IVO Robot: A New Social Robot for Human-Robot Collaboration

Javier Laplaza
IRI (CSIC-UPC)
Barcelona, Spain
jlaplaza@iri.upc.edu

Nicolás Rodríguez
IRI (CSIC-UPC)
Barcelona, Spain
nrodriguez@iri.upc.edu

J. E. Domínguez-Vidal
IRI (CSIC-UPC)
Barcelona, Spain
jdominguez@iri.upc.edu

Fernando Herrero
IRI (CSIC-UPC)
Barcelona, Spain
fherrero@iri.upc.edu

Sergi Hernández
IRI (CSIC-UPC)
Barcelona, Spain
shernand@iri.upc.edu

Alejandro López
IRI (CSIC-UPC)
Barcelona, Spain
alopez@iri.upc.edu

Alberto Sanfeliu
IRI (CSIC-UPC)
Barcelona, Spain
sanfeliu@iri.upc.edu

Anaís Garrell
IRI (CSIC-UPC)
Barcelona, Spain
agarrell@iri.upc.edu

Abstract—We present a new social robot named IVO, a robot capable of collaborating with humans and solving different tasks. The robot is intended to cooperate and work with humans in a useful and socially acceptable manner to serve as a research platform for long-term Social Human-Robot Interaction. In this paper, we proceed to describe this new platform, its communication skills and the current capabilities the robot possesses, such as, handing over an object to or from a person or performing guiding tasks with a human through physical contact. We describe the social abilities of the IVO robot, furthermore, we present the experiments performed for each robot's capacity using its current version.

Index Terms—Human Robot Interaction, Social Robots, Human Robot Collaboration
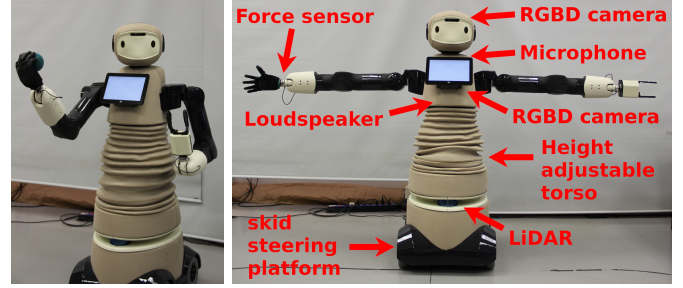
Fig. 1. **IVO robot.** *Left:* Front picture with the robot grabbing a rubber ball. *Right:* Robot with its arms stretched out.

I. INTRODUCTION

More and more robots are gradually leaving the laboratories and factories and starting to populate the streets of our cities. A trend that, far from reversing, is set to accelerate in the coming years. Examples of robots in University environments, museums or hospitals [1]–[3] are becoming increasingly common. This means that robots must be prepared to actually interact with the people they encounter in real operating situations without the typical constraints of laboratory experiments or controlled factory environments.

In our case, the interests of our research group range from the development of robust algorithms for navigation in urban environments, such as social accompaniment of humans [4], [5], and even handover tasks among others [6]. Because of this, we need a multipurpose robot capable of working properly in multiple scenarios. We intend to make the interactions of the robot with the human as natural as possible, and acting according to the established social conventions, so we need

the robot to have a humanoid appearance and to have several ways to communicate with the human.

For that reason, we present the new IVO robot, Fig. 1, with a friendly humanoid external appearance to increase the naturalness with which any human interacts with it [7], as well as elements which will be needed to successfully perform the tasks which our research group wants to do with it.

Our future work with the IVO robot will involve field trials with the robot in its actual context of use. Such research poses interaction and human-robot collaboration challenges. On the one hand, the robot should be robust in order to autonomously perform its goals and to interact with multiple people for large periods of time. On the other hand, to test several approaches of human-robot collaboration over time, researchers should be able to modify conditions of the robot's physical aspect and attitudes. As it has been stated previously, we are particularly interested in collaboration between humans and robots in a social accepted manner.

In the remainder of the paper, the context of use is presented in Sec. II. Sec. III explains the collaborative tasks the robot is capable to perform. Finally, conclusions are given in Sec. IV.

II. CONTEXT OF USE

Advances in robotics are targeted at specific populations – elderly people, those with some disabilities, and others. We

Item	Measurement	Item	Measurement
Height (maximum)	1.55 m	Platform width	0.49 m
Height (minimum)	1.20 m	Platform length	0.70 m
Arm length	1.00 m	Wheelbase	0.41 m
Arm (end-effector)	1.10 m	Hand payload	1.00 kg
Palm to gripper	2.40 m	Gripper payload	1.20 kg

desire to design robots that can interact and collaborate with non-trained volunteers. To accomplish that, we are interested in how a robot can perform different collaborative tasks.

As it has been mentioned, we decide to work on a robot that would collaborate with humans. These tasks include handing over an object to a person, guiding the robot through physical contact and communication with humans using gestures.

In this vein, it is assumed that the height of the robot should be adjustable to accommodate people of different heights. In addition, it is considered that the robot should possess both a robotic hand and a gripper in order to perform manipulation tasks which require both dexterity and a good grip. Finally, it is concluded that the robot should possess several ways of communicating with the human with whom the robot is interacting. Therefore, it was decided to provide the robot with both a speaker with which to express itself using natural language, and a touch screen on the torso through which being able to maintain a visual-based communication.

Furthermore, the robot is provided with two robotic arms (7 degrees of freedom each). One of them ends in a robotic hand with 5 fingers of 3 links each. The other ends in a gripper. Both end-effectors have force/torque sensors at the wrists to be able to detect the possible exchange of physical forces which occur in manipulation tasks. As for navigation, a skid-steering type platform is chosen to allow the robot to move in both indoor and outdoor environments. In this sense, a 16-channel LiDAR is added at the front, a 2D laser at the rear and an RGBD camera on the torso to detect possible obstacles which the robot may encounter along the way. Additionally, another RGBD camera is added on the forehead that can be used both to detect obstacles and to detect the human.

Table I summarizes the robot's main dimensions to be taken into account. The adjustable height allows it to adapt to the human it is interacting with, as well as being able to manipulate with both arms up to heights greater than $2.0\ m$. The payload of each arm allows it to handle light tools as well as to deliver or receive small objects. Finally, the wheelbase allows it to navigate at moderate speeds (up to $1.0\ m/s$).

III. IVO'S COLLABORATIVE TASKS

In this section, we proceed to describe the three main capabilities of the IVO robot we are working on at the moment.

A. Handover

IVO's dual arm configuration can be used to perform handovers between a human and IVO. This handover operation can be used to interchange objects between robots and humans,

such as tools, food, or any object required by the partner to complete a specific task. There are usually two agents in the handover operation, the *giver* and the *receiver*. The *giver* goal is to deliver a certain object to the *receiver*. The handover task is finished when the object is in the hand of the *receiver* and the *giver* is no longer in contact with the object.

The robot can use the left gripper when it has to play the *giver* role, since the gripper provides a safer fixation point than the right robot hand. On the contrary, the right robot hand can be used when the robot plays the *receiver* role.

1) Experiments: We worked previously on handovers using IVO, [6], IVO was used to collect data of humans performing handovers with the robot. These humans were filmed using the RGBD camera inside IVO's head. The 2D skeleton of the human is extracted using OpenPose [8], and then the 3D pose is computed using the depth information of the camera, computing the distance to each joint located in the image.

The dataset was created in order to train a deep neural network able to predict the T future motion frames of a human during the handover task given the N previous frames. More specifically, this predictor is able to predict up to $2.5\ s$ of human's future trajectory using the previous $5.0\ s$.

We are currently working on incorporating this prediction as a conditioning factor of the IVO's navigation planner in a similar way as Koppula et al. [9] did. In that work, authors calculated object affordances to predict each possible future using Anticipatory Temporal Conditional Random Fields (ATCRFs) and, with this prediction, conditioned the actions that the robot could or could not do. In our case, we plan to use the person's motion prediction to estimate the delivery point so that the robot can start planning up to $2.5\ s$ in advance in line with the work done by Huang et al. [10].

Not only that, but we also hope to make the robot capable of reacting earlier to variations in the human's trajectory, thus, avoiding abrupt stops in its movement. This is inspired by [11] in which they incorporated a prediction of the motion of nearby humans as an input to their navigation planner to avoid the "freezing robot" problem.

B. Guiding Robot

The arms have 7-DOF manipulator, having a 6-DOF force-torque sensor between the wrist joint and the end-effector.

If used appropriately, the F/T sensor can offer some interesting human-robot interaction capabilities to offer more natural ways of interacting with IVO, specially in terms of direct physical interaction (DPI) using the F/T measures to move IVO mobile base and the manipulator joints to adjust the grip with the human that is exerting the force.

DPI has been widely studied in robotics applications as a way of increase user experience in a human-robot interaction. Following the reasoning in [12], IVO with its double manipulator configuration can be an excellent example to test DPI effectiveness, intuitiveness and safety.

We implemented in IVO a kinematic controller with grip adjustment that is based on [13]. Our approach used force measures projecting the force into the robot's base frame. The
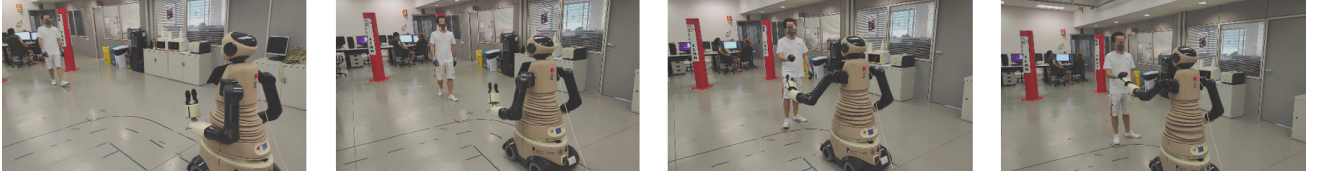
Fig. 2. **Handover task:** Example of a sequence recorded for the dataset. The robot moves forward and raises its left arm to take the object from the human, who is also walking towards the robot. In this case, the robot detects its human counterpart using the RGBD camera on its head.

Fig. 3. **Guiding the robot task:** Sequence of movements turning the robot using physical contact. The robot is controlled using only the F/T sensor.

projected force into the plane XY will define the linear and angular speed of the robot, whereas Z projected force will serve to regulate the vertical position of the torso.

We could not directly implement a system that included the gravity and friction compensation, because IVO's manipulators have an embedded controller that is integrated inside its operative system. Therefore, we chose an strategy to build over this built-in controller our guiding controller.

Torque information was used to adjust the grip of the human by moving the 3-DOF wrist. The full movement of the joints is limited to a $[-\frac{\pi}{6}, \frac{\pi}{6}]$ range from the home position. The maximum angle limitation is a simple technical solution to avoid mechanical demanding configurations in which an external force can provoke some damage over the joints of the arm, specially joints overheating.

However, this limitation has a further purpose. The aim of arm movements for this application is to compensate small changes at the human side, to increase the comfort in the interaction. A bigger range can even increase the complexity of the interaction, increasing robot response and inducing the human to continuously adapt the fit with IVO's end-effector.

Fig. 4. **Guiding controller algorithm:** Diagram of the implemented algorithm for the guiding controller.

A common approach for the controller design is to translate forces into kinematic commands that will vary the position in the end [14]. For our application we have implemented a controller that will consider non-zero forces or torques as an impulse for the robot base or the joints. Kinematic command for the base is computed using a PD controller whose gains will set the magnitude transformation between force and speed.

In spite of the known faults of PD controllers [15], being sensitive to noise, this approach was chosen to increase global system response before fast variations introduced by the human. Hence, the robot will be able to power up its acceleration or deceleration when the human is trying to modify its speed.

Noise overall effect over the system is damped by a physical and software filter. The physical filter is introduced by the human hand, that naturally adapts to small changes. On the other side, software filter is introduced in the normalization phase, neglecting small effects by reducing the numerical precision after normalizing the measure in the range $[-1, 1]$.

Extending the algorithm description, torso and arm controllers implement a simple proportional controller. In contrast with base movement, both arm and torso movement is aimed to adjust the fit with the hand. This adjustment can be uncomfortable for the human if the movements are fast or frequent. Therefore, a derivative term that will increase the frequency of the movements is undesirable to our prior objective.

The general logic scheme followed in the algorithm implementation can be seen in Fig. 4.

1) Experiments: In order to test IVO's guiding system, we decided to use two experimental cases to evaluate its performance: *(i)* Open field guiding; and *(ii)* guiding through a door. These scenarios will help to test system maneuverability and its precision.

a) Open field experiment: The set for this case scenario was an open space at a laboratory with no obstacles. The executed movement was a rotation of the robot, combining it with a small linear displacement so that the eccentricity corresponded with an ellipse.
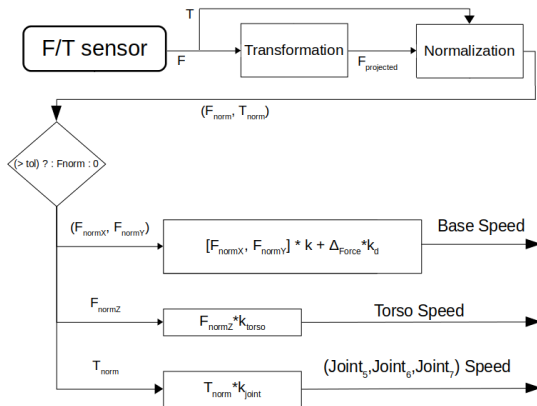
Fig. 5. **Accompanying the robot passing through a door:** Sequence of movements with the human guiding the robot through the door.
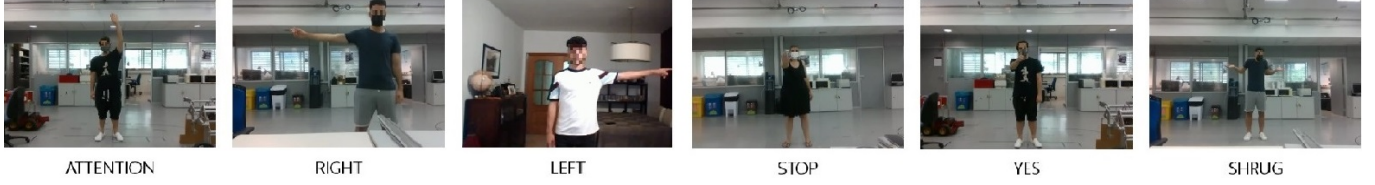

Fig. 6. **Natural non-verbal communication:** Samples of gestures recorded for the Body Gesture dataset.

At Fig. 3, we can see a sequence of pictures that show the temporal evolution of the turning. This sequence should be read from left to right.

We can see using floor marks that the turning was performed at a small space, letting the person turn the robot without moving from the focus point. Hence, system maneuverability could be tested.

b) Door experiment: In contrast with the previous experiment, in this one we will try to check the precision of IVO's response when small adjustments are being performed in the guiding, so that it can pass through an evacuation door.

The width of the door is approximately of 110 cm, whereas the human and IVO together have a width of about 140 cm. Therefore, the person should lead the way ahead of the robot while the robot arm allows the needed angle adjustment so that the grip can be considered natural.

The sequence of movements can be seen at Fig. 5. In the sequence we can see how the human progressively moves ahead of the robot while the wrist adjust the angle of the end-effector to increase the comfort in the interaction.

C. Communication skills

One of the most important skills for a robot during a cooperation task is communication. The robot needs to clearly identify the message being sent by the human in order to answer with the proper feedback.

Humans can communicate with other humans using different channels. The most used communication channel is voice, so we will split the different channels in two main families:

1) Verbal: This is the most explicit communication channel between humans. IVO has a microphone and speakers installed, so it is able to understand human voices, but also to provide audio feedback to the human partner.
2) Non-Verbal: Although non-verbal communication is more implicit, studies point that about 70% of communication between humans is considered non-verbal [16]. We will enumerate some of the channels that IVO can exploit to communicate under this branch:
(a) Body gestures, *(b)* Sign language, *(c)* Facial expression, and *(d)* Written expression

1) Experiments: So far, we have only explored the communication reach of the body gesture. We created a dataset of humans performing different communication gestures in the most natural way. We could not find any similar dataset, since we found that other body gesture datasets were too specialized in a certain task or they were designed for very different platforms, as drones. In our dataset, we defined the following gestures: Attention, Greeting, Move Right, Move Left, Stop, Continue, Turn Back, Move Back, Come, Yes, No and Shrug.

It is worth mentioning that during the creation of the dataset, there was no guidelines to perform the gestures. The non-trained volunteers were only told to express each order or idea to the robot, letting them decide what was the most natural way to provide the information with their body. The descriptions in the list were only an average description of the gestures recreated by the human volunteers.

This dataset was then used to train a very simple neural network in order to classify the gestures (in our first approach, we only used "static" gestures with no motion: attention, left, right, stop, yes, shrugh, random and static). The classifier output was then fed to IVO, which had to answer by moving its head towards a certain direction or opening or closing its hand according to the gesture.

IV. CONCLUSIONS

In this paper, we have presented the IVO robot, a new social robot designed to collaborate with humans, and capable of performing long-term interactions. We introduced three overarching tasks for the development of this robot: handover task, guiding the robot through physical contact and non-verbal communication skills.

Our future research will evaluate the success of the design in relation to our collaborative goals through extensive field tests to understand change in environments, volunteers, and physical and social context. Moreover, we are currently working on enhancing the aforementioned communication skills of the robot, expressing feelings and understanding people's emotions.

REFERENCES

[1] Maria Kyrarini, Fotios Lygerakis, Akilesh Rajavenkatanarayanan, Christos Sevastopoulos, Harish Ram Nambiappan, and Kodur Chaitanya. A Survey of Robots in Healthcare. *Technologies*, 9:8, 2021.

[2] Josef Guggemos, Sabine Seufert, and Stefan Sonderegger. Humanoid robots in higher education: Evaluating the acceptance of Pepper in the context of an academic writing course using the UTAUT. *British Journal of Educational Technology*, 51, 2020.

[3] Takamasa Iio, Satoru Satake, Takayuki Kanda, Kotaro Hayashi, Florent Ferreri, and Norihiro Hagita. Human-Like Guide Robot that Proactively Explains Exhibits. *International Journal of Social Robotics*, 12(2):549–566, 2020.

[4] Anais Garrell, Luis Garza-Elizondo, Michael Villamizar, Fernando Herrero, and Alberto Sanfeliu. Aerial social force model: A new framework to accompany people using autonomous flying robots. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 7011–7017. IEEE, 2017.

[5] Alex Goldhoorn, Anaís Garrell, René Alquézar, and Alberto Sanfeliu. Searching and tracking people in urban environments with static and dynamic obstacles. *Robotics and Autonomous Systems*, 98:147–157, 2017.

[6] Javier Laplaza, Albert Pumarola, Francesc Moreno-Noguer, and Alberto Sanfeliu. Attention deep learning based model for predicting the 3D Human Body Pose using the Robot Human Handover Phases. *30th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 161–166, 2021.

[7] Min Kyung Lee, Jodi Forlizzi, Paul E. Rybski, Frederick Crabbe, Wayne Chung, and Josh Finkle. The Snackbot: Documenting the Design of a Robot for Long-Term Human-Robot Interaction. In *Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction (HRI)*, page 7–14, 2009.

[8] Zhe Cao, Gines Hidalgo, Tomas Simon, and Shih-En Wei. OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(1):172–186, 2021.

[9] Hema S. Koppula and Ashutosh Saxena. Anticipating Human Activities Using Object Affordances for Reactive Robotic Response. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(1):14–29, 2016.

[10] Chien-Ming Huang and Bilge Mutlu. Anticipatory robot control for efficient human-robot collaboration. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 83–90, 2016.

[11] Adarsh Jagan Sathyamoorthy, Jing Liang, Utsav Patel, Tianrui Guan, Rohan Chandra, and Dinesh Manocha. DenseCAvoid: Real-time Navigation in Dense Crowds using Anticipatory Behaviors. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 11345–11352, 2020.

[12] Tiffany L. Chen and Charles C. Kemp. Lead me by the hand: Evaluation of a direct physical interface for nursing assistant robots. In *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 367–374, 2010.

[13] François Ferland, Arnaud Aumont, Dominic Létourneau, and François Michaud. Taking your robot for a walk: Force-guiding a mobile robot using compliant arms. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 309–316, 2013.

[14] Zijian Wang and Mac Schwager. Kinematic multi-robot manipulation with no communication using force feedback. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pages 427–432, 2016.

[15] Lin Cai, A.B. Rad, W.L. Chan, and Kai-Yuan Cai. A robust fuzzy PD controller for automatic steering control of autonomous vehicles. In *12th IEEE International Conference on Fuzzy Systems (FUZZ '03)*, volume 1, pages 549–554, 2003.

[16] Tim O'Sullivan, John Hartley, Danny Saunders, Martin Montgomery, and John Fiske. *Key concepts in communication and cultural studies*. Routledge, UK, 1994.