Noname manuscript No.
(will be inserted by the editor)

A Tangible Interface for Transferring Skills

Using Perception and Projection Capabilities in Human-Robot Collaboration Tasks

Davide De Tommaso · Sylvain Calinon · Darwin G. Caldwell

Received: date / Accepted: date

Abstract Our research focuses on exploring new modalities to make robots acquire skills in a fast and user-friendly manner. In this work we present a novel active interface with perception and projection capabilities for simplifying the skill transfer process. The interface allows humans and robots to interact with each other in the same environment, with respect to visual feedback. During the learning process, the real workspace is used as a tangible interface for helping the user to better understand what the robot has learned up to then, to display information about the task or to get feedback and guidance. Thus, the user is able to incrementally visualize and assess the learner's state and, at the same time, focus on the skill transfer without disrupting the continuity of the teaching interaction. We also propose a proof-of-concept, as a core element of the architecture, based on an experimental setting where a pico-projector and an rgb-depth sensor are mounted onto the end-effector of a 7-DOF robotic arm.

Keywords Human-Robot Interaction · Learning by Demonstration · Tangible Interfaces · Augmented Reality

Fig. 1 Conceptual illustration where a robot can actively perceive objects and project information about a task progressively learned with the help of the user.

1 Introduction

While accuracy and speed have, for a long time, been at the top of the agenda for robot design and control, the development of new actuators and control architectures is now bringing a new focus on passive and active compliance, energy optimization, human-robot collaboration, easy-to-use interfaces and safety. The considerable growth of the number of service robots has brought machines closer to the human, involving aspects of daily life. The cooperation between robots and people without technical skills is becoming even more common in different fields and applications. Therefore, the classic methods for interfacing with the robot do not satisfy the new requirements of the modern world in which the final user should not need to be an expert-programmer to use the interface. Instead of stand-alone programming, dynamic bidirectional models of interaction are required, in which the robot (learner) actively acquires the task demonstrated by the user (teacher).

Our recent studies have specifically addressed the issue of finding new user-friendly physical interfaces in order to reduce the complexity gap between humans

Davide De Tommaso · Sylvain Calinon · Darwin G. Caldwell
Department of Advanced Robotics, Istituto Italiano di Tecnologia, via Morego, 30, 16163 Genova E-mail: {davide.detommaso, sylvain.calinon, darwin.caldwell}@iit.it

and machines and to speed-up the skill transfer. The aim of this paper is to propose a novel architecture in Learning from Demonstration (LfD) scenarios, based on a new configuration of both active perception and projection technologies, allowing the system to establish a bidirectional communication channel for skill transfer. Fig. 1 illustrates the concept of the proposed approach. The proposed interface is designed to provide a visual augmented operating space shared between the learner and the teacher, as shown in the schematic in Fig. 2. The aim is to share a common understanding of the task needing to be transferred, by using the operating space as a tangible interface where the task features are graphically superimposed in an augmented reality fashion. Consequently, the teacher can understand what the robot is learning, by observing the surrounding environment, and eventually refine or rectify on-line the skill whenever a robot mistake occurs. This adaptive learning process enables the user to always be aware of the learner's state and to continue the task without interrupting the teaching phase.

In order to exploit the perceptual and projective features of the proposed system, our experimental setup jointly adopts an rgb-depth sensor and a pico-projector, both mounted onto the end-effector of a robot arm. Adopting such a mobile configuration, instead of a fixed setup, leads to a number of key advantages, such as: a) an extended field of view due to the different angle views reachable by the robotic arm, b) the possibility to actively handling occlusions and facilitating tracking of task-relevant features and also c) the adaptive multi-resolution for perception and projection features.

For demonstrating the feasibility of the proposed architecture, the development of the core framework with an experimental setup are also shown. The goal of the experiment, is to make the robot project an image (depicting a user interface) onto a planar surface in a desired position by following the user's guidance. It consists of two steps. In the first phase, the user's body is detected while he/she is pointing towards the desired plane in which the projected interface has to be shown. In the second phase, the image is projected at the basis of the pointed region taking into account the user's head position, to guarantee that the interface is oriented towards him. This setup contains the core features to accomplish the next scenarios that we envisaged for future work involving the proposed architecture. Since no robot learning is implemented at this stage of the development of the interface, in order to give a more practical view of the paper, we present possible scenarios where the system can be applied, by describing what kind of interaction and information is related to a particular skill transfer.

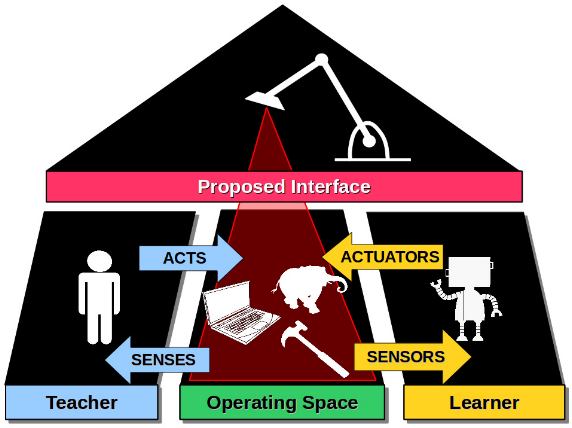


Fig. 2 Architecture overview of a typical learning scenario, in which the Robot/Learner and the User/Teacher interact with each other in the same Operating Space.

The remainder of the paper is organized as follows. In Section 2, we focus on the role of user interfaces in the learning contexts and we outline our motivation in designing a robot interface in the context of robot LfD. In Section 3, we review recent works in Human-Computer Interaction (HCI), Human-Robot Interaction (HRI) and robot learning that go along this line of research. We also give an overview of techniques such as Tangible Interfaces, Mixed Reality and Spatial Augmented Reality by focusing on the learning aspect. The details on the architecture and the technology adopted, as well as the conducted experiment is described in Sections 4 and 5. In Section 6, we highlight possible scenarios where the proposed interface can be employed. Possible extensions to some existing works are also presented. Lastly, conclusions and future works are presented in Section 7.

2 Motivations

For learning processes that require natural human interaction for transferring skills to robots, the design and development of interfaces between the teacher and the learner play a key role. In LfD strategies, expertise in robotics should not be expected of the final user. This makes it necessary to develop a shared communication protocol for transferring skills from humans to machines. Although several studies have investigated the social and technological aspects of the Human-Robot Interaction, many issues remain largely unexplored.

From the social point of view, transferring skills is more complex than a one-way step process, as the learner usually requires feedback from the instructor after the demonstrations and reproduction attempts. A good instructor maintains a mental model of the learner's state in order to support the learner's need and use different teaching methods such as demonstrations, verbal instructions, attention cues and gestures [16]. Alternatively, the teacher may provide additional demonstrations useful for task generalization.

We emphasize in in [4] emphasize the importance of providing an active role to the human-teacher in the learning process. The effectiveness and the generalization of the acquired skill do not depend only on the number of demonstrations but mostly on the pedagogical quality of these. The way to transfer a skill may be affected by the different nature of the learner and the teacher involved in the interaction and by several psychological factors related to the user during the teaching process. In the Human-Robot Interaction learning process, the way of giving several demonstrations of the task and the way of refining the learned skill by observing new reproduction attempts are often considered separately. We propose in [4] a learning paradigm to allow the user-teacher to incrementally see the results of demonstrations. This attempts to establish a two-ways interaction during the teaching process and to allow the user to feel involved in the task acquisition process.

In the past decades, research studies have focused on the critical task of finding an approach to enable communication between humans and robots by common natural methods that people use when teaching each other. Natural human-robot interaction means can extend the use of robots to cooperative tasks. In [41], Thomaz and Breazeal highlight the fact that in an efficient teacher-learner interaction, the two partners cooperate to simplify each other's role. While the teacher maintains a mental model of the learner's current understanding of the task to appropriately support the learner's needs, the learner helps the teacher by making the skill acquisition process transparent (e.g., by communicating its internal state, by revealing the level of knowledge, or by demonstrating the current mastery of the skill). In other terms, the teacher and learner cooperate with reciprocal roles: the user guides the interaction through a scaffolding process including feedback, structuring successive experiences, regulating the complexity of information, and guiding the learner's exploration, while the robot helps the user by improving the learning experience through transparency [41]. Indeed, natural methods for teaching offer an increasing range of possible interactions that come along with the development of new interfacing capabilities.

Humans and Robots should be able to communicate with a shared protocol, to share a common description of the task and to give/receive feedback through common cues, made accessible by both partners in an appropriate visualization format.

The above considerations have been the starting point of our research, by exploring novel methods that encapsulate hardware and software components in an interface designed for efficient interaction in robot skill acquisition.

3 Related Work

In order to automatically design a robot controller, a robot learning algorithm is required for reproducing a specific task, demonstrated by humans, and for generalizing it in new conditions. In the learning phase, the task model is acquired by decoding inputs from the robot's perception senses and by finding relationships with a set of relevant task features. Then, in the reproduction phase, the robot needs to make decisions and generalize the acquired skill to new situations by transforming the learned task to appropriate actions.

3.1 Learning from Demonstration (LfD)

In the last thirty years, different learning techniques for robotics applications have been studied. Learning from Demonstration (LfD), Programming-by-Demonstration (PbD), or Imitation Learning, study how the human demonstrations can enable the robot to reproduce the underlying task. This skill transfer process, apparently easy for humans, hides different key problems. One research question is how humans should provide the demonstrations to the robot to model a new skill, e.g. by manual teleoperation [38,23], kinesthetic teaching [20, 5], vision [25,31], gestures [42] or natural language [26]. Another important research question is to find ways for the robot to extract the task constraints and generalizing in different contexts [29,30,11].

LfD was born to meet the requirements of manufacturing robotics for reducing or eliminating the heavy work of programmers who had to develop software each time a new robot skill was required. Instead, LfD takes the perspective that skills can be transferred by imitation, and that once a skill has been correctly acquired, the robot can adapt the learned task to different situations. Therefore, further information (input task parameters) from the teacher are often needed to accomplish the task. Input task parameters can be provided: (1) directly by a single teacher in a multi-robot context [7], by requesting additional information through a query-based approach [14]; (2) indirectly through screen-based interfaces or using Virtual Reality [2]; or (3) automatically by using machine learning techniques to extract the important features characterizing the skill [9].

3.2 Tangible Interfaces for Learning Applications

The proposed work aims at achieving an effective and fast skill transfer between human and robot in an incremental learning context. To accomplish these requirements, the teacher should access the learner's state directly on the work scene during the task's demonstrations. For this reason, we have drawn inspiration from previous works on *Tangible User Interfaces* (TUIs) to design the system architecture. A TUI establishes a bidirectional channel of embodied interactions and perceptions between the user and the environment in which the physical interactions change its digital state [19]. While a Graphical User Interface (GUI) provides the tools to access digital information by traditional input/output devices (e.g. mouse, keyboard, display), a TUI instead provides a close coupling between real and virtual objects or landmarks. A practical example could be a tangible-PC desktop represented by a real desk, in which each real object has a digital correspondent, like a cube for an icon or a basket for the trash. In such a way, physical objects/surfaces can be used to visualize and interact with computer-generated information. This two-ways mapping between real and digital world enables the user to achieve the desired effect without interacting through computer interfaces, but directly acting on a real object semantically related to such effect. In the previous example, the real action of throwing a cube in the basket could digitally correspond to the deletion of its corresponding element from the PC.

The social and cognitive implications resulting from the use of these interfaces led us to apply these interaction paradigms to learning processes. From the cognitive point of view, less effort is required for the user to understand how the system works, because a tangible scenario can involve natural and intuitive interactions [10]. The use of physical materials in a learning context, exploiting haptic senses, could also speed-up the gained task's knowledge [24], rather than using virtual objects in visual interfaces [15]. Indeed, as a consequence of such intuitive interaction, the attention of the user can be focused on the completion of the task. As a consequence, time can be reduced by the rapid understanding of the task space.

3.3 Mixed Reality in Human-Robot Interaction

Augment Reality (AR) enables users to see virtual graphical elements superimposed on real objects. While TUIs allow the user to see graphical elements projected on the surrounding environment, AR applications enable solely the illusion to see virtual object mapped in the real-world. This important distinction is due to the fact that AR need a couple of computing interfaces (camera-display) to show the user a mixed reality.

When Augmented Reality meets Tangible Interfaces, Spatial Augmented Reality turns out. Bimber and Raskar [3] describe the way to visualize information in the task space by using a projective system. The recent technology of hand-held projectors promises a rapid growth of applications, for enabling the user to interface with computers or robots.

In [37] and [40], the authors use pico-projectors to visualize augmented digital information over real objects as HCI interfaces. A lightweight mobile camera projector unit is used in [37] to augment a paper map with additional information. This virtual map, the *Map Torchlight*, is tracked over the paper map and can precisely highlight points of interest, streets, and areas to give directions or other guidance information. In [40], a digital pen embedded with a spatially-aware miniature projector is used to explore the interaction design space of a paper document, providing the user with immediate access to additional information and computational tools. HRI interfaces with the help of hand-held projectors have also been studied, such as a robotic control interface for visualizing manipulation tasks [18], as an interface for controlling the robot without the direct manipulation by the user [21], or an alternative to the anthropomorphic interface using a projected display to interact with the user [34]. We also took inspiration from *LuminAR* [27], a project redefining the original concept of a desk lamp. Combining the technology of robotics and computer science, the authors use the light from a pico-projector, mounted on an articulated robotic arm, to show digital information to the user directly over the desk or any other surface. The joint use of a camera allows the user to interact with this virtual interface through hands motion, such as reading emails or navigating through a website.

Recently, Vogel *et al* [8] tackled the safety issue in Human-Robot collaboration task by using a projector-based solution. The authors propose a spatial augmented reality interface able to establish a physical safety area in a shared workspace between users and robots, by using a camera projector pair. The projective device gives feedback to the user about the safe working area, by projecting virtual barriers directly aligned with the real portion of space. The perception device helps the system to actively monitor the physical state of the user and the robot within the safety area, by changing position, shape and orientation of the projected image dynamically. Another recent work about projective interfaces in the field of wearable computing is *OmniTouch* [17], in which Harrison *et al* suggest an innovative way to access digital information everywhere. *OmniTouch* is a wearable device that enables the user to interact with a GUI projected on any physical surface by using gestures. By exploiting the perception capabilities of the rgb-depth sensor, the system is able to detect suitable surfaces in which to project the GUI, by using a pico-projector, and to interact with fingers like a mouse pointer.

4 Overview of the Proposed Interface

We propose a tangible interface for skill transferring in human-robot collaboration tasks. The system uses a structured light system which combines a pico-projector and an rgb-depth sensor. Instead of using fixed tracking and traditional displays, a robot manipulator is endowed with perception and projection capabilities (see Fig. 3). We consider a compliant lightweight robotic arm as an interface that can move, perceive and act on its environment. Here, the aim consists of providing the user an assistance tool during the skill transferring to facilitate the teaching process by enhancing the robotic tool with augmented reality capabilities. The system is designed to be able of sharing a common understanding of the task between the teacher and the learner, by establishing a bidirectional interaction during the skill transferring. In one direction, the interface provides outputs by projecting information, for visually enriching the environment with virtual features according to the task to transfer. The projective device is thought to highlight objects of interest for the task or to visualize guidance for the user, such as trajectories or landmarks directly in the surrounding work space. While this issue has been studied in the context of mobile robotic platforms, it presents new challenges in the context of manipulation and human-robot collaborative skills learning. In the other direction, the feature of perceiving RGB and depth information, make it able to acquire inputs from the objects in the scene and from human's guidance. State-of-the-art computer vision techniques can be used for human and object detection and tracking, although the paper does not provide new contribution in this direction. We will exploit for this issue the very active development of state-of-the-art algorithms and software (see e.g. Point Cloud Library [36]) enacted by the spread of devices such as the Microsoft Kinect. Such devices can track on-line both color and distance information, and are small enough to be mounted on the robot's end-effector.

The combination of these devices in a mobile robot configuration brings a few key features to light. While a fixed camera system has a static field of view, a robotic setup, such as that proposed, can project/detect at various places and under various angles. In other, the capability of detecting the human body by using the Kinect, would allow the robot to actively handle occlusions. In such case, the user would not need to care of being in the field of view, nor to place the camera and projector in the workspace prior to the experiment. Finally such configuration also offers adaptive multi-resolution tracking and projection features. Indeed, the human detection needs different field of views than the object detection. For example if one wants to project on a large surface (e.g. to have an overview of the objects involved in an assembly task), the robot can move back to increase its field of view. Then, if a more precise information on a particular object is needed (e.g. projecting information on the position of screws and threads), the robot can move closer to the object of interest.

For designing a learning scenario, the complete system takes into account the following entities: the *User* (human teacher), the *Robot* (machine learner) and the *Operating Space* (see Fig. 2). In a typical Robot Learning from Demonstration context, the User/Teacher needs to transfer a particular skill to a Robot/Learner by providing several examples of the task. The place where the interaction is accomplished is the Operating Space, which includes also the other physical parts of the environment perceived by the sensors and reachable by the robot hands.

5 Experimental Setup

In this section we describe the implementation of the core components of the tangible interface, without focusing on a particular learning scenario.

The experiment consists of two phases. The first, that we call *User Pose Detection* (Mode 1), has the purpose of finding the human pose in the Operating Space. In this phase, we move the robot while gravity-compensated, by exploiting its compliant control capabilities (the robot is physically moved as if it had no weight and no motor in its articulations), until a human pose is detected. The second phase, that we call *Adaptive Projection* (Mode 2), consists of projecting a user-interface (represented here as a colored box) on a suitable surface. In order to make the projected information easy to read, the geometric perspective of the surface being used is also considered.

For our experimental setup, instead of a fixed setup, we follow a similar approach to Alenyà *et al* in [1]. They conducted an experiment of plants monitoring in order to extract visual features from leaves for indicating the state of the plant. The authors exploit the use of a color and time-of-flight cameras as perceptual devices rigidly mounted at the end-effector of a robotic arm. The controlled robot enables the development of

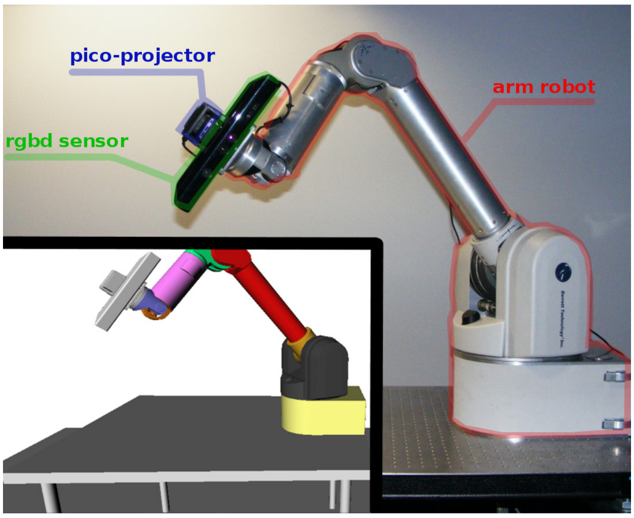
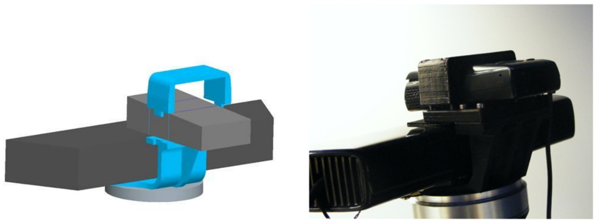


Fig. 3 A laser pico-projector and an rgb-depth sensor are mounted on the Barrett WAM arm robot.

an autonomous procedure to extract a 3D model of the leaves, by mixing the color and the depth information from the different cameras. While Omnitouch [17] and LuminAR [27] are HCIs used to display information in the environment, one of the novelty brought by the plant monitoring application of Alenyà [1] consists of a motorized vision system. These three works are specifically designed for short projecting distances, while one of the proposed features of the proposed interface consists of an adaptive multi-resolution for perception and projection, as previously explained in the Section 4.

5.1 Hardware

As shown in Fig. 3, our experiment is implemented with a 7-DOFs Barrett WAM robot manipulator, a laser AAXA L1V2 pico-projector, and the Microsoft Kinect as rgb-depth sensor. The arm robot is mounted on a wheeled table, that we use as flat surface for projecting information. On the end-effector of the robot, we mounted a rigid tool for holding the Kinect and the pico-projector. Such a support tool has been built with a 3D printer to meet the requirement of low-weight and rigidity. As shown in Fig. 5, the support tool is fixed to the metallic plate of the Barrett WAM. A calibration process is required prior to the experiment, as we discuss in Section 5.2. Since the human has a fundamental role in the interaction that we consider, a reliable algorithm for human tracking is required. Nowadays the Microsoft Kinect presents advantages to accomplish this task in a robust way and at low cost [39]. Even if the Kinect was not specifically designed for human-robot interaction tasks, in the last years, several works have successfully contributed to research in HRI and HCI fields. The Kinect device has brought a huge amount of open-source developments, making it possible to access state-of-the-art computer vision algorithms, such as detection and tracking (see e.g. PCL [36]). The interest shown from the scientific community, mostly in HCI, Computer Vision and Robotics, has significantly increased in a short period of time, due to the fact that it introduced a non expensive solution to replace previously available range sensors. Thanks to the infrared sensors, used by the depth estimation method, the device can be used even in low light conditions. Such feature is very important in our setup because the current technology of pico-projectors requires a usage in low light conditions for better results. Another advantage, that will be exploited in the next works, is the possibility to collect 3D mesh data of the surrounding environment. The current state-of-the-art pico-projector we have selected has a small size (4.2” x 2.1” x 1.2”, for a weight of 170 grams). An important hardware feature that is relevant in our application is the use of the laser technology, which enables auto-focus. In these conditions, it is possible to project information on surfaces at any distance, without taking into account focal adjustment.



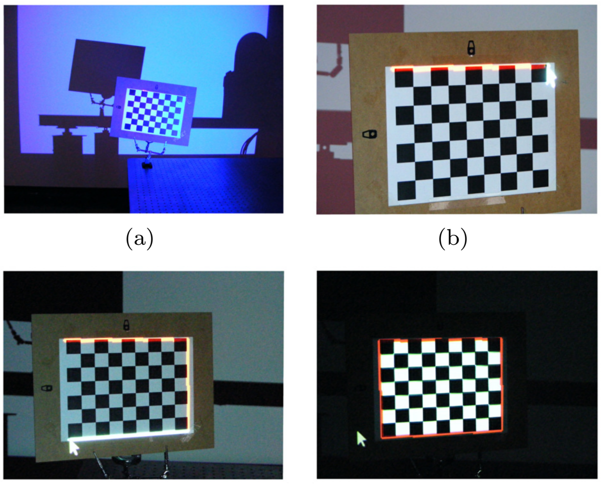
(a) (b)

Fig. 4 The support tool mounted on the end-effector plate of the Barrett WAM. The CAD model in (a), and the real support in (b).

5.2 Calibration

5.2.1 Kinect Calibration

The Microsoft Kinect consists of a structured light system which uses dots projected patterns to estimate 3D depth information about the scene. It is endowed with a RGB standard camera, an infrared (IR) sensor and an IR projector. The projector illuminates the scene with a spotted light pattern detected by the IR sensor. Depth information for each point is directly preprocessed by the device through stereo triangulation. A calibration process is required to reduce the inaccuracy about the depth estimation and the correspon-



(a) (b) (c) (d)

Fig. 5 (a) The printed chessboard is inside the frustum of both the IR sensor and the pico-projector. (b) and (c) The four points are manually selected by the projected mouse pointer. (d) Once the quadrilateral is defined, a projected chessboard is projected over the printed one by using homography transformation.

dence errors between RGB and IR sensors. For estimating the extrinsic and intrinsic parameters of the sensors, we used a stereo based calibration procedure [32]. A stereo calibration takes as input several couples of images from two cameras, in which a planar chessboard is tracked/visualized by the two devices [12]. As output of this process, intrinsic parameters for each camera or sensor and geometric transformation (extrinsic parameters) between them can be estimated.

5.2.2 Kinect-Projector Calibration and Mapping

A Kinect output is the *Depth Map*, in the form of a 640×480 matrix of grayscale values, where pixels encode depth information. Thus, the depth value z can be decoded for each pixel (x, y) of such matrix, therefore a 3D point (x, y, z) in the IR frame is defined. Finally, the correspondent RGB value of each 3D point can be performed by the calibration parameters already estimated. The same approach used for the Kinect's calibration is exploited for calibrating the pico-projector. Instead of considering the previous procedure for calibrating the IR and the RGB sensor, we replace the set of RGB images with a set of image coming from the pico-projector. Since a projector does not get images from its sensor, we project the same chessboard pattern superimposed to the printed chessboard. The resulting effect is to see a projected chessboard aligned with the printed one (see Fig. 5(d)). The process requires the printed chessboard to be inside the frustum of both the IR sensor and the pico-projector (see Fig. 5(a)). The way we collect images from the pico-projector calibration consists of two steps. Firstly, we need to select 4 points which correspond to the 4 corners of the printed chessboard. We can achieve this easily by projecting a mouse cursor and selecting manually the correct points on the real chessboard (see Fig. 5(b) and 5(c)). For each corner selected we keep track of the correspondence pixel in the projected image, by obtaining as final result a quadrilateral. We then warp the original chessboard image (same pattern as the printed one) to the quadrilateral through homography. We obtain as result a virtual chessboard with the same perspective as the real one. For each chessboard pose, we collect both IR images and the correspondent images to project similarly as for RGB images. Then, by giving this information as input of the calibration algorithm used for the Kinect, we also obtain the intrinsic parameters of the pico-projector and the extrinsic parameters between the IR sensor and the pico-projector. Thus, in our experiment, the pico-projector sensor is calibrated with the IR sensor of the Kinect. Therefore, we can estimate a 3D correspondence between the two sensors by the relation

$$X_p = T_{ir2proj} X_{ir}, \tag{1}$$

by using the transformation matrix $T_{ir2proj}$ defined with the extrinsic parameters. Here, the aim is to project light to a desired 3D world point. If X_{ir} is a 3D point in the IR reference, we are looking for the corresponding (u_p, v_p) pixel coordinates of the image to project, i.e.

$$\begin{pmatrix} u_p \\ v_p \end{pmatrix} = \begin{pmatrix} \frac{X_{proj_x} f x_p}{X_{proj_z}} + p x_p \\ \\ \frac{X_{proj_y} f y_p}{X_{proj_z}} + p y_p \end{pmatrix}, X_p = \begin{pmatrix} X_{proj_x} \\ X_{proj_y} \\ X_{proj_z} \end{pmatrix}. \tag{2}$$

The terms fx_p, fy_p, px_p, py_p are referred to the intrinsic parameters of the pico-projector.

5.2.3 End-Effector Tool Calibration

The fixed frame of reference of the Operating Space is defined by the robot's base. The position and orientation of the end-effector, are computed by direct kinematics, while we need the transformation between the IR sensor of the Kinect and the End-Effector plate. To do so, we collect a series of corresponding 3D points from the two different frames of reference, (IR sensor and End-Effector), and estimate the geometric transformation, by using the Procrustes method [22], which estimates the rotation, scaling and offset between two sets of data points.

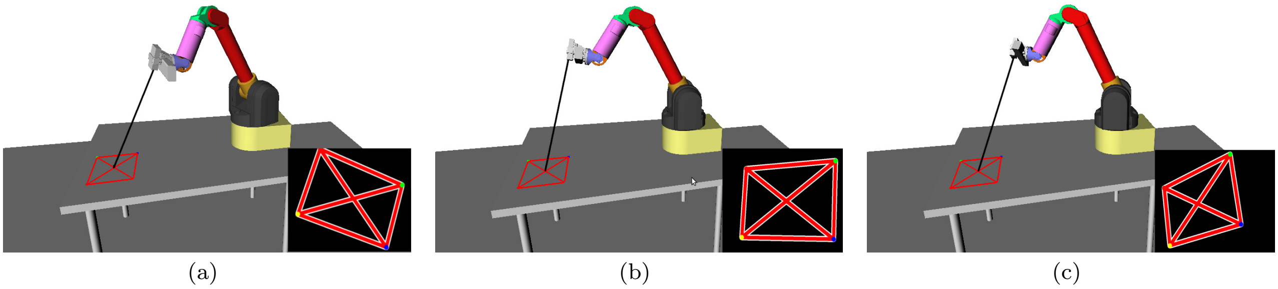


Fig. 6 In (a), (b) and (c) three different robot's poses are shown with their respective projected images in the right corners.

5.3 Robot Control

The robot is controlled by inverse dynamics solved with a recursive Newton-Euler formulation [13]. A gravity compensation force is added to each link, and a wrench command in the Cartesian space is used to keep the orientation of the end-effector towards the center of the projection plane. The joint forces f_i at each joint $i \in \{1, \ldots, n\}$ are calculated as

$$f_i = f_i^a - f_i^e + \sum_{j \in c(i)} f_j,$$

where f_i^a is the next force acting on link i, f_j with $j \in c(i)$ are the forces transmitted by the child $c(i)$ of link i, and f_i^e are the external forces defined as

$$f_i^e = F_T + F_G.$$

In the above equation, $F_T = [0, M_T]^\top \in \mathbb{R}^6$ is the vector of force and momentum requested to accomplish the task (only applied at the end-effector, i.e. when $i = n$), and $F_G = [f_G, 0]^\top \in \mathbb{R}^6$ is the gravity compensation force. Tracking of the desired orientation in Cartesian space is insured by a resolved-acceleration control scheme [28].

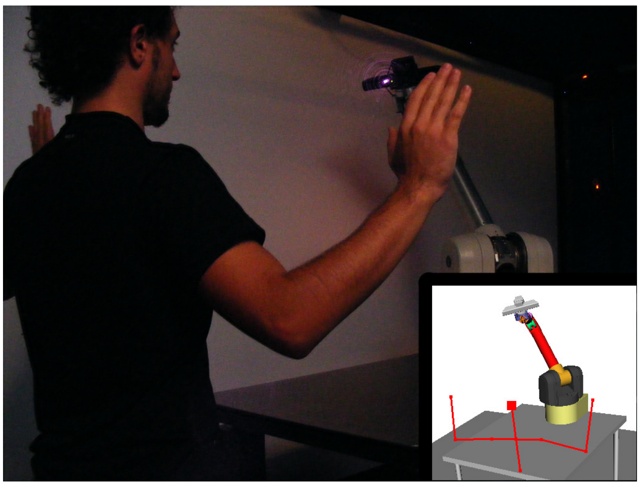


Fig. 7 Control Mode 1: The system detects the pose of the user.

5.3.1 Control Mode 1: User Pose Detection

The aim of this interaction phase is to detect the user's pose in the frame of the robot's base (see Fig. 7). The tracking and detection of the user's upper body is based on the state-of-the-art proprietary OpenNI framework and NITE middleware [33,35]. The Cartesian coordinates of each body part represent the position and orientation of the human skeleton. The projecting space is the planar table in which the robot is fixed. In the current implementation, by using the gravity compensation control capabilities, the robot is manipulated by a human assistant for the user localization.

5.3.2 Control Mode 2: Adaptive Projection

Once the data about the human pose are correctly acquired, an *Adaptive Projection* method can be applied to localize the position and orientation of the projected image. We call *Adaptive Projection* the process that combines a control policy of the robot and capabilities of perception and projection for adapting projected information on surfaces that are comfortable for the user, by using both perspective transformations and robot control. The center of the projected image is the point resulting from the intersection of the vector d_1 and the surface of interest (here the table, see Fig. 8). The d_1 vector defines the direction from the elbow and the hand, while d_2 defines the direction between the user's head and the resulting center mentioned above. The position of the projected image is estimated by the vector d_1, while its rotation is performed by d_2 in order to make the interface readable by the user (see Fig. 9). As shown in Fig. 5(a), if the quadrilateral which defines the projecting area is fully or partially inside the frustum of the pico-projector, the image is correctly aligned with the chosen perspective. The robot in this mode continuously modifies the orientation of the end-effector such

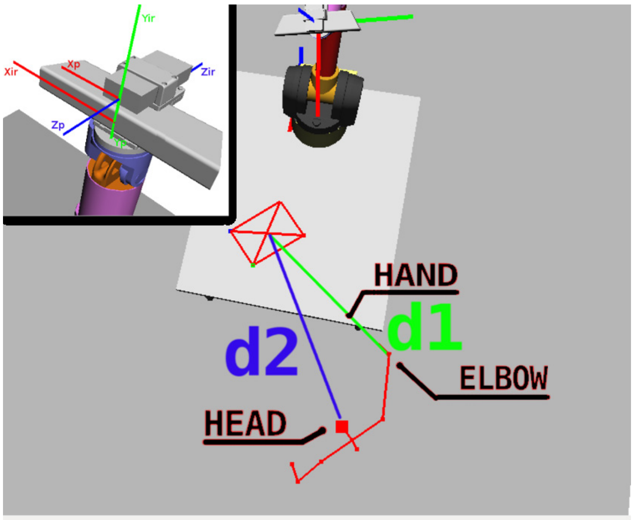


Fig. 8 Estimation of position and orientation of the projected image. In the upper left corner are shown the two frame of reference of IR sensor and the pico-projector.

that it is maintains its orientation towards the center of the projected image. As it is shown in Fig. 6, despite the robot moves around the space, the end-effector points towards the projecting surface, while the projector actively changes the geometric transformation to follow the required perspective. To demonstrate this capability, the robot is either moved manually by the user, while being gravity-compensated, or it can move autonomously by following a desired predefined path.

6 Illustrative Scenarios

In this section we present some prototype scenarios as future applications of the proposed system. Our aim is to highlight the potential of the interface and the classes of possible features, by suggesting some possible ways of implementing the different scenarios, but by remaining loose on the technical aspects involved by these challenges.

6.1 The kitchen assistant

Usually, during the preparation of a recipe (see Fig 10), the user is fully involved in the task using both hands and eyes. He might need to go around the kitchen to accomplish different subtasks, such as moving from the work table to the oven or from the fridge to the stove. According to the features of the human tracking and the control policies described in Section 5.3, we would like the robot is able follow the user and find automatically appropriate planes to project useful information for completing the recipe. Several issues need to be considered to accomplish this task. First, in order to take into account the safety of the user, machine learning techniques can be exploited for avoiding the perceived human body (e.g. in [6]). The task of finding the best projecting surfaces, on the basis of the user poses, can be learned from the system as a result of a set of training with a decreasing level of user's guidance. Third, once the user selects a cake's recipe, he should show to the vision system each ingredient separately to let it register the relevant visual features to make it able, during the interaction, to recognize the objects from the whole 3D scene. The user can point his arm towards the corresponding object on the table, whenever the system requests information about one of the ingredient of the recipe. Then, the system should be able to establish a relationship between a segmented point cloud and its semantic meaning in the task, by setting a label (e.g. "butter", "eggs") for each of those. As shown in Fig. 10 the system should detect objects involved in the *Operating Space* and overlays visual recipe's information (such as to indicate the required amount of butter or milk). In order to accomplish the issue of finding the user pointed objects or surfaces, geometrical intersection between arm's direction and segmented 3D point cloud can be performed. The 3D point cloud libraries available for the Kinect, like PCL [36], could for example be used for helping to solve the tracking and detection challenges by using state-of-the-art algorithms (e.g. feature estimation, surface reconstruction and segmentation). Such example shows how the system's perception could be used here to establish an input channel that allows the user to select specific areas of the *Op-*

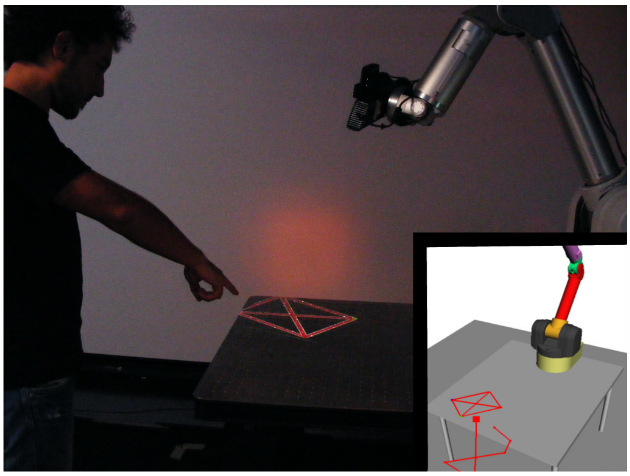


Fig. 9 Once the user pose is detected, the relative position and orientation of the projected interface can be determined. By pointing towards the selected surface area, the user can visualize the projected interface automatically oriented towards him.

Fig. 10 Illustrative scenario of a robot assistant in the kitchen. Cooking often requires the use of both hands. With hands in flour, the user could not use a standard external interface to easily follow the recipe while doing the task. Here, a possible implementation could be to project information on the required amount of milk and butter and the list of ingredients without interrupting the task (namely, by having the text oriented towards his sight).

erating Space to indicate directions or to describe trajectories simply using his arm or hand like a pointing devices. Moreover, the projector could be used to superimpose task information, directly on the *Operating Space*, for providing the user with useful information about the recipe to prepare and the relevance of the objects for the task.

6.2 The *pizzaiolo* robot: an interactive visualization of learned movement as 3D flow fields

A critical issue in Learning from Demonstration is to visualize the skill that the robot has learned in the robot's environment, prior to executing it on the real robot for safety reasons. Virtual reality techniques and robot simulators have the drawback that the whole robot's environment, namely the objects involved in the interactions and the robot itself, need to be modeled. Accuracy errors of the synthesized model might introduce discrepancy between real and simulated movements. Physics and dynamics of the system also have to be taken into account to develop the simulator, which is sometimes difficult to develop. Even if powerful frameworks and tools are nowadays available to simulate robot environments, physical GUI user interactions are still required. Indeed, several operations such as: zoom-in/out, view point changes, removing occluding objects, are performed by the user through input devices (mouse, keyboard, touch-screen) causing interruptions during the teaching process.

Our work in LfD takes the perspective that the development of compliant actuators will bring gradual changes in the way skills and motions are represented by the algorithms in LfD. The machine learning tools that have been developed for precise reproduction of reference trajectories need to be re-thought and adapted to these new challenges. For planning, storing, controlling, predicting or re-using motion data, the encoding of a robot skill goes beyond its representation as a single reference trajectory that needs to be tracked or set of points that needs to be reached. Instead, other sources of information need to be considered, such as the local variation and correlation in the movement.

One perspective of our work is that a skill acquired from human demonstrations can be transferable to a robot by representing it as a combination of local flow fields. Such representation aims at generalizing the movement by computing control commands that are valid within local regions describing the current state of the skill. The robot can thus continue its movement without replanning, even if it is faced with continuous sources of perturbation. Compared to a standard reference trajectory, the representation as vector field also brings new challenges in the way the movement acquired by the robot can be visualized and analyzed by the user (if the user prefers to visualize it before running it on the real robot).

The proposed system could be used in this context to project trajectories or flow fields by using planar surfaces in the environment. In this way, the user can select the surface of interest and move around the robot to see different views while keeping his/her gaze towards the robot's workspace. The following scenario takes inspiration from one of our recent LfD experiment in which the robot learns to roll out a pizza dough by gradually gen-

Fig. 11 Illustrative scenario of an interactive skill analysis process. A section of the 3D flow field related to the task is projected on the board, while the user moves the planar surface in the manipulator's workspace.

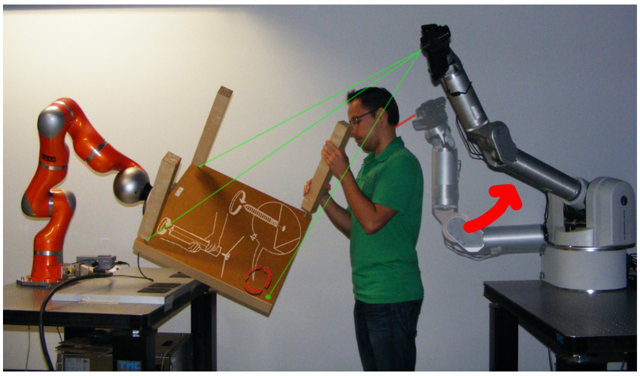


Fig. 12 Illustrative scenario of a collaborative assembly task where the user is screwing a leg of the table and occluding the view of the system interface. The system interface should here react by moving sideways to avoid the user's occlusion while projecting information on the table top held by another robot (a 7 DOFs Kuka Lightweight Arm).

eralizing the movement in the form of a 3D flow field. As shown in Fig. 11 the user holds a cardboard in a desired position/orientation around the manipulator, and a section of the 3D flow fields is computed in real-time and projected on the hand-held surface. The user can assess the robot's current understanding of the task by visualizing the generalization capabilities of the robot without having to reproduce the movement in several initial configurations.

6.3 Collaborative Assembly Task: active projection handling occlusion

The human presence plays a fundamental role for the proposed interface, due to the fact that all input/output communication channels involve the user as sender and receiver of the different information flows. In addition, the physical position and orientation of the user in the Operating Space drives the modalities to project and to capture data. To make the process transparent, learning information about the task should be presented in an area of the workspace that is convenient for the user. Also guidance needs to be perceived from the correct angle of view. One example is when occlusions occur. If the user occludes the frustum of the projective device and make it infeasible to project information on a particular surface, the proposed system could be used to bypass the obstacle and find an appropriate projection posture for the current situation. To assess reproduction trials, the robot can also actively place the sensors in order to track the important task features. If the task requires to move a tool in a given plane, the robot arm can move its sensors in order to track the tool's path along the plane that is of relevance for this specific task.

This type of scenario is highlighted in a collaboration task in which the robot helps the human to cooperatively assemble an *IKEA* table. In such a scenario, visual guidelines could for example be projected on the table to be mounted by using the proposed interface. In Fig. 12, the user changes his posture to assemble one of the legs. The robot should manage the occlusion by changing its point of view to be able to continue projecting information on the desired surface. User's guidelines could also be shown as interactive explanations of the current step of the assembly, such as highlighting positions on the table to drill the table legs (see Fig. 12).

7 Conclusion

We presented a novel tangible interface in the context of LfD for the assessment of the robot skills acquisition through active sensing and interactive data visualization. We discussed the importance of designing interfaces for non-expert users, who should not need to be skilled in robotics and computer programming to socially teach skills to robots. We implemented a prototype to demonstrate the technical feasibility of the proposed interface, by combining a perceptive and a projection device mounted on an actively compliant robot manipulator. Then, we emphasized the approach by introducing a set of concept scenarios in which the proposed interface may be applied. In the context of social bidirectional teaching interactions we believe that the proposed system can help the instructor for giving or receiving feedbacks in a natural and intuitive manner, helping him keep his focus of attention to the task without disrupting the teaching activity. The proposed experiment opens new research perspectives that will be part of our future work. Firstly, the proposed architecture needs to be tested in a realistic context, by identifying and collecting the parameters to measure the quality of the teaching process and analyzing the resulting data in user studies. Such results can be compared to the performances obtained in the same scenario without the help of the proposed active interface. By taking insight from affective computing studies, psychological factors may be similarly taken into account to study the role of the device as a social actor in the teaching interaction. Secondly, a selection of the envisaged scenarios will be selected for a complete implementation of the scenario. Lastly, since only planar surfaces have been considered, future work will consider projection on more complex surfaces. We will study the option of projecting instead of 2D images projected on a planar surface, 3D textures on curve-shaped objects.

References

1. Alenyà, G., Dellen, B., Torras, C.: 3d modelling of leaves from color and tof data for robotized plant measuring. In: Proceedings of IEEE International Conference on Robotics and Automation (ICRA), pp. 3408–3414 (2011)
2. Aleotti, J., Caselli, S.: Grasp recognition in virtual reality for robot pregrasp planning by demonstration. In: Proceedings of IEEE International Conference on Robotics and Automation (ICRA), pp. 2801–2806 (2006)
3. Bimber, O., Raskar, R.: Spatial augmented reality: A modern approach to augmented reality. In: In Proceedings of Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH. ACM, New York, NY, USA (2005)
4. Calinon, S., Billard, A.: What is the teacher's role in robot programming by demonstration? - Toward benchmarks for improved learning. Interaction Studies. Special Issue on Psychological Benchmarks in Human-Robot Interaction **8**(3), 441–464 (2007)
5. Calinon, S., Guenter, F., Billard, A.: On learning, representing and generalizing a task in a humanoid robot. IEEE Trans. on Systems, Man and Cybernetics, Part B **37**(2), 286–298 (2007)
6. Calinon, S., Sardellitti, I., Caldwell, D.G.: Learning-based control strategy for safe human-robot interaction exploiting task and robot redundancies. In: Proc. IEEE/RSJ Intl Conf. on Intelligent Robots and Systems (IROS), pp. 249–254. Taipei, Taiwan (2010)
7. Chernova, S., Veloso, M.: Confidence-based multi-robot learning from demonstration. International Journal of Social Robotics **2**(2), 195–215 (2010)
8. Christian, V., Maik, P., Christoph, W., Norbert, E.: Towards safe physical human-robot collaboration: A projection-based safety system. In: Proceedings of International Conference on Intelligent Robots and Systems IEEE/RSJ, pp. 3355–3360. San Francisco, CA, USA (2011)
9. Delson, N., West, H.: Robot programming by human demonstration: the use of human variation in identifying obstacle free trajectories. In: Proceedings of IEEE International Conference on Robotics and Automation (ICRA), pp. 564–571 (1994)
10. Dourish, P.: Where the action is: the foundations of embodied interaction. MIT Press, Cambridge, MA, USA (2001)
11. Ekvall, S., Kragic, D.: Learning task models from multiple human demonstrations. In: Robot and Human Interactive Communication. ROMAN. The 15th IEEE International Symposium on, pp. 358–363 (2006)
12. Faugeras, O.D., Toscani, G.: The calibration problem for stereo. In: Proceeding of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), pp. 15–20 (1986)
13. Featherstone, R., Orin, D.E.: Dynamics. In: B. Siciliano, O. O. Khatib (eds.) Handbook of Robotics, pp. 35–65. Springer, Secaucus, NJ, USA (2008)
14. Friedrich, H., Muench, S., Dillmann, R., Bocionek, S., Sassin, M.: Robot programming by demonstration (RPD): Supporting the induction by human interaction. Machine Learning **23**(2), 163–189 (1996)
15. Gillet, A., Sanner, M., Stoffler, D., Olson, A.: Tangible augmented interfaces for structural molecular biology. IEEE Comput. Graph. Appl. **25**, 13–17 (2005)
16. Greenfield, P.: Theory of the teacher in learning activities of everyday life. In: Everyday Cognition: It Development in Social Context
17. Harrison, C., Benko, H., Wilson, A.D.: Omnitouch: wearable multitouch interaction everywhere. In: Proceedings of the 24th annual ACM symposium on User interface software and technology, UIST, pp. 441–450. ACM, New York, NY, USA (2011)
18. Hosoi, K., Dao, V.N., Mori, A., Sugimoto, M.: Visicon: a robot control interface for visualizing manipulation using a handheld projector. In: Proceedings of the international conference on Advances in computer entertainment technology, pp. 99–106. ACM, New York, NY, USA (2007)
19. Ishii, H., Ullmer, B.: Tangible bits: towards seamless interfaces between people, bits and atoms. In: Proceedings of the SIGCHI conference on Human factors in computing systems, CHI, pp. 234–241. ACM, New York, NY, USA (1997)
20. Ito, M., Noda, K., Hoshino, Y., Tani, J.: Dynamic and interactive generation of object handling behaviors by a small humanoid robot using a dynamic neural network model. Neural Networks **19**(3), 323–337 (2006)
21. Kazuki, K., Seiji, Y.: Extending commands embedded in actions for human-robot cooperative tasks. International Journal of Social Robotics **2**(2), 159–173 (2010)
22. Kendall, D.G.: A survey of the statistical theory of shape. Statistical Science **4**(2), 87–99 (1989)
23. Kheddar, A.: Teleoperation based on the hidden robot concept. IEEE Transactions on Systems, Man, and Cybernetics, Part A **31**(1), 1–13 (2001)
24. Klemmer, S.R., Hartmann, B., Takayama, L.: How bodies matter: five themes for interaction design. In: Proceedings of the 6th conference on Designing Interactive systems, DIS, pp. 140–149. ACM, New York, NY, USA (2006)
25. Kuniyoshi, Y., Inaba, M., Inoue, H.: Teaching by showing: Generating robot programs by visual observation of human performance. In: Proceedings Intl Symposium of Industrial Robots, pp. 119–126. Tokyo, Japan (1989)
26. Lauria, S., Bugmann, G., Kyriacou, T., Klein, E.: Mobile robot programming using natural language. Robotics and Autonomous Systems **38**(3-4), 171–181 (2002)
27. Linder, N., Maes, P.: LuminAR: portable robotic augmented reality interface design and prototype. In: Adjunct proceedings of the 23nd annual ACM symposium on User interface software and technology, UIST, pp. 395–396. ACM, New York, NY, USA (2010)
28. Luh, J.Y.S., Walker, M.W., Paul, R.P.C.: Resolved-acceleration control of mechanical manipulators. IEEE Trans. on Automatic Control **25**, 468–474 (1980)
29. Nicolescu, M.N., Mataric, M.J.: Natural methods for robot task learning: Instructive demonstrations, generalization and practice. Second International Joint Conference on Autonomous Agents and Multi-Agents and Multi-Agent Systems Melbourne, Australia
30. Ogawara, K., Takamatsu, J., Kimura, H., Ikeuchi, K.: Extraction of essential interactions through multiple observations of human demonstrations. IEEE Transactions on Industrial Joint Conference on Autonomous Agents and Multiagent Systems pp. 667–675 (2003)
31. Okada, K., Kino, Y., Inaba, M., Inoue, H.: Visually-based humanoid remote control system under operator's assistance and its application to object manipulation. In: Proc. IEEE Intl Conf. on Humanoid Robots (Humanoids). Karlsruhe, Germany (2003)
32. OpenKinect: Kinect calibrations. Online available at http://openkinect.org/wiki/Calibration
33. OpenNI: User Guide (2010). Online available at http://openni.org/documentation

34. Park, J., Kim, G.J.: Robots with projectors: an alternative to anthropomorphic hri. In: Proceedings of the 4th ACM/IEEE international conference on Human Robot Interaction, HRI, pp. 221–222. ACM, New York, NY, USA (2009)
35. Primesense: NITE 1.3 Algorithms notes (2010). Online available at http://primesense.com
36. Rusu, R.B., Cousins, S.: 3D is here: Point cloud library PCL. In: Proceedings of IEEE International Conference on Robotics and Automation (ICRA). Shanghai, China (2011)
37. Schöning, J., Rohs, M., Kratz, S., Löchtefeld, M., Krüger, A.: Map torchlight: a mobile augmented reality camera projector unit. In: Adjunct Proceedings of the 27th International Conference on Human Factors in Computing Systems, CHI EA, pp. 3841–3846. ACM (2009)
38. Segre, A.M.: Machine learning of robot assembly plans. Kluwer Academic Publishers, Norwell, MA, USA (1988)
39. Shotton, J., Fitzgibbon, A.W., Cook, M., Sharp, T., Finocchio, M., Moore, R., Kipman, A., Blake, A.: Realtime human pose recognition in parts from single depth images. In: Proceeding of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1297–1304 (2011)
40. Song, H., Grossman, T., Fitzmaurice, G., Guimbretiere, F., Khan, A., Attar, R., Kurtenbach, G.: Penlight: combining a mobile projector and a digital pen for dynamic visual overlay. In: Proceedings of the 27th international conference on Human factors in computing systems, CHI, pp. 143–152. ACM, New York, NY, USA (2009)
41. Thomaz, A.L., Breazeal, C.: Teachable robots: Understanding human teaching behavior to build more effective robot learners. Artif. Intell. **172**(6-7), 716–737 (2008)
42. Voyles, R., Khosla, P.: A multi-agent system for programming robotic agents by human demonstration. AI and Manufacturing Research Planning Workshop (1998)

Davide De Tommaso is a Ph.D. student at the Advanced Robotics Department, Istituto Italiano di Tecnologia (IIT), where he has joined the Learning & Interaction Group in 2011. He received a M.S. degree in Computer Engineering from the University of Catania, Italy, in 2010. He is actually working on developing multi-modal interfaces in human-robot learning scenarios. His research interests involve human-computer and human-robot interaction, augmented reality and robot learning.

Sylvain Calinon is a Team Leader at the Department of Advanced Robotics, Istituto Italiano di Tecnologia (IIT), where he leads the Learning & Interaction Group since 2009. He received a PhD on robot programming by demonstration in 2007 from the Learning Algorithms and Systems Laboratory (LASA), Ecole Polytechnique Federale de Lausanne (EPFL), Switzerland, which was awarded by the international Robotdalen scientific award, ABB award and EPFL-Press distinction. From 2007 to 2009, he was a postdoctoral research fellow at LASA-EPFL, where he maintains research collaborations as a visiting researcher. His research interests cover robot learning, imitation and human-robot interaction.

Darwin G. Caldwell is a Director at the Istituto Italiano di Tecnologia in Genoa, Italy, and a Visiting/Honorary Professor at the Universities of Sheffield, Manchester, Bangor and Kings College, London. His research interests include innovative actuators and sensors, haptic feedback, force augmentation exoskeletons, dexterous manipulators, humanoid robotics (iCub), bipedal and quadrupedal robots (HyQ), biomimetic systems, rehabilitation robotics, telepresence and teleoperation procedures, and robotics and automation systems for the food industry. He is the author or co-author of over 200 academic papers, 8 patents and has received awards at several international conferences and events including, ICRA, IROS, ICAR, Humanoids, IAV and Virtual Concepts. He is on the editorial board of Journal for Robots, International Journal of Social Robotics and Industrial Robot and Technical Editor of the Trans. in Mechatronics.