis and a demonstration video, please refer to [11].

IV. EXAMPLE II: OPENING A SLIDING DOOR THROUGH VISION-FORCE COMBINATION

This experiment is very similar to the previous one in the sense that two complementary sensors (vision and force) are used for a door opening task. However, in this case, we deal with a sliding door and a more complex robot: a mobile manipulator composed of a PA-10 arm, endowed with a three-fingered Barrett Hand, and mounted on an ActivMedia PowerBot mobile robot (the UJI Service Robot).

A. Planning the task, hand and grasp frame

The structural model of the new door is shown in Figure 4. The only difference with the previous case is that the task is now specified as a negative velocity along X axis of frame O . However, as before, we choose to specify it as a positive translation velocity along Z axis of the frame G , so that we can set ${}^G\mathbf{M}_T = \mathbf{I}_{4 \times 4}$, without the need to know the door geometric model.

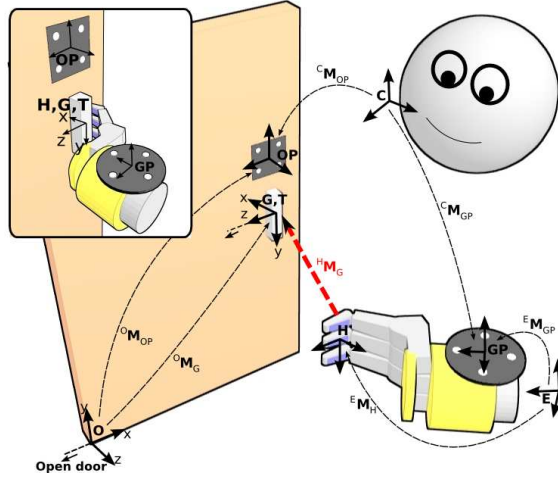


Fig. 4. Specification of the sliding door opening task with the proposed formalism (top), and the mobile manipulator at Jaume-I University opening a sliding door through force and vision combination (bottom).

The Barrett Hand offers more advanced capabilities than the parallel jaw gripper. A task-oriented grasp planner [12] selects a hook preshape as the more suitable hand configuration for the intended task, and the hand frame is set to the inner part of the fingertips, as shown in Figure 4.

The grasp frame is also set by the task-oriented grasp planner to the right part of the handle. Then, according to the specification of the hand and grasp frames, the desired relationship between both is ${}^H\mathbf{M}_G = \mathbf{I}_{4 \times 4}$, which means that the robot has to use the fingertips to make contact with the right face of the handle.

B. Task Execution

Once the task has been specified, it is performed by the same methods explained in the previous experiment, supporting the claim that the robot is not specifically programmed for one particular task. Instead, the same algorithms are applied, but its execution depends on the task specification and the multisensor information that the robot receives during execution. As in the previous example, an external camera tracks the robot hand and the object simultaneously and a position-based visual servoing control is performed

in order to reach and keep the desired relative hand-object configuration, ${}^H\mathbf{M}_G$. At the same time, force control is used for dealing with vision errors and ensuring a successful execution of the task, even in the presence of uncertainties and errors.

V. EXAMPLE III: GRASPING A BOOK THROUGH FORCE-TACTILE COMBINATION

Now, the sensor-based compliant physical interaction framework is applied to the task of taking out a book from a bookshelf, using the UJI Service Robot and force-tactile combination. The goal of the task is to extract a book from a shelf, while standing among other books. The approach is to do it as humans do: only one of the fingers is used, which is placed on the top of the target book and is used to make contact and pull back the book, making it turn with respect to the base, as shown in Figure 6. In this task, the force/torque sensor is used to apply a force towards the book and avoid sliding, whereas a tactile array provides detailed information about the contact, and helps estimating the hand and grasp frame relationship. As shown in Figure 5, there is one tactile array on each of the fingertips. This sensor consists of an array of 8×5 cells, each of one can measure the local pressure at that point.

A. Planning the task, hand and grasp frame

In Figure 5, a representation of the book grasping task, including the necessary frames, is shown. There are two possibilities for the task frame in this case. The first is to set it to the book base (frame T' in Figure 5), so that the task is described as a rotation velocity around this frame. The second possibility is to set the task frame on the top edge of the book (frame T in Figure 5), so that the task is described as a negative translational velocity along X direction. We have opted for the second solution, because, in this case, the task frame coincides with the grasp frame, and, then, there is no need to know the book model. In the first case, the height of the book should be known in order to transform the task velocity from the task frame to the hand frame. By adopting the second solution, we make the approach general for any book size. Two references are set in the task frame, \mathbf{v}^* and \mathbf{f}^* . The first one is set to a negative velocity in X axis, in order to perform the task motion, whereas \mathbf{f}^* is set to a force along Z axis. This force is needed in order to make enough pressure on the book surface and avoid slip. We have set it to 10 N for our particular system, but it depends on the friction coefficient between the fingertip and the book. For small friction, a bigger force would be needed. Therefore, \mathbf{S}_f is set to $\mathbf{diag}(0, 0, 1, 0, 0, 0)$.

For this task, we define a special hand posture where one of the fingers is slightly more closed than the other ones, so that we can easily make contact on the top of the book with one finger, as shown in Figure 5. The hand frame is set to the inner part of the middle finger fingertip, just in the centre of the tactile sensor. The hand frame pose with respect to the robot end-effector, ${}^E\mathbf{M}_H$, is computed from hand kinematics.

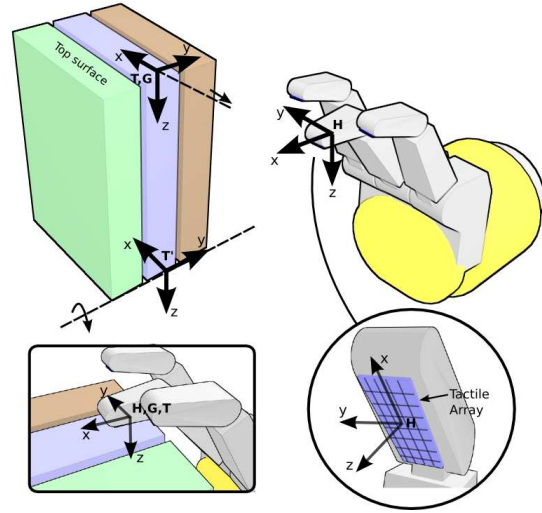


Fig. 5. Frames involved in the book grasping task. The tactile array is used to estimate the relationship between the hand and the grasp frame, ${}^H\mathbf{M}_G$.

The fingertip has to make contact on the top of the book. Therefore, we set the grasp frame to the book top surface, which could be located by vision or range sensors. The desired relationship between the hand and the grasp frame, ${}^H\mathbf{M}_G$, is set to the identity.

B. Task execution

In this case, the task is performed by combining force and tactile feedback. Tactile information is used to estimate and improve the contact between the hand and the book, whereas force feedback is used in order to cope with uncertainties and ensure that a suitable force is performed on the book surface so that there is no sliding.

1) *Estimating hand-book relative pose:* Contact on the book is performed with the tactile array. Depending on the sensor cells that are activated, the relative pose between the sensor surface and the book can be estimated. It is not possible to compute the complete relative pose only with tactile sensors, because they only provide local information when there is contact. However, we can obtain a qualitative description of the relative pose. For example, if there is contact with the upper part of the sensor, but not with the lower part, we can deduce that the sensor plane is rotated around Y axis with respect to the book top plane.

All the tactile cells lie in the XY plane of the hand frame. We consider that the finger is completely aligned with the book surface when there are cells activated on each of the four XY quadrants of the hand frame, i.e., all the tactile sensor surface is in contact. If there is contact on the upper half of the sensor, but not on the lower half, or vice versa, we consider that there is a rotation about Y axis, between the sensor (hand frame) and the book surface (grasp frame). Similarly, a rotation around X axis can be detected.

2) *Improving the grasp:* The goal of this process is to align the finger (tactile sensor) surface with the book surface, taking as input the qualitative description of the relative pose, described in the previous point. We follow a reactive

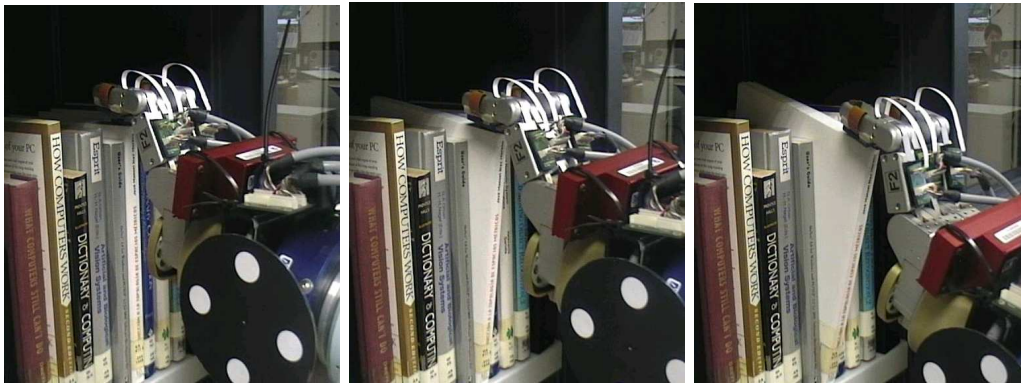


Fig. 6. The robot grasping the book by means of force and tactile-based continuous estimation of hand-to-object relative pose.

approach, where fingertip rotation around X and Y axis of the hand frame is continuously controlled, in order to obtain contact on each of the XY quadrants of the hand frame. With this approach, the behaviour of the robot is completely reactive to the tactile sensor readings. The goal is to keep the sensor plane always parallel to the book top plane, thus ensuring that ${}^H\widehat{M}_G = \mathbf{I}_{4 \times 4}$.

3) *Task motion and coping with uncertainties:* According to the task description, the task motion is performed by moving the hand along negative X axis of the task frame, while applying a force along Z axis. This motion makes the book turn with respect to the base, as shown in Figure 6. Note that, as the fingertip moves backwards and the book turns, the tactile sensor may lose contact with the lower part. This situation is detected by the qualitative pose estimator, and corrected with the control strategy described in the previous point, so that the hand frame is always aligned with the grasp frame, ensuring that task motion can successfully be transformed to end-effector coordinates by equation 1. Figure 6 shows a sequence of the robot performing the task.

VI. CONCLUSION

We have shown three different examples of robotic execution of everyday chores, built on top of a new vision-force controller [2] and a general framework for specifying multisensor compliant physical interaction tasks, presented on a separate paper [1]. Two door-opening tasks with different robotic systems and a book grasping task have been implemented making use of external vision-force control and force-tactile integration. The three examples exhibit a reasonable degree of robustness, in the sense that the use of force feedback allows to deal with uncertainties and errors. External vision-force control allows to avoid local minima which is one of the main drawbacks of the existing approaches. The implementation of these examples in very different robotic systems during a short period of time shows the suitability of the framework for versatile specification of disparate multisensor physical interaction tasks.

As future research, we would like to use the proposed framework for the specification and compliant execution of several common tasks in home environments, based on integrated visual, tactile and force feedback. We think that the

integration of multiple and disparate sensor information for hand-to-object pose estimation is a key point for successful and dependable robotic physical interaction.

ACKNOWLEDGMENT

The authors would like to thank Universitat Jaume-I/Bancaja, under project PI1B2005-28, and Generalitat Valenciana, under projects CTBPRB/2005/052 and GV-2007-109 for their invaluable support in this research.

REFERENCES

- [1] M. Prats, P.J. Sanz, and A.P. del Pobil. A framework for compliant physical interaction based on multisensor information. In *IEEE Int. Conf. on Multisensor Fusion and Integration for Intelligent Systems*, Seoul, Korea, 2008. [Submitted].
- [2] Y. Mezouar, M. Prats, and P. Martinet. External hybrid vision/force control. In *Intl. Conference on Advanced Robotics (ICAR'07)*, Jeju, Korea, 2007.
- [3] R. Bajcsy. *Integrating vision and touch for robotic applications*. Trends and Applications of AI in Business, 1984.
- [4] Lars Petersson, David Austin, and Danica Kragic. High-level control of a mobile manipulator for door opening. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2333–2338, vol 3, Takamatsu, Japan, October 2000.
- [5] C. Ott, C. Borst, U. Hillenbrand, B. Brunner, B. Buml, and G. Hirzinger. The robutler: Towards service robots for the human environment. In *Video Proc. International Conference on Robotics and Automation*, Barcelona, Spain, 2005.
- [6] U. Hillenbrand, B. Brunner, Ch. Borst, and G. Hirzinger. The robutler: a vision-controlled hand-arm system for manipulating bottles and glasses. In *35th International Symposium on Robotics*, Paris, France, 2004.
- [7] H. Bruyninckx and J. De Schutter. Specification of force-controlled actions in the 'task frame formalism': A synthesis. *IEEE Trans. on Robotics and Automation*, 12(5):581–589, 1996.
- [8] S. Lee et al. Robust recognition and pose estimation of 3d objects based on evidence fusion in a sequence of images. In *IEEE Int. Conf. on Robotics and Automation*, Rome, Italy, 2007.
- [9] E. Marchand and F. Chaumette. Virtual visual servoing: a framework for real-time augmented reality. In *EUROGRAPHICS 2002*, volume 21(3), pages 289–298, Saarebrücken, Germany, 2002.
- [10] P. Martinet and J. Gallice. Position based visual servoing using a nonlinear approach. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, volume 1, pages 531–536, Kyongju, Korea, 1999.
- [11] M. Prats, P. Martinet, A.P. del Pobil, and S. Lee. Vision/force control in task-oriented grasping and manipulation. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, San Diego, USA, 2007.
- [12] M. Prats, A.P. del Pobil, and P.J. Sanz. Task-oriented grasping using hand preshapes and task frames. In *Proc. of IEEE International Conference on Robotics and Automation*, pages 1794–1799, Rome, Italy, April 2007.