Journal of Ambient Intelligence and Humanized Computing
https://doi.org/10.1007/s12652-019-01676-6

ORIGINAL RESEARCH

Development and experimental validation of algorithms for human–robot interaction in simulated and real scenarios

A. Freddi[1] · S. Iarlori[1] · S. Longhi[1] · A. Monteriù[1]

Received: 18 December 2018 / Accepted: 26 December 2019

Abstract
The development of robots, which can safely and effectively interact with people and assist them in structured environments, is an open research problem whose importance has been growing rapidly in the last years. Indeed working in shared environments with human beings, these robots require new ways to achieve human–robot interaction and cooperation. This work presents an approach for performing human–robot interaction by means of robotic manipulators. The interaction is composed by three main steps, namely the selection, the recognition and the grasping of an object. The object selection is recorded on the base of a gesture execution, realized by the user in front of a RGB-D camera and associated to each particular object. The object recognition is achieved by means of the RGB cameras mounted on the two manipulator arms, which send the workspace information to a specific classifier. With the aim of realizing the grasping step, the object position and orientation are extracted in order to correctly rotate the gripper according to the object on the desk in front of the robot. The final goal is to release the grasped object on the hand of the user standing in front of the desk. This system could support people with limited motor skills who are not able to get an object on their own, playing an important role in structured assistive and smart environments, thus promoting the human–robot interaction in activity of daily living.

Keywords Human–robot interaction · Human–robot cooperation · Robotic manipulators

1 Introduction

Human–robot interaction (HRI) is a field of study dedicated to design robotic systems to be used by or with people. The interaction requires communication between robots and humans and can be realized in several forms, depending whether the human and the robot are in close proximity to each other or not.

The communication and the interaction can be separated into two main categories:

- remote interaction during which the human and the robot are not co-located and are, spatially or even temporally, separated;
- proximate interactions during which the humans and the robots are co-located (for example, service robots may be in the same room as humans).

Within these two categories, it is useful to distinguish between applications that require mobility, physical manipulation, or social interaction. Remote interaction with mobile robots is referred to teleoperation, while with a physical manipulator is defined as telemanipulation. Proximate interaction with mobile robots may fall within the robot assistant domain and may include a physical interaction. Social interaction may be either remote or proximate (cognitive and/or physical interaction). In the last years, the HRI field is in constant growth due to the attention given to this topic and the great progresses realized about close-proximity activities (Feil-Seifer and Matarić 2005; Bekele et al. 2016; Clodic et al. 2017).

For the assistive robots, the goal is to create close and effective interactions with a user, to provide assistance

✉ S. Iarlori
sabrina.iarlori@gmail.com; s.iarlori@univpm.it

A. Freddi
a.freddi@univpm.it

S. Longhi
s.longhi@univpm.it

A. Monteriù
a.monteriu@univpm.it

[1] Department of Information Engineering, Università Politecnica delle Marche, Via Brecce Bianche, 60131 Ancona, AN, Italy

Published online: 11 January 2020

and achieve measurable progress in convalescence, rehabilitation and learning. In a near future, service robotic platforms will be machines that cooperate with human-beings to assist people in daily activities at home, providing for an easier and healthier life (D'Onofrio et al. 2018; Cavallo et al. 2013). In this context, assistive robots, in particular for older adults, can be categorized into two main subgroups:

- physically assistive rehabilitation robots that provide functional and physical assistance, as shared transportation of loads or passing objects (Lee et al. 2016);
- socially assistive robots (or social robots) that aim to improve the quality of life and can have measurable social interactions with people (Sciutti and Sandini 2017).

Despite physical and/or social purposes, an assistive robot has to interact with the user in some ways (Romer et al. 2005). On the social expectation, the robots main application areas will change from the typical industry field to daily living environments in the near future, where many communication skills for the service robot have been proposed as gesture or voice recognition (Fujii et al. 2014).

The present work proposes an assistive robotic architecture which allows the interaction between a user and a robot manipulator, where an RGB-D vision system is used as communication instrument. The proposed interaction would lead to the robot picking an object on behalf of the user, who may find it difficult to do it by him/herself. The user who stands in front of the camera can select a desired object, by means of specific gestures. Each gesture is associated to one of the objects available within the robotic manipulator workspace. The defined gestures should be as natural for the user as possible to reduce training or learning of a specific set of gestures (Canal et al. 2016). Moreover, the robot should understand gestures as a human would understand another humans's gesture, and should reply to them in real time. Static gestures are those in which the user place his/her limbs in a specific position and stands for a while, without any dynamics or involved movement. In this case, the transmitted information is obtained through a static pose configuration related to a single object choice. The object, selected by the user, is then recognized by the robot by an histogram of orientation gradient (HOG) algorithm and a support vector machine (SVM) classifier. The object position and orientation are extracted in order to realize the grasping task, at the end of which the robot manipulator detects, by means of the RGB-D sensor, the user hand and places the object on it to complete the interaction task. The proposed architecture has been implemented by using the robotic operating system (ROS) framework and tested in the Gazebo simulator, where a model of the robot and the RGB-D camera are represented. Finally, the interaction task has been tested in the real environment of the university laboratory.

The working steps of the whole system are summarized as follows:

- the user selects the object he/she needs by a specific gesture associated to each object;
- the system recognizes the desired object through the HOG algorithm and the SVM classifier;
- the object position and orientation in the workspace are extracted to realize the grasping task;
- the object is placed on the user hand localized through the Microsoft Kinect camera (Chen et al. 2013).

The paper is organized as follows. The works that report similar applications are described in Sect. 2. The main architecture is presented in Sect. 3 with the hardware, Baxter robot and Microsoft Kinect sensor camera, and software related to the robot manipulator tasks. The trials on the simulator and the experimental tests are shown in Sect. 3, while the results, conclusions and future works are addressed in Sects. 4 and 5.

2 Related works

Different applications of HRI are proposed in the literature, aiming at providing various kind of supports for people with special needs in structured and assistive environments (Foresi et al. 2018). In the last years different solutions for supporting people at home have been developed, based mostly on the monitoring and analysis of human motion for robot control or human activity recognition in assistive environments (Sarcevic et al. 2019; Poncela et al. 2018). Vision based systems are employed as interaction technologies to classify gestures using machine learning methodologies as in (Ding and Chang 2017) for controlling robots in HRI process (Mendes et al. 2017). Different approaches have focused on gesture elicitation that identified the preferred gesture and body language communication to interact with robots, as in Pons and Jaen (2019), where the body language used by children to communicate with ground robot has been investigated. The goal of this kind of studies is to make use of a minimal number of sensors and, at the same time, being a low cost solution to identify human behaviors that are used to adapt the way a robot performs (Le et al. 2016). A HRI system, able to recognize gestures pointing at an object, was developed in Canal et al. (2016) using a dynamic time warping (DTW) approach based on gesture specific features computed from depth maps. The pointed location is estimated in order to detect candidate objects the user refers to by the use of a Kinect sensor. Another approach, proposed in Raheja et al. (2018), has been investigated localizing the 3D position of upper human body skeletal joints and tracking the skeleton. This was achieved with real time constraints using Microsoft Kinect sensor which has the ability to estimate human joint location which results invariant to pose, clothing, body shape, etc. The developed method allows selecting the object using hands and different body configuration. A gesture recognition module is also in Tellaeche et al. (2018) developed to be used in collaborative human robot application in safe environments to command robots: in this case the presence of the involved robots in a different task proves that they can be easily integrated also in working environment. The commands can be modeled as static gestures, maintained for a short predefined period of time.

Different works investigate other approaches for technical gestures recognition using RGB-D camera in a collaborative task between a robot and an operator. In Coupeté et al. (2015), technical gestures are performed to enable the robot to understand which task has been just executed in order to anticipate human actions. The same task of selecting an object and taking it, can be achieved also by integrating a brain computer interface (BCI) with a robotic manipulator, as proposed in Foresi et al. (2019): it consists in selecting an object via an assistive interface and moving it by means of a robotic arm. Also for activities of daily living as dressing, assistive robots can provide their contribution as in Chance et al. (2016), where a compliant robotic arm has been employed. In detail, the Baxter robot was used to dress one arm of a jacket, by tracking the joints location and calculating their trajectory. Another kind of approach, proposed in Morales et al. (2017), shows a human–robot interaction in structured smart environments, where different robot systems cooperate to realize a task. Vision based systems, as Microsoft Kinect, are not only employed for the motion and gesture analysis, but also for recognizing the different types of objects that are involved in the tasks as in Wen et al. (2019). In Mettel et al. (2019), an application of the Kinect sensor has been investigated for detecting potentially dangerous objects, in a user walking path, on the base of scene analysis in a depth image. Moreover, the capability of manipulating objects based on vision systems is essential for robot applications in the context of human robot interaction and cooperation (Wen et al. 2019).

The paper presents integrate mobile robots and robotic manipulators systems as in Achic et al. (2016), where an electrical wheelchair with an embedded robotic arm assists a user to realize a task as picking up a cup of water. With the same scope, in Tsui et al. (2008) the authors equipped a wheelchair with a robotic arm and a graphical interface to select an object to place on a shelf. In the literature, there are also projects addressing the development of assistive robotics, e.g., the ACCRA (Agile Co-Creation for Roots and Aging) project, which aims to promote the independent living to support daily life management (D'Onofrio et al. 2018).

3 Materials and methods

3.1 Case of study

In the proposed work, a human–robot interaction between a user and a robot manipulator has been realized to simulate an assisting task for people with special needs. In detail two different objects, namely a bottle and a box, have been chosen to be recognized, located, grasped and placed on the user hand. For the pick and place phase, the Baxter robot was employed to use its two arms, separately. Once the user is in front of the desk where the desired objects are presented, a Microsoft Kinect sensor is adopted to recognize the user gesture by means of the joints position and orientation estimated by the Tracking Skeleton System.

The task is composed by different phases: a first one related to the user detection in front of the kinect with the recognition of the gesture associated to the desired object, a second one related to the object recognition by a dedicated algorithm, and a pick and place phase that completes the motor task, leaving the object on the user hand (see Fig.1); all phases are implemented in the robotic operating system (ROS).

The object detection and recognition, together with pick and place were preliminarily tested in a simulator environment, the Gazebo 3D simulator, where the robot, the desk and the used objects are reproduced to test the algorithm and the pick and place. Also the Kinect sensor camera is introduced in Gazebo environment but, since it was not possible to reproduce the subject in Gazebo, the release point of the object is fixed a priori.

3.2 Hardware: anthropomorphic Baxter robot and Kinect

The Baxter robot and its two robotic arms, illustrated in Fig.2a, have been chosen to realize the interaction with the user for the pick and place phase. The Baxter robot communicates directly with the controller via a ROS node created to manage the pick and place. The controller provides joint angle trajectories for Baxter's 7 degree-of-freedom arms in order to take the object selected by the user.

The chosen RGB-D vision sensor camera is the Microsoft Kinect v1 (see Fig.2b) that allows to compute a 3D structure of the scene, with a good invariance against illumination changes, color and texture. Microsoft for Windows v1 sensor employs the structured light (SL) cameras with the triangulation system (Chen et al. 2013). The

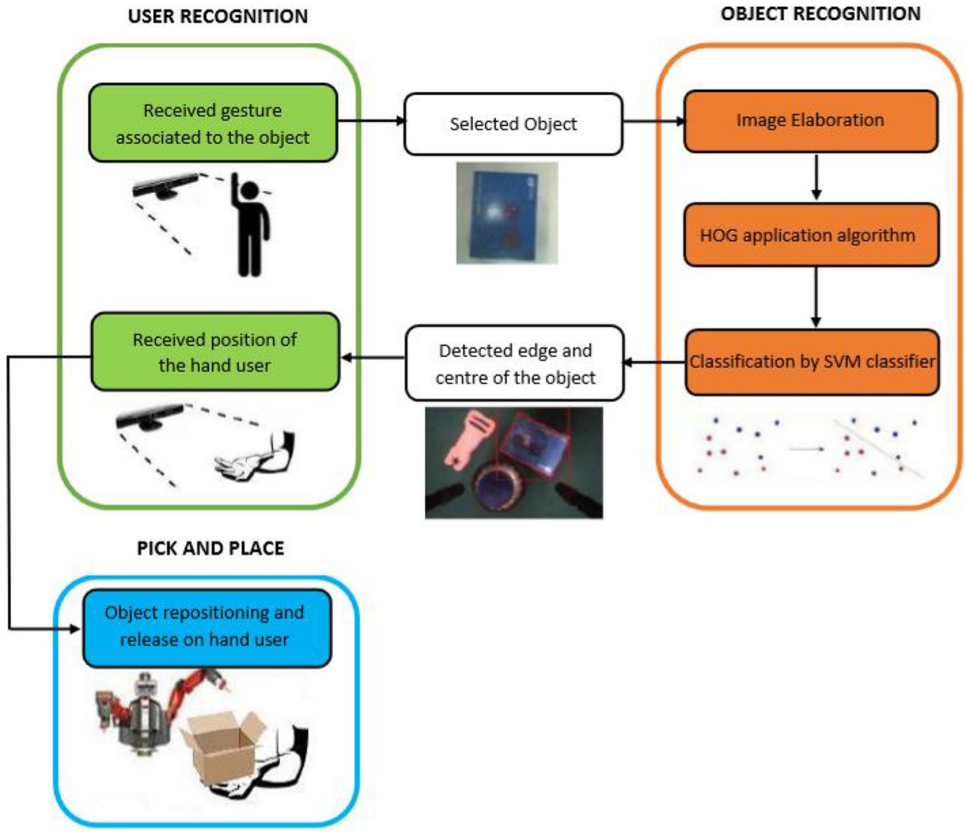


Fig. 1 Scheme of the proposed interaction task, structured in the user and object recognition phases and the pick and place on the user hand

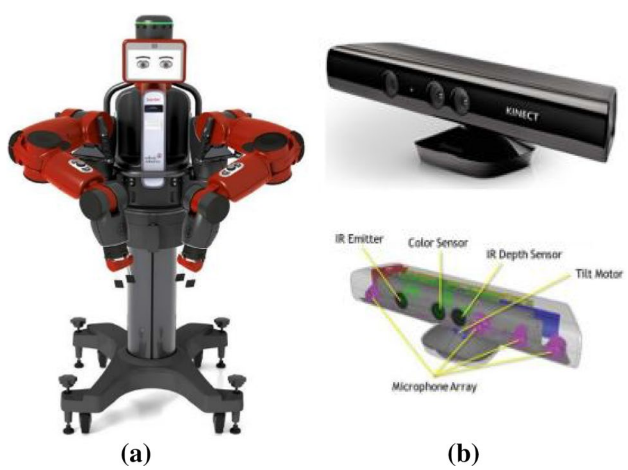


Fig. 2 The hardware setup: **a** the Baxter research robot; **b** the kinect sensor camera

Kinect v1 provides a depth map resolution (640×480), allowing to recognize thin objects and solving some ambiguity problems. Moreover, Kinect v1 is a cheap, unobtrusive and easy to set up sensor that can be used both at home or structured environment. The depth features allow the recognition of different people and different body parts in the field of view, while the resolution permits to identify the 3D points of 20 distinct body parts at 30 fps, that have been considered for the gesture recognition.

3.3 Software: ROS and Gazebo

The software architecture is composed by ROS and the Gazebo simulator, chosen for their easy interfacing and possibility of integration.

ROS is a robotic middleware (i.e., collection of frameworks for robot software development) and, although it is not an operating system, it provides services designed for heterogeneous computer cluster such as hardware abstraction, low-level device control, implementation of commonly used functionality, message-passing between processes and package management. Running sets of ROS-based processes are represented in a graph architecture where processing takes place in nodes that may receive and send messages, and control multiple sensors and actuators (Quigley et al. 2009). All the implemented algorithms are managed by ROS structure nodes.

While ROS serves as the interface for the robot, Gazebo is a 3D simulator by which it is possible to create a 3D scenario on the computer with robots, obstacles and many other objects; it also uses a physical engine for illumination, gravity, inertia, etc.. Gazebo was designed to evaluate algorithms, for many applications; in fact, it is essential to test the developed robot applications, like error handling, battery life, localization, navigation and grasping.

3.4 Simulator implementation

All the algorithms related to the object recognition by means of the kinect sensor camera, were tested in the Gazebo simulator (see Fig. 3). The first step was to add the external Kinect system, Fig. 4, in Gazebo by a predefined ROS package and the recognition of the kinect reference system of the frame by the Baxter robot. Moreover, it was possible to convert the pixel coordinates of the kinect system in the Baxter system coordinates in order to realize the pick and place task of objects such as a box and a cylinder, for the simulator interface.

In order to integrate the Kinect sensor in the Gazebo simulator, the first problem was that of assigning a static position to the Kinect in order to guarantee a complete vision without obstacles of the desk. An image of the Kinect vision in the simulated environment is reported in Fig. 4b where it is possible to focus the object on the desk. Moreover, it was necessary to delete the attribution of gravity for the Kinect sensor, allowing the position of the sensor everywhere in Gazebo. At the same time, it was necessary to realize the interaction between the Baxter and the Kinect camera, expressing the coordinates of the kinect position

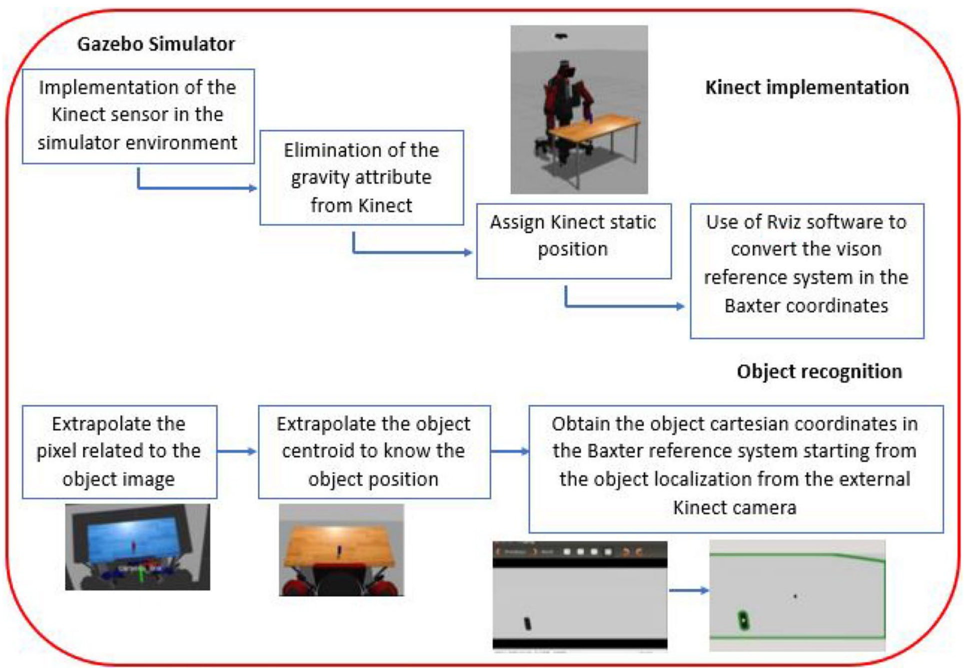
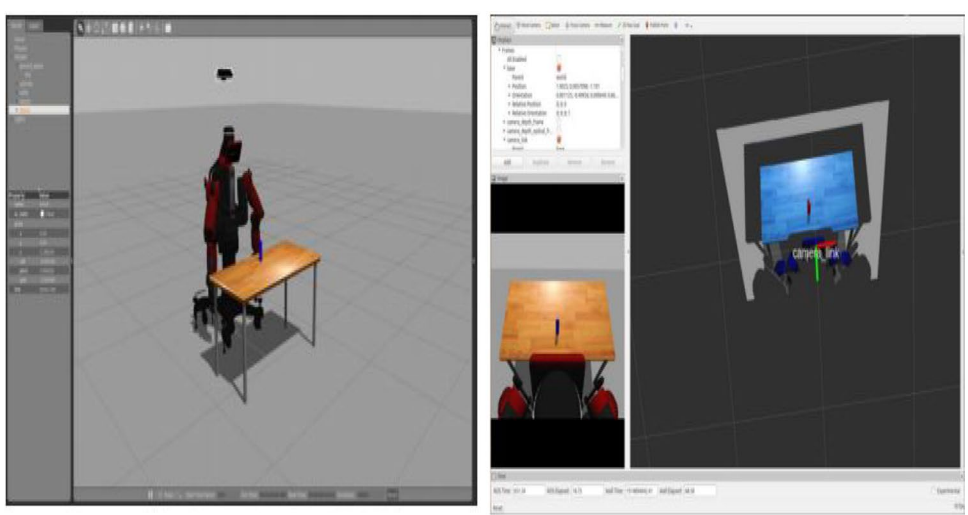


Fig. 3 Description of all the steps realized in the Gazebo simulator related to the integration of the Kinect sensor and the algorithm for object recognition and pick and place



(a) (b)

Fig. 4 **a** The Kinect and the Baxter robot integrated in the simulator environment; **b** the Kinect sensor camera vision from a top view upon the Baxter robot with respect to the Baxter reference system. In order to realize this step, the conversion from Euler angles to quaternions was requested for Rviz that is a specific software in ROS to visualize the recorded frames acquired by Kinect.

Moreover, an image of the object to grasp is acquired by Kinect in order to extract the object centroid, the edge and the center in the pixel image, so that conversion from the pixel vision system to the cartesian coordinates observed from the base of the Baxter can be performed. This step is necessary to realize the pick and place in the robot reference system.

3.5 User recognition

The algorithm for the user recognition allows to cooperate with the robot by means of the gesture recognition and is able to locate the user hand where the object has to be placed. In order to interact with the Microsoft Kinect sensor, it is necessary to communicate through the open source OpenNI (Open Natural Interaction) library package. This package can be easily integrated in the ROS system and lunched by a package *openni_camera* to acquire and publish the camera raw data. The package *openni_tracker* allows the real-time tracking of the user joints framed by Kinect. Also in the Gazebo simulator all the reference systems related to the joints are mapped and reproduced by the broadcasting of the reference systems transformation. To realize the body tracking, the user has to reproduce the *Psi* pose in front of the camera (see Fig. 5) standing with the legs slide apart, elbows at right angles and hands pointing upwards.

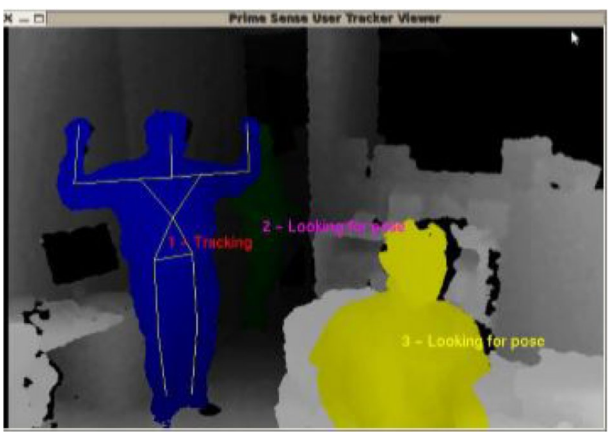


Fig. 5 Description of the *Psi* pose that the subject has to assume for the calibration phase in front of the kinect

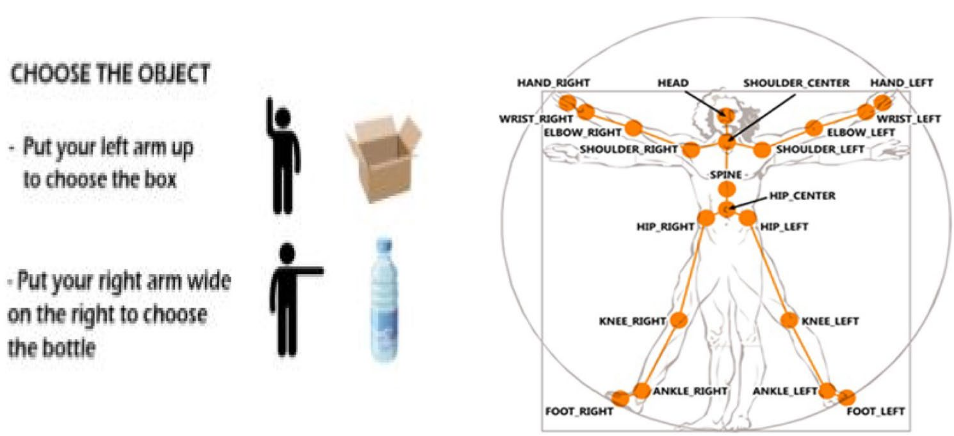
Two kind of gestures have been associated to the different objects on the desk: the position of the left arm stretched upwards is the command to get the box, and the right arm stretched outwardly is that for the bottle (see Fig. 6a). The two configurations are strictly related to the position of the shoulder, the elbow and the wrilst joints for the two arms as they are described in Fig. 6b. The algorithm ignores any other configuration of the skeleton joints not included in the command list. In order to avoid erroneous gesture, it has been set that the gesture has to be kept for two seconds.

Finally, the estimation of the hand position to receive the object has been calculated and a tolerance area of ±2 cm of diameter plus the accuracy suggested by the producer has been considered (see Fig. 7) during the placing of the object.

3.6 Object recognition

For the object recognition phase, the Baxter Research Robot is equipped with two colour cameras with an effective resolution of 640x400 pixels, located on each arm, namely "*left_hand_camera*" and "*right_hand_camera*" (see Fig. 8a). The goal of this stage is to recognize via the gripper camera, one among the two objects presented in Fig. 8b, which are physically placed in the robot workspace, on a desk 40 cm under the robot arm camera. Once the user selects the object, it is recognized from the image captured

Fig. 6 **a** Gesture for object selection; **b** body joints recognized by kinect camera



(a) **(b)**

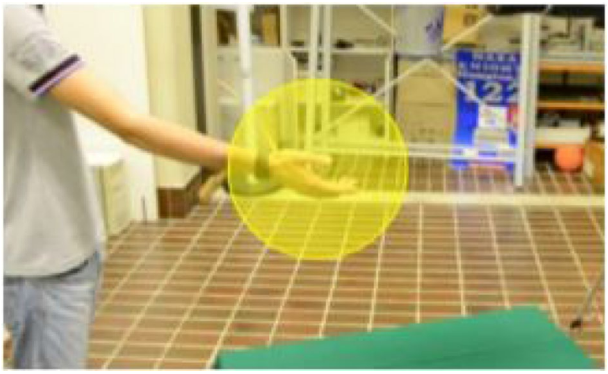


Fig. 7 User hand image acquired by the kinect sensor. The yellow circle indicates the area of tolerance for the object to be placed on the user hand

by the Baxter camera, the object image is then inscribed within a rectangle whose center point falls within the object image itself.

In order to realize this goal, a solution composed by the application of the algorithm of histogram of orientation gradient (HOG) and the support vector machine (SVM) classifier is proposed. After a training phase, with all the images related to the possible objects, the SVM is able to recognize an object on the base of the feature extraction obtained by the HOG, from the analysis of the acquired image.

The first step was to acquire the single snapshot (i.e., image) of the objects on the desk, the box and the bottle, by one of the two cameras mounted on each of the robot arms, and an image with all objects on the desk. The model creation process involves a large amount of positive and negative images of the objects by which it is possible to realize the training of the classifier. To obtain a good classification, it is better to have a huge amount of negative images related to the different objects with respect to the object to identify. During the training phase, images of the object that has to be found, are chosen as positive samples. All the positive and negative images were realized centering the object in the new image and limiting the background of each image. For each object, different images have been cropped and resized (100 × 100 pixels), the original images have been converted to gray-scale and resized, editing them by different transformations to get more positive images, for a total of approximately 2000 positive images. The image processing was related to blur filter (smoothing, median, bilateral filter), brightness (plus and minus) and contrast. By means of the HOG algorithm, it is possible to extract the features of each object in order to produce a feature vector. The features are extracted both for the positive and the negative images.

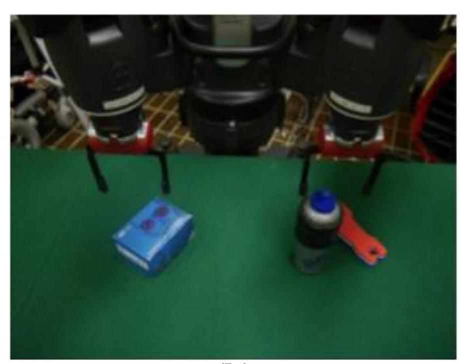
Once the features are obtained, it is possible to train the SVM classifier. For each object to classify, all the features are inserted in a matrix that includes the information related to the positive and negative images. For each object a model of classifier is generated in order to determine if an object is present or not in the image. In the testing phase, once the image from the Baxter arm camera is acquired, it is necessary to determine if the object is present or not in the scene, so in the snapshot to analyze, a sliding window of 100×100 pixels with a step size of 20×20 pixels scans the image in order to find and recognize the desired object. After every window scanning, the features are extracted through the HOG algorithm and checked with the SVM that will find or not the desired object. When the whole snapshot is analysed, a factor of 1, 1 is introduced and then the procedure is iterated until the dimension of the sliding window is bigger than the snapshot. Once this phase is completed, it is possible to have track of all the positive occurrences with the relative scores. The window with the maximum confidence score is that with the highest probability to find the desired object.

3.7 Pick and place algorithm via Baxter robot

The movement of the Baxter arm for picking the object and placing it to a different position is performed by applying the inverse-kinematics (IK) pick and place. This method combines a simulated IK service to obtain the joint angles solutions for a given cartesian orientation endpoint, and for a controlled position movement together with grasping and releasing services. The position of

Fig. 8 **a** Position of the RGB camera under the left arm; **b** image of different objects taken by the *left_hand_camera*

(a)



(b)

the object on the desk is calculated as the center of the designed rectangle that inscribes the desired object, recognized among all those presented in the database by the SVM classifier. Also the orientation of the object on the desk has been considered in order to address with the same angle the arm gripper. This has been realized by knowing the pose of the Baxter arm, its height from the table where the object is located, the camera calibration factor and using the following formula

$$B = (Ip - Cp) \times cc \times d + Bp + Go \tag{1}$$

where B is the Baxter point in x or y direction, Ip is the image pixel, Cp is the centre image pixel: height/2 (x direction) or width/2 (y direction), cc is the camera calibration factor (= 0.0029), d is the distance from the table (=0.30m), Bp is the Baxter point in pixels coordination and Go is the gripper offset. It is important to note, that the calculation in x-direction is for the front/back movement of the Baxter arm, while the calculation in y-direction is for the left/right movement of the Baxter arm. The setup of the Baxter Robot Pick and Place experiment, with an old and new position related to the box, can be seen in Fig. 9. For the proposed task the new position was represented by the user hand localized by means of the kinect sensor.

The test has been conducted for the two considered objects: the box and the bottle. While for the box grasping, the gripper was directly positioned upon the object at a fixed distance along the z axis in correspondence of the center of the rectangle that inscribes the object, for approaching the bottle, the gripper is slowly moved around the object in order to avoid possible collision that can affect the grasping.

4 Results

Following the preliminary tests in the Gazebo simulator, different experimental tests have been realized in the laboratory. The obtained results are reported and commented. About the calibration and the joint skeleton tracking, no errors have been recorded and the tracking was correctly completed for the subject. The user calibration was tested on different people with various light conditions and the results were always satisfying. The object recognition and its coordinates position extraction have been successfully realized for the two considered objects. Even if the light conditions have affected in some analysis the edge extraction, this problem has not influenced the final object grasping. As reported in Table 1, the results for the box and the bottle recognition are affected by the different parameters configuration tested for the classifier training, and used to generate the file model for each single object. The combined parameters are related to the number of positive and negative images, the cropping and smoothing actions and the downscaling. The results are related to different test realized on white and green desks in order to evaluate the best conditions for object recognition. In general, the object grasping resulted stable and no object loss ever occurred for all the realized tests. A minor question is connected with the field of view of the kinect sensor. Sometimes, to raise the hand, the robot arm can stretch blocking the camera view in the area where the subject can position the hand. This condition verified rarely in the proposed test but it constitutes a specific critical issue.

To evaluate the accuracy estimation of the user hand position, a comparison with the object placing from its position to a fixed point in the Baxter reference system was considered. In Cremer et al. (2016) the results related to different tests, about object manipulation and in particular pick and place task, were

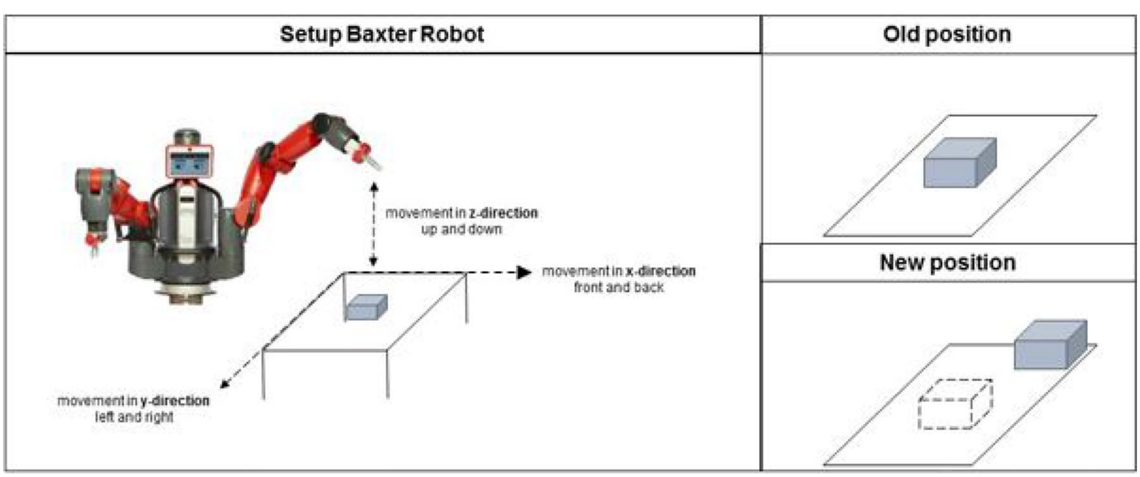


Fig. 9 Description of the pick and place service realized by Baxter robot with an example of the old and new position of the box

Table 1 Comparison among performance results of Baxter robot for object recognition related to the used box and bottle on white and green desks

Object recognition						
BOX		Recognition percentage (%)	BOTTLE		Recognition percentage (%)	Parameter combination
White desk	Green desk		White desk	Green desk		
4/5	5/5	90	5/5	3/5	80	1
4/5	5/5	95	3/5	4/5	70	2
4/5	5/5	90	3/5	5/5	80	3
4/5	5/5	90	3/5	5/5	80	4
3/5	5/5	80	2/5	3/5	50	5
4/5	5/5	90	3/5	4/5	70	6
5/5	5/5	100	3/5	4/5	70	7

Table 2 Comparison among performance results of Baxter robot for object manipulation and those obtained for the presented task

	On the performance of the Baxter research robot 2016 (Cremer et al. 2016)	Grasping of the objects released on the user hand
Error vector		
Mean (mm)	1.4	2.9
Standard deviation (mm)	0.6	1.2

presented. The main index chosen to evaluate the algorithm performances is the norm of the vector that measures the distance between the desired position and that reached by the robotic arm. On the base of this analysis, in this study the data related to 20 tests of object placing have been acquired varying the choice of the object and moving the hand in different positions. The values obtained in our tests did not overcome the threshold of ± 5 mm fixed by the producer, and are compatible with the values reported in the study (Cremer et al. 2016) even if higher than those as it is possible to see in Table 2. For the sake of completeness, a validation study about the estimation performance of the kinect sensor for the joint position has been reported. With respect to the study conducted in Mobini et al. (2013), the RMSE is of 17 mm for a left hand joint pose estimation, realized by Kinect v1, at a distance from 1 to 3 m. An analogous study has been conducted by Mishra et al. (2015) and in our department (Capecci et al. 2016) to test the validation of the Kinect sensor with respect to a gold standard system, namely the Vicon motion capture system. The obtained results are comparable with those presented in the literature.

5 Conclusion and future works

The work, presented in this paper, proposed the development of an architecture for human–robot interaction towards assisted daily living environments. In details, the following three main aspects were faced:

- the development of the simulator for the robot arm in order to test the implemented algorithms;
- the development of a gesture recognition algorithm for the object selection;
- the interaction between the manipulator robot and the user hand with the object transfer from the robot manipulator to the user hand.

All the implemented algorithms have been tested in a simulator environment, namely Gazebo, where also the Kinect sensor, that represented the external camera, was introduced. The results related to the tests in Gazebo for the object recognition and for the picking were satisfying and they were confirmed when the task was reproduced in the laboratory directly from a user. Also the recognition of the gesture and the localization of the hand, where the object is placed, fell into the tolerance limits of the sensor accuracy.

There are some possible improvements to realize:

- in the context of assistive architecture systems, using the Kinect for the gesture recognition could be a limit as in the case of a person, forced on a wheelchair, for the calibration phase. For this step, the person would be indeed asked to stay standing in the well known *Psi* Pose, characterized to have the arms pointing upwards. Moreover, the placing position of the object could present some issues if the subject is sat on a wheelchair;
- the impossibility of representing the subject in the Gazebo simulator limits the study of the interaction in the simulator environment, thus limiting the possibility to test the algorithm related to the gesture recognition only experimentally.

Another aspect for future works is the improvement of the database that contains all the possible objects and that can be enriched with more objects of common use. Moreover, the database could add some information for the Baxter, related to the objects available for the pick and place.

References

Achic F, Montero J, Penaloza C, Cuellar F (2016) Hybrid BCI system to operate an electric wheelchair and a robotic arm for navigation and manipulation tasks. In: 2016 IEEE workshop on advanced robotics and its social impacts (ARSO), pp 249–254. https://doi.org/10.1109/ARSO.2016.7736290

Bekele E, Wade J, Bian D, Fan J, Swanson A, Warren Z, Sarkar N (2016) Multimodal adaptive social interaction in virtual environment (MASI-VR) for children with autism spectrum disorders (ASD). In: IEEE virtual reality (VR), pp 121–130. https://doi.org/10.1109/VR.2016.7504695

Canal G, Escalera S, Angulo C (2016) A real-time human-robot interaction system based on gestures for assistive scenarios. Comput Vis Image Underst 149:65–77. https://doi.org/10.1016/j.cviu.2016.03.004 **(special issue on Assistive Computer Vision and Robotics - Assistive Solutions for Mobility, Communication and HMI)**

Capecci M, Ceravolo M, Ferracuti F, Iarlori S, Longhi S, Romeo L, Russi S, Verdini F (2016) Accuracy evaluation of the kinect v2 sensor during dynamic movements in a rehabilitation scenario, vol 2016, pp 5409–5412. https://doi.org/10.1109/EMBC.2016.7591950

Cavallo F, Aquilano M, Bonaccorsi M, Limosani R, Manzi A, Carrozza MC, Dario P (2013) On the design, development and experimentation of the astro assistive robot integrated in smart environments. In: 2013 IEEE international conference on robotics and automation, pp 4310–4315. https://doi.org/10.1109/ICRA.2013.6631187

Chance G, Camilleri A, Winstone B, Caleb-Solly P, Dogramadzi S (2016) An assistive robot to support dressing - strategies for planning and error handling. In: 6th IEEE international conference on biomedical robotics and biomechatronics (BioRob), pp 774–780. https://doi.org/10.1109/BIOROB.2016.7523721

Chen L, Wei H, Ferryman J (2013) A survey of human motion analysis using depth imagery. Pattern Recogn Lett 34:1995–2006. https://doi.org/10.1016/j.patrec.2013.02.006 **(smart Approaches for Human Action Recognition)**

Clodic A, Pacherie E, Alami R, Chatila R (2017) Key elements for human–robot joint action. Springer, Berlin, pp 159–177. https://doi.org/10.1007/978-3-319-53133-5_8

Coupeté E, Moutarde F, Manitsaris S (2015) Gesture recognition using a depth camera for human robot collaboration on assembly line. Proced Manuf 3:518–525. https://doi.org/10.1016/j.promfg.2015.07.216 **(6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) and the Affiliated Conferences, AHFE 2015)**

Cremer S, Mastromoro L, Popa DO (2016) On the performance of the baxter research robot. In: 2016 IEEE international symposium on assembly and manufacturing (ISAM), pp 106–111. https://doi.org/10.1109/ISAM.2016.7750722

Ding IJ, Chang YJ (2017) Hmm with improved feature extraction-based feature parameters for identity recognition of gesture command operators by using a sensed kinect-data stream. Neurocomputing 262:108–119. https://doi.org/10.1016/j.neucom.2016.11.089 **(online Real-Time Learning Strategies for Data Streams)**

D'Onofrio G, Fiorini L, de Mul M, Fabbricotti I, Okabe Y, Hoshino H, Limosani R, Vitanza A, Greco F, Giuliani F, Guiot D, Senges E, Kung A, Cavallo F, Sancarlo D, Greco A (2018) Agile co-creation for robots and aging (accra) project: new technological solutions for older people. Eur Geriatr Med 9(6):795–800. https://doi.org/10.1007/s41999-018-0106-7

Feil-Seifer D, Matarić M (2005) Defining socially assistive robotics. Robotics 2005:465–468. https://doi.org/10.1109/ICORR.2005.1501143

Foresi G, Freddi A, Monteriù A, Ortenzi D, Pagnotta DP (2018) Improving mobility and autonomy of disabled users via cooperation of assistive robots. In: IEEE international conference on consumer electronics (ICCE), pp 1–2. https://doi.org/10.1109/ICCE.2018.8326291

Foresi G, Freddi A, Iarlori S, Longhi S, Monteriù A, Ortenzi D, Proietti Pagnotta D (2019) Human–robot cooperation via brain computer interface in assistive scenario. In: Casiddu N, Porfirione C, Monteriù A, Cavallo F (eds) Ambient assisted living. Springer, Berlin, pp 115–131

Fujii T, Lee J, Okamoto S (2014) Gesture recognition system for human–robot interaction and its application to robotic service task. Lect Notes Eng Comput Sci 2209:63–68

Le HH, Loomes MJ, Loureiro RCV (2016) Group interaction through a multi-modal haptic framework. In: 2016 12th international conference on intelligent environments (IE), pp 62–67. https://doi.org/10.1109/IE.2016.18

Lee HR, Tan H, Šabanović S (2016) That robot is not for me: addressing stereotypes of aging in assistive robot design. In: 25th IEEE international symposium on robot and human interactive communication (RO-MAN), pp 312–317. https://doi.org/10.1109/ROMAN.2016.7745148

Mendes N, Ferrer J, Vitorino J, Safeea M, Neto P (2017) Human behavior and hand gesture classification for smart human-robot interaction. Proced Manuf 11:91–98. https://doi.org/10.1016/j.promfg.2017.07.156 **(27th international conference on flexible automation and intelligent manufacturing, FAIM2017, 27–30 June 2017, Modena, Italy)**

Mettel MR, Alekseew M, Stocklöw C, Braun A (2019) Designing and evaluating safety services using depth cameras. J Ambient Intell Humaniz Comput 10:747–759. https://doi.org/10.1007/s12652-018-0898-9

Mishra A, Skubic M, Abbott C (2015) Development and preliminary validation of an interactive remote physical therapy system, vol 2015. https://doi.org/10.1109/EMBC.2015.7318332

Mobini A, Behzadipour S, Foumani M (2013) Accuracy of kinect's skeleton tracking for upper body rehabilitation applications. Disabil Rehabi Assist Technol. https://doi.org/10.3109/17483107.2013.805825

Morales Y, Miyashita T, Hagita N (2017) Social robotic wheelchair centered on passenger and pedestrian comfort. Robot Auton Syst 87:355–362. https://doi.org/10.1016/j.robot.2016.09.010

Poncela A, Coslado F, García B, FernÁndez-Carmona M, Ariza J, Peinado G, Demetrio C, Sandoval F (2018) Smart care home system: a platform for eassistance. J Ambient Intell Human Comput. https://doi.org/10.1007/s12652-018-0979-9

Pons P, Jaen J (2019) Interactive spaces for children: gesture elicitation for controlling ground mini-robots. J Ambient Intell Human Comput. https://doi.org/10.1007/s12652-019-01290-6

Quigley M, Conley K, Gerkey B, Faust J, Foote T, Leibs J, Wheeler R, Ng A (2009) Ros: an open-source robot operating system, vol 3

Raheja JL, Chandra M, Chaudhary A (2018) 3d gesture based real-time object selection and recognition. Pattern Recogn Lett 115:14–19. https://doi.org/10.1016/j.patrec.2017.09.034**(multimodal Fusion for Pattern Recognition)**

Romer GRBE, Stuyt HJA, Peters A (2005) Cost-savings and economic benefits due to the assistive robotic manipulator (arm). In: 9th international conference on rehabilitation robotics, 2005. ICORR 2005, pp 201–204. https://doi.org/10.1109/ICORR.2005.1501084

Sarcevic P, Kincses Z, Pletl S (2019) Online human movement classification using wrist-worn wireless sensors. J Ambient Intell Humaniz Comput 10:89–106. https://doi.org/10.1007/s12652-017-0606-1

Sciutti A, Sandini G (2017) Interacting with robots to investigate the bases of social interaction. IEEE transactions on neural systems and rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society. https://doi.org/10.1109/TNSRE.2017.2753879

Tellaeche A, Kildal J, Maurtua I (2018) A flexible system for gesture based human-robot interaction. Proced CIRP 72:57–62. https://doi.org/10.1016/j.procir.2018.03.017 **(51st CIRP Conference on Manufacturing Systems)**

Tsui K, Yanco H, Kontak D, Beliveau L (2008) Development and evaluation of a flexible interface for a wheelchair mounted robotic arm. In: 3rd ACM/IEEE international conference on human–robot interaction (HRI), pp 105–112. https://doi.org/10.1145/1349822.1349837

Wen Z, Liu D, Liu X, Zhong L, Lv Y, Jia Y (2019) Deep learning based smart radar vision system for object recognition. J Ambient Intell Humaniz Comput 10(3):829–839. https://doi.org/10.1007/s12652-018-0853-9