INTRO: A Multidisciplinary Approach to Intelligent Human-Robot Interaction

Aleksandar Jevtić, Eric Lucet
Robosoft S.A.
Bidart, France
{aleksandar.jevtic, eric.lucet}@robosoft.fr

Alex Kozlov, Jeremi Gancet
Space Applications Services
Zaventem, Belgium
{alex.kozlov, jeremi.gancet}@spaceapplications.com

Abstract— **This paper presents the interactive robotics concept being developed by the INTRO research network. The aim is to create a new generation of intelligent mobile robots that operate in close interaction with humans in unstructured, dynamically changing environments. The INTRO network consists of a team of researchers, from academia and industry, which create a multidisciplinary framework that entails Cooperative Robot Learning, Cognitive Human-Robot Interaction (HRI), and Intelligent Interface Design. The robotic system being developed will be tested in two application scenarios: the Robot Waiter, and the Urban Search and Rescue (USAR). For these scenarios, two different robotic platforms are used in the implementation stage. This paper presents an overview of the obtained research objectives, and proposes a framework for the integration of work and the implementation of the expected results. Finally, the paper describes a potential impact through development and use of research results and proposes future lines of research.**

Keywords – human-robot interaction; learning; human-robot interface; manipulator arm; robot safety; mobile robots; autonomous robots.

I. Introduction

Human-Robot Interaction (HRI) is an interdisciplinary study of interaction dynamics between humans and robots. In recent years, this field has received attention from a wider scientific community and it has been greatly supported by both public and private funding. Robots have become part of our everyday lives; they clean our homes, work as restaurant waiters, perform monitoring and surveillance, assist in search and rescue missions, explore outer space, and so forth. These robots are usually designed for a specific mission, and their interaction with humans is limited. They are built to perform simple tasks that do not involve human-robot cooperation.

The aim of the INTRO research network is to create a new generation of intelligent mobile robots that operate in close interaction with humans in unstructured, dynamically changing environments. The proposed interactive robotics concept aims to contribute to the areas of Cooperative Robot Learning, Cognitive HRI, and Intelligent Interface Design, and poses a big challenge in terms of integration of results and their implementation on a single robotic platform. The resulting technological developments should provide more intuitive HRI and scalability in terms of a wider range of applications.

II. State of the Art

Various scientific disciplines contribute to the field of HRI, which brings collaboration of scientists and engineers of different professional backgrounds. HRI combines research from Computer Science and Engineering, but also human sciences, including cognitive, behavioural, and social sciences [1]. Great research efforts are being made because of high expectations and potentially large impact that HRI may have on quality of living. Robots could replace humans in carrying out work that is physically hard, dangerous, or involves repetitive and boring tasks. Applications in which the cooperation between humans and robots is expected are even greater challenges.

HRI is an essential part of robot-assisted applications. The most popular interaction method for hazardous environments is through some form of remote teleoperation or telepresence [2, 3]. This has the obvious advantage of not directly endangering personnel in hazardous areas. The main aim is to achieve remote perception and situational awareness. Another interaction paradigm is direct, proximate interaction and human-robot collaboration (HRC) [4, 5]. In this case interaction capabilities include human gesture recognition, voice recognition and natural language interpretation. Functionalities for useful human-robot collaboration include cognitive spatial reasoning, perspective taking, and behaviours for sequencing and executing high level tasks. Dynamic autonomy is also expected, which would allow the robot to dynamically adjust its behaviours depending on the situation.

Natural and efficient human-robot collaboration is crucial to advancing robotic applications. One significant challenge that is common in any form of HRC is synchronization of actions between human and robot [6, 7]. Current HRC is fairly limited to very rigid and structured turn based interaction. Research in timing-based HRC aims to improve the fluency and efficiency of joint HRC making the interactions more natural, coordinated and effective [8]. Joint attention is also an

This work has been financed by the EU funded Initial Training Network (ITN) in the Marie-Curie People Programme (FP7): INTRO (INTeractive RObotics research network), grant agreement no.: 238486

important function for effective human-robot collaboration [9]. Joint attention requires shared visual attention, attention detection, attention manipulation and intention recognition. For visual attention visual saliency detection is needed; this is done by detecting faces and movements. Pointing gesture generation is a highly suitable method of attention manipulation, used by the robot to direct the attention of the interacting partner. For attention detection applicable methods include eye gaze following and pointing gesture recognition; this is used by the robot to understand what the partner's attention is focused on.

Improved robot performance can be achieved through learning and adaptive control techniques. Learning from Demonstration (LfD) [10, 11] is an approach to robot learning through demonstrations of desired behaviours. The concept is to map desired robot actions (policies) from examples or demonstrations. The robot then uses learned policies to execute appropriate actions based on the current world state, to reproduce demonstrated behaviours. LfD can be applied to different levels of robot control, from low-level motion control to complex high-level behaviours. Another important kind of learning is self-exploration through motor-babbling, in order to acquire self-awareness, learn body maps and autonomously obtain control and movement capabilities [12].

When interacting with humans, the robot should provide application-relevant information and feedback on its operational state. This is done through interface, but also through actions that the robot can perform and human can intuitively understand [13]. Designing appropriate interface for an autonomous robotic system requires an understanding of the users and of their expectations of such a system [14]. System scalability in terms of different user profiles (age, gender, cultural background, technical education, etc.) is one of the goals for development of intuitive interface.

Safety may be considered the main obstacle for a larger involvement of robots in human society. Different standards have been proposed to guarantee safe presence of robots in our daily activities. One approach is to create intrinsically safe robot hardware structure and the other addresses the limitations in robot's autonomy. Safety requirements become more rigorous for applications that entail close cooperation between humans and robots, especially when robots are expected to manage unplanned interactions with initially unknown environmental features [15].

In the HRI system development several factors need to be taken into account. Multi-modal perception is needed to create knowledge about the robot's environment and humans it interacts with. This involves data fusion from different types of sensors and real-time signal processing. Main HRI inputs are vision and speech [16]. Autonomous mobile robots also use low-level sensing that permits them to detect obstacles and safely navigate in unstructured environments. Sensors that measure physiological signals, such as heart rate or blood pressure, could provide insight into the emotional state of a user. Input from the sensors is important to create awareness about the robot's environment but in order to provide more natural, human-like interaction fast signal processing is a must.

III. Multidisciplinary Approach to HRI

A. *Problem statement*

The study of HRI contains a wide variety of challenges, some of them related to the concepts general to HRI, and others dealing with specific use of robot systems that interact with humans in particular application domain. Intuitive interaction, interface design, and end-user's expectations are relevant factors to be considered for a wider acceptance of robots in our society. Technical challenges include sensor data fusion and real-time signal processing. Learning and adaptability can improve robot's performance and allow it to perform a greater variety of tasks. Moreover, HRC can improve joint efficiency. These are some of the main challenges in HRI that INTRO project addresses, which will be evaluated in two application scenarios.

B. *Application scenarios*

1) *Robot Waiter scenario*

The goal of this scenario is to test the robot in the role of a waiter in close interaction with humans placed in a dynamically changing environment of a bar. In order to achieve an efficient and human-like interaction and cooperation, the robot is expected to have a high degree of autonomy, intelligent interface, high-level sensing abilities, safe manipulator arm, and to be capable of pattern recognition and knowledge extraction in order to learn about its environment.

In the basic scenario, Robot Waiter takes orders and delivers drinks to the customers that are randomly seated at the tables. The robot detects a raised hand at a table from the far end of the bar, which is recognized as a call to place an order. The robot estimates the position of the customer and computes the optimal path on the preloaded map of the bar. It then safely navigates between the tables to approach the costumer. Face and eye-gaze direction are detected in order to recognize customer's attention. The facial features are stored in the robot's database for future identification. The customer places an order using speech recognition or the robot-mounted touch screen. Upon taking the order, the robot safely navigates to the kitchen. The bartender prepares the drink found in the robot's order list and hands it out to the robot that uses the manipulator arm and the gripper to take the drink and puts it on the mounted tray. The robot takes the ordered drink to the previously stored table location. It recognises the customer's face and announces that the drink is ready. It again detects the customer's attention and safely hands out the drink using its arm and gripper.

2) *Urban Search and Rescue scenario*

The premise of the scenario is that a team of human USAR agents and a mobile robot assistant have been deployed to search for victims (scope is constrained to a single human and the robot). The human and the robot navigate the ruins while searching for victims to extract. Detailed tasks of the scenario, in execution order, are as follows:

- The robot uses visual saliency detection to continuously search for victims' faces and/or movement during exploration.
- Upon successful detection, the robot generates a pointing gesture with its manipulator, in the general direction of the detected victim. The purpose of this is to inform the human counterpart of the detection.
- An inherent assumption in this scenario is that the victim is partially buried under a pile of small pieces of brick, concrete and other rubble, i.e. lightly trapped.
- The robot is brought in the vicinity of the desired position for the extraction task. Using intention recognition, the scenario continues in one of two alternative ways depending on the intention recognition cue:
    - Intention recognition cue: the human picks up a piece of rubble and offers it to the robot. Recognising the intention the robot picks up the rubble, moves it and deposits it to a suitable clearing. The robot then repositions itself ready to pick up the next piece.
    - Intention recognition cue: the human directs the robot with a pointing gesture. The robot then begins clearing out a particular area of the rubble independently (the area pointed at). The difference is that instead of being passed the rubble from the human, it independently works on its own area of the rubble while the human works on a different area.
- Continue clearing the rubble until it is removed and the victim is uncovered. Continue finding and extracting victims until the mission is terminated.

C. *Objectives of INTRO*

INTRO project targets the development of interactive robotic concept comprising the three main sub-fields of Cooperative Robot Learning, Cognitive HRI and Intelligent Interface Design. The work has been organized in smaller-scale projects that address the following topics:

- Learning inspired by cognitive psychology.
- Sequence learning and imitation.
- Behaviour and intention recognition.
- Robotic gesture analysis.
- Emotional interaction.
- Robot safety and interaction.
- Failure detection and recovery.
- Interface design.

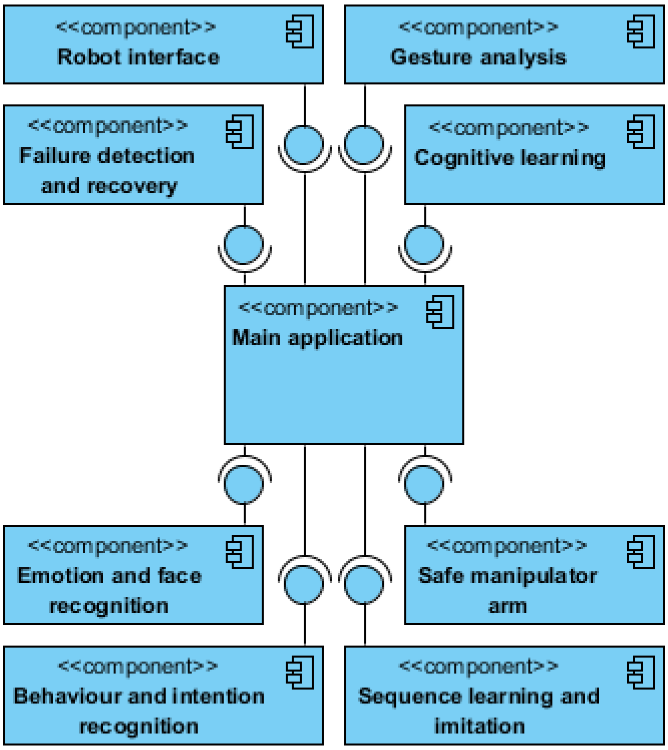

Figure 1. Generic system component diagram

IV. Integration and Expected Results

The research results from the above-mentioned projects will be integrated into two described application scenarios. Two different robotic platforms will be used for the integration and implementation of hardware and software components. A modular system architecture as shown in fig. 1 is proposed to allow reusability of the components. The selection of components will be made with respect to each scenario's requirements.

A. *System components*

Main Application: This component is the central module that integrates individual components functionalities and manages the communication between the components and existing robot services (developed in robuBOX [17] or ROS [18]). The communication is represented with the provided interfaces on the components' side (ball symbol), and required interfaces on the Main Application side (cup symbol). Nevertheless, some components will require input from other components and the communication channels into those components will exist as well. The direct communication between the components will also be possible, and this will be considered in the implementation stage taking into account the system's performance and easier integration of work.

Cognitive learning: This component deals with time-critical tasks. Its objective is to ensure smooth operation of the manipulator arm and the gripper during human-robot collaborative actions. The movement of the robot arm should be perceived as natural by the human user.

Figure 2. RobuLAB10 mobile robotic platform

Sequence learning and imitation: This component will use high-level and low-level controllers to incrementally teach the robot new skills from demonstrations. The component acquires information from high-level sensors (e.g. low-level behaviour, emotion, gesture, face), and forms a high-level context by means of a semantic network. This will allow the robot to learn and perform high-level behaviours by combining a set of parametrised behaviours into sequenced behaviour. The same mechanism will also be used to monitor robot's cognitive state.

Behaviour and intention recognition: This component consists of two modules: Behaviour Recognition and Intention Recognition. Behaviour Recognition module will provide a mechanism for selecting the action performed by the human from a set of learned actions in the database. This refers to motor behaviours involving an object such as approaching an object, moving an object, pointing at an object, and for selecting the target object of the current action, in case of multiple objects in the scene. Inputs of the behaviour recognition module are: position of the hand of the person interacting with the robot, and position of the objects in the scene. Output is the ID of the most probable action performed on the most probable target. Intention Recognition module will be implemented for recognition of motivations and desires of an interacting human. Such a module will represent the mental state of the person and the ID of the behaviour that the person is performing. It is important to distinct the behaviour and intention. Behaviour refers to the motor action and intention refers to the "goal" or the desire of the person, i.e. "why" the person is performing that motor action.

Gesture analysis: Three modules make this component, and these are: Gesture Recognition, Gesture Learning, and Gesture Synthesis. Gesture Recognition module takes a video stream as an input, and outputs known gestures as it recognises them. Gesture Learning module is responsible for adding new gestures, which cannot be matched to any of the known gestures, to the list of known gestures. Gesture Synthesis module generates motor commands in order to reproduce one of the known gestures.

Emotion and face recognition: This component consists of four operational modules that extract information from the input video image. Face Detection module extracts facial features from the input video frame and generates a rectangle that marks the location of the face. Face Recognition module uses low-resolution facial features to determine a person's identity. Emotion Recognition module uses high-resolution facial features such as eye corner, eye centre, eye brow, mouth corner, and nostrils shape, in order to recognise person's emotional state. Eye-Gaze Detection module detects 3-dimensional orientation of the face and the position of the eyes in order to determine the eye-gaze direction. Recognised faces, their associated facial features, and emotional patterns are stored for future identification.

Failure detection and recovery: The functionality of the component focuses on fault detection and identification in the robotic system. At a fundamental level, this means detecting hardware faults. Specifically this involves localising the fault and subsequently reassessing the capabilities of the robot with the fault in account. At a deeper level the study will look into detecting faults in the interaction which might have been caused by the human, and errors caused by a sudden change in the operating environment of the robot.

Safe manipulator arm: The manipulator consists of an arm and a gripper. Different manipulators will be acquired for two scenarios. For the Robot Waiter scenario, the goal is to build intrinsically safe arm with bio-inspired gripper with variable compliance for the operational and non-operational mode. For the USAR scenario, the arm and the gripper will be acquired from the commercially available hardware and the emphasis will be given to manipulation of the objects in the time-critical tasks.

Figure 3. Costumised robuLAB10 platform

Figure 4. Husky A200 Mobile Platform

Figure 5. Invenscience ARM 2.0

Robot interface: Interface component will be developed to provide relevant information to the users about the service and robot's operational state. For the Robot Waiter scenario, different user access levels will be defined for the supervisor and the end-users. Supervisor interface will be used to monitor the state of the robot and administrate the robot's knowledge base. End-user interface will allow an end-user to directly interact with the robot. Both modules will receive input from touch screen and speech recognition, but also from high-level sensors such as emotional recognition, face recognition, and intention and behaviour recognition. In the USAR scenario, external interface will in addition allow remote navigation of the robot when deemed necessary.

B. Robot Waiter system integration

The robotic platform that will be used for the integration of work into the Robot Waiter scenario is robuLAB10 developed by Robosoft (see fig. 2). It is a mobile platform with two propulsive wheels. The platform uses PURE low-level controller and robuBOX software kit (based on Microsoft Robotics Developer Studio) for the development of robot's services, both developed by Robosoft. The implementation of the software components will be carried out in C#.

The robuLAB10 platform is designed to ease the development of mobile robots [19] and it has been customised to fulfil the INTRO project's objectives related to cognition (learning, interaction, etc.), behaviour and gestures (see fig. 3). It uses a 270-degree laser range finder for mapping and navigation. Infrared and ultrasound sensors can be added for object detection. The customised platform includes a mounted tray that can be drilled to add the arm with its controller (or additional set of batteries, sensors, etc.). Additional connectors allow the arm integration. The communication for the arm control can be performed with PURE through the CAN bus. A 60mm gap between the tray and the mobile platform makes additional space for wiring (electric, RJ45), among other things. The provided power supply is 24V, 10A.

A rigid structure including three tubes and a desk permits placing a tablet PC for high-level robot applications, but it is also used to support a microphone and speakers. On the top of this structure a separate pan-tilt movement unit was added with a fire-wire camera that has zoom and auto-focus functionalities.

C. USAR robot system integration

The development and integration of the software components shall be carried out through the ROS (Robot Operating System) robotic framework, in C++. With regard to software deployment, there are two physical processors involved: the base station control PC and the onboard processing PC on the robot. All of the components shall be deployed on the onboard processing PC on the robot. The exception to this is teleoperation module. This will be deployed on the base station PC, to facilitate wireless remote control of the robot.

The hardware robotic framework for the integration scenario consists of two components: a mobile robotic platform, and a robotic manipulator. A mobile robotic platform should possess navigation capabilities for rugged and partially unstructured USAR environments. The requirements include navigation over different types of building rubble terrain, navigation over varying surface gradients up to 30 degree inclination and unstructured obstacles up to 10 cm in height, and a 20 kg payload capacity under reasonable working conditions.

A robotic manipulator that will be deployed on the mobile platform should be capable of light pick and place manipulation tasks. The requirements include payload capacity of at least 1 kg at full extension, reach capacity of at least 1 m, rudimentary gripper capable of grasping small simple shapes, and dynamic feedback on the joint angle positions.

As the mobile platform, the Husky A200 by Clearpath Robotics was chosen (see fig. 4). This robot satisfies the necessary scenario requirements. Clearance of 13 cm is sufficient to navigate over rubble terrain and 10 cm high obstacles. The platform provides sufficient payload capacity and drive power to navigate over variable terrain gradients.

As the manipulator, the Invenscience ARM 2.0 was chosen (see fig. 5). It has excellent reach and payload capacity. A gripper is included which is sufficient for the desired manipulation tasks. Potentiometers at joints are used to obtain sensory feedback on joint angles. The mobile platform has sufficient payload capacity to support the arm.

With regard to the exteroceptive sensing hardware, a Hokuyo URG-04LX laser range finder will be used for navigation, together with the embedded wheel encoders on the mobile robot. A Point Grey Bumblebee XB3 stereo vision camera will be used as a vision sensor for both navigation and manipulation tasks.

V. IMPACT AND FUTURE WORK

Introduction of robots to our everyday lives is a slow but inevitable process, and development of HRI technologies is of utmost importance in terms of safety, efficiency, and overall acceptance. Advantages of the INTRO research network lie in close collaboration of researchers from both academia and industry. This facilitates transfer of research results and their integration into new industrial technologies. The goal is to develop robotic systems that will be accessible to a larger number of end-users, and in a wider range of applications. Scalability will be achieved through a modular system architecture that permits selective use of hardware and software components, but also development of new system functionalities.

Future work in the INTRO network will cover several complementary domains. The contributions will be made to robot learning, human-robot collaboration, safety, and intelligent interface design. These however include several modalities, specifically addressing face detection and recognition, emotion recognition, gesture identification and synthesis, joint attention, behaviour recognition, manipulator and griper design and control, failure detection and recovery, among others. The developments will be implemented on an operational system and demonstrated in a specific application scenario.

REFERENCES

[1] D. Feil-Seifer and M. J. Matarić, “Human-robot interaction,” in Encyclopedia of Complexity and System Science, Robert A. Meyers, ed., Springer Tracks in Robotics 54, Springer, 2009.

[2] J. Burke, R. R. Murphy, M. Coovert and D. Riddle, “Moonlight in Miami: A field study of human-robot interaction in the context of an urban search and rescue disaster response training exercise,” Human-Computer Interaction, pp. 85-116, 2004.

[3] M. Baker, R. Casey, B. Keyes and H. A. Yanco, “Improved interfaces for human-robot interaction in urban search and rescue,” 2004 IEEE International Conference on Systems, Man and Cybernetics, pp. 2960-2965, 2004.

[4] P. Hinds, T. L. Roberts and H. Jones, “Whose job is it anyway? A study of human-robot interaction in a collaborative task,” Human-Computer Interaction, pp. 151-181, 2004.

[5] W. Kennedy, et al., “Spatial representation and reasoning for human-robot collaboration,” Proceedings of the national conference on artificial intelligence, pp. 1554-1562, 2007.

[6] M. Cakmak, S. S. Srinivasa, M. K. Lee, S. Kiesler and J. Forlizzi, “Using spatial and temporal contrast for fluent robot-human hand-overs,” Proceedings of the 6th international conference on Human-robot interaction, pp. 489–496, 2011.

[7] S. Glasauer, M. Huber, P. Basili, A. Knoll and T. Brandt, “Interacting in time and space: Investigating human-human and human-robot joint action,” 2010 IEEE RO-MAN, pp. 252–257, 2010.

[8] G. Hoffman and C. Breazeal, “Effects of anticipatory perceptual simulation on practiced human-robot tasks,” Autonomous Robots, pp. 403-423, 2009.

[9] F. Kaplan and V. V. Hafner, “The challenges of joint attention,” Interaction Studies, vol. 7, no. 2, pp. 135-169, Jan. 2006.

[10] B. Argall, S. Chernova, M. Veloso and B. Browning, “A survey of robot learning from demonstration,” Robotics and Autonomous Systems, pp. 1-15, 2009.

[11] B. Fonooni, T. Hellström and L. E. Janlert, “Learning High-Level Behaviors From Demonstration Through Semantic Networks,” in Proc. of the 4th International Conference on Agents and Artificial Intelligence (ICAART 2012), 2012 – to appear.

[12] G. Schillaci and V. V. Hafner, “Random Movement Strategies in Self-Exploration for a Humanoid Robot,” in Proc. of the 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI2011), pp. 245-246, March 2011.

[13] S. Bodiroza, G. Schillaci and V. V. Hafner, “Robot Ego-sphere: An Approach for Saliency Detection and Attention Manipulation in Humanoid Robots for Intuitive Interaction,” in Proc. of the 11th IEEE-RAS International Conference on Humanoid Robots, pp. 689-694, 2011.

[14] G. Doisy and J. Meyer, “Expectations regarding the interaction with a learning robotic system,” in Proc. of the ACM/IEEE International Conference on Human-Robot Interaction (HRI 2011), pp. 41-44, 2011

[15] C. Harper, M. Giannaccini, R. Woodman, S. Dogramadzi, T. Pipe and A. Winfield, “Challenges for the hazard identification process of autonomous mobile robots,” in 4th Workshop on Human-Friendly Robotics (HFR 2011), Enschede, Netherlands, 2011.

[16] C. Granata, M. Chetouani, A. Tapus, P. Bidaud and V. Dupourque, “Voice and Graphical based Interfaces for Interaction with a Robot Dedicated to Elderly and People with Cognitive Disorders,” in 19th IEEE International Symposium in Robot and Human Interactive Communication (Ro-Man 2010), 2010.

[17] D. Salle, M. Traonmilin, J. Canou, and V. Dupourque, “Using Microsoft robotics studio for the design of generic robotics controllers: the robuBOX software,” in ICRA 2007 Workshop Software Development and Integration in Robotics - ”Understanding Robot Software Architectures”, Roma, Italy, April 2007.

[18] M. Quigley et al., “ROS: an open-source Robot Operating System,” ICRA Workshop on Open Source Software, 2009.

[19] C. Granata and Ph. Bidaud, “Interactive person following for social robots,” in Proc. of the 11th International Conference on Climbing and Walking Robots and the Support Technologies for Mobile Machines (CLAWAR 2011), Paris, France, Sep. 2011.