Augmented reality-aided tele-presence system for robot manipulation in industrial manufacturing

Lorenzo Peppoloni*and Filippo Brizzi and Emanuele Ruffaldi and Carlo Alberto Avizzano

PERCRO, Tecip Institute, Scuola Superiore SantAnna, 56010 Pisa, Italy

Abstract

This work investigates the use of a highly immersive telepresence system for industrial robotics. A Robot Operating System integrated framework is presented where a remote robot is controlled through operator's movements and muscle contractions captured with a wearable device. An augmented 3D visual feedback is sent to the user providing the remote environment scenario from the robot's point of view and additional information pertaining to the task execution. The system proposed, using robot mounted RGB-D camera, identifies known objects and relates their pose to robot arm pose and to targets relevant to the task execution. The system is preliminary validated during a pick-and-place task using a Baxter robot. The experiment shows the practicability and the effectiveness of the proposed approach.

CR Categories: I.2.9 [Robotics]: Operator Interfaces— [H.5.1]: Multimedia Information Systems—Artificial, Augmented, and virtual realities; I.4.8 [Scene Analysis]: Motion—Sensor Fusion

Keywords: Tele-operation and Tele-presence, Augmented and Mixed Reality, Tracking and Sensing.

1 Introduction

Thanks to the recent advancements in the robotics research fields, robots are now allowed to perform a plethora of demanding manipulation tasks. Despite the significant progress there are scenarios where human intelligence is still necessary. The problem is usually addressed with approaches falling in two main categories. The operator gives remote guidance to a robotic manipulator through a haptic devices or through wearable interfaces, which can have a haptic feedback. The main challenge for the first approach, which usually employs also a visual feedback on a monitor, lies in the difficulty to coordinate the movements of a teleoperated device with the ones of the human upper limb. This may impair the user's ability to understand the relative position of the remote robot arm with respect to other objects and to anticipate future positions of the manipulator. On the other hands wearable haptic devices, such as exoskeletons, provide a more natural interface for the operator. Despite the high level of embodiment, exoskeletons fail to mimic the entire range of human arm movements, have usually limited adjustability and cannot be employed for extended period because of their weight. Moreover, operator's safety has to be guaranteed usually limiting the responsiveness of the system and the comfort of the interaction. Solution presenting less obtrusive wearable technologies have been already presented using in particular singularly IMUs and EMG signals. Usually in all of the proposed works the visual feedback is obtained allowing the robot and the operator to share the same workspace. This work proposes a system with a high level of embodiment for telepresence in industrial task using IMUs and EMG to generate control signals for remote robot teleoperation. The operator's movements and muscle effort are captured through a compact wearable interface. Such a solution does not suffer from the problem of lack of anthropometric adaptation of exoskeletons, since a calibration procedure allows the system to be adjusted on different subjects. The system is capable of recognizing objects relevant to the task execution in the remote environment and to register object target poses. These information are used to provide a 3D augmented feedback showing the robot perspective of the manipulated environment and simultaneously representing objects, robot and task information to improve the manipulation performance. The feedback is sent back to the user through an head mounted display (HMD). The system is completely integrated in the Robot Operating System (ROS) and can be used with every ROS-compatible robotic manipulator. A preliminary assessment of the system is given with a pick-and-place task performed with a Baxter Robot. Compared to existing works with a similar approach we further progress the control architecture fusing information from both IMUs and EMG sensors and the embodiment component adding an immersive visual feedback. Moreover our approach is platform independent, while usually similar systems are designed to work with a particular robotic platform. The paper is structured as follows. Section 2 gives an overview of the related works. Section 3 presents the architecture of the framework. Section 4 describes in details all the different components and algorithms of the system ROS architecture. In section 5 the experimental setup used as a benchmark is presented along with the results of the tests. The results are then discussed in section 6, and final remarks about the system and further developments are presented.

2 Related Work

In the context of telepresence in industrial robotics several approaches using haptic interfaces have been presented. In [Radi et al. 2010; Kron and Schmidt 2003] the authors present systems with multimodal feedback, both visual and haptic. Systems with a higher level of embodiment can be found in [Folgheraiter et al. 2012]. These works present systems with multimodal feedback where the operator controls a remote robot with a wearable exoskeleton. An immersive feedback of the remote scene is given to the user. In the field of less obtrusive wearable interfaces to control robots, works can be found using EMG and IMUs. In [Vogel et al. 2011] the authors present a machine learning technique to map the EMG signals to the Cartesian coordinates. The system is then used to control a robotic manipulator. An approach based on IMUs is presented in [Khassanov et al. 2014]. The authors propose a framework for tra-

*e-mail:l.peppoloni@sssup.it

DOI: http://dx.doi.org/10.1145/2821592.2821620

jectories generation during pick and place tasks.

3 System Overview

The system described in this paper consists of a ROS-integrated framework for human-in-the-loop control of a robot with real-time augmented immersive visual feedback. Figure 1 shows the architecture of the system. The operator side and the robot side are not co-located and are linked through the ROS Control Unit. The user is sensed with a wearable device sending synchronized EMG and IMUs signals through Bluetooth to the control unit. The ROS-integrated control unit reconstructs, in real-time, the pose of the user's arm and extracts information about muscle activations through EMG features computation. These information are then used to generate a control signal for the remote robot. The robot-side environment is perceived through a RGB-D Kinect camera. The camera stream is augmented by the ROS Control Unit with additional task information and it is sent back to the user in the form of an immersive visual feedback through an HMD.

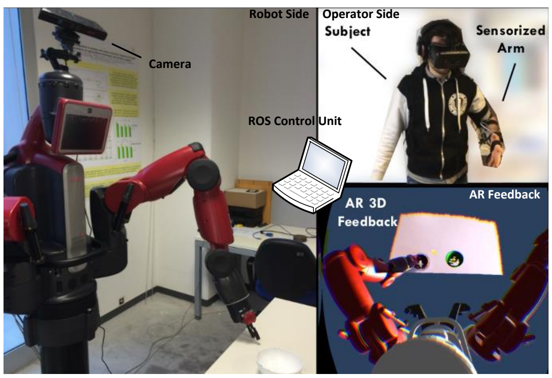

Figure 1: *The system architecture. Synchronized raw sensors data are captured and sent via Bluetooth to the host PC by the wearable device. Gathered data are used to extract motion and muscle efforts information. The computed data are used by a ROS control node to operate a robot arm. An augmented visual feedback of the scenario with which the robot is interacting is provided to the user.*

3.1 Operator Side

On the operator side information about the user's motion and muscle effort are gathered by the wearable device presented in [Avizzano et al. 2014]. Three IMUs and a surface EMG sensor with 8 channels are attached to the user's arm. IMUs signals are sampled at 100Hz, EMG signals are acquired at 4kHz, filtered and then downsampled to 300Hz, which is enough to capture the major part of the human muscle effort. The signals synchronization is guaranteed by the wearable device. Sensors raw data are sent through Bluetooth to the ROS Control Unit. The operator side comprises also the HMD for the 3D AR feedback. The HMD is an Oculus Rift DK2 (1080p at 60Hz with global positioning) and it is connected to the ROS Control Unit via USB connection.

3.2 ROS Control Unit

The ROS Control Unit acts as the central unit of the system. Sensors data are used to reconstruct operator's upper limb motion and compute EMG features. The reconstructed motion and EMG signals are used by to generate the control signal for the robot on its side. The perception stream from the robot side is used for object recognition and to generate the AR feedback for the operator side.

3.3 Robot Side

The robot side comprises the teleoperated robot and the environment in which it is acting. A Kinect camera captures the remote scenario from the robot's view point. The perceptual stream is sent to the ROS Control Unit. The system is not dependent on a particular robot, since it is implemented to allow usability with every ROS-compatible robotic platform.

4 The ROS Control Unit

The structure of the ROS Control Unit is shown in Figure 2. Several components interact at this level. The Motion Reconstruction and EMG Features modules process the raw data from the wearable device. The ROS Control Node generates the control signal for the remote robot from the reconstructed motion and muscle effort and computes forward kinematics data to be used for the augmented feedback. The Object Recognition Module recognizes objects in the remote scene using the Kinect camera stream. The AR Feedback Module generates the augmented feedback sent to the operator.

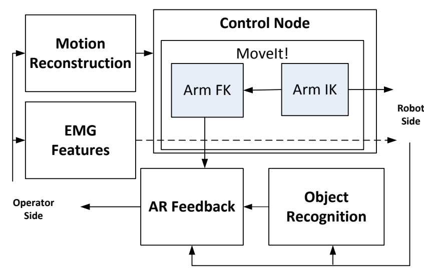

Figure 2: *Raw sensors data are used to reconstruct operator's upper limb motion and compute EMG features. The reconstructed motion and EMG signals are used to generate the control signal for the robot side. The perception stream from the robot side is used for object recognition and to generate the AR feedback sent to the operator side together with the robot arm forward kinematics (FK) computed by ROS MoveIt!. MoveIt! computes also the robot arm inverse kinematics (IK), from which the desired joint angles are extracted. The information about the muscle activation are directly sent to the driver as a control signal for the gripper.*

4.1 Motion Reconstruction and Muscle Effort

IMUs signals are fused exploiting an Unscented Kalman Filter (UKF) to reconstruct online the posture of the user's arm. Motion information are published in ROS at a frequency of 100Hz. For further information about the algorithms the reader can refer to [Peppoloni et al. 2013; Ruffaldi et al. 2014]. Surface EMG array sensors monitor the activity of the operator's forearm flexors (flexor carpi radialis, the palmaris longus and the flexor carpi ulnaris). The raw signal is windowed, low-pass filtered to remove motion artefacts, then it is rectified and high-pass filtered, computing the linear envelope. The value of the linear envelope is published in ROS at a rate of 5Hz.

4.2 ROS Control Node

The ROS node for the robot control acts as an interface from the information published by the pose reconstruction algorithm and muscle activation node to the remote manipulator driver ROS wrapper. The architecture of the control node is shown in Figure 2. The implemented control law matches the 3D pose of the operator's hand in the operator's shoulder reference frame with the robot end effector pose in the robot base frame. This approach allows the system to be as flexible as possible regarding the usability with different robotic platforms, since mapping the operator's upper limb joints to the robot ones, is not a feasible solution when using robotic arms with a kinematic structure not similar to the human's arm one. The control node provides also information about the desired robot pose used to augment the visual feedback. The pose is obtained through the Moveit![1] forward kinematics.

4.3 Objects Recognition and AR Feedback

Since we are dealing with assembly and in general industrial tasks, it is crucial to provide the operator the greatest possible level of information about the remote environment in which he is acting. In order to do so we provide the system with an object recognition module based on the ROS object_recognition_tabletop package [2]. A ROS node providing a ROS service to record and store target poses for objects has been added to the package. The stored poses can be sent to the AR feedback module to add meshes of target objects in the target poses to the virtual scene. All the information from the Object Recognition module are sent to the AR feedback module.

The AR feedback component has been developed using a newly created ROS-integrated framework for high-performance AR, called Compact Components (CoCo), composed of a core library and of several specialized modules. This framework is motivated by the high requirements of visual feedback. The AR feedback can be divided in two different components according to the type of conveyed information. The first module, using the information provided by the control node, superimposes the animated robot model to the scene, complementing the camera view of the real robot. This helps the user to predict the remote robot arm movements and its pose in the remote environment. This module is also is also responsible of moving the Virtual View Point from its initial pose according to relative translations along the axes and relative roll, pitch and yaw angles computed from the HMD tracking. The virtual scene is rendered as a 3D mesh obtained from the undistorted RGB image and the depth provided by the camera. The distortion parameters are directly obtained from the Kinect camera, through the *openni_camera* ROS stack. The second module provides virtual fixtures [Rosenberg 1993] which have been proven to be very helpful in teleoperation tasks [Xia et al. 2012]. Virtual representation of the objects are displayed according to the information received by the object recognition node: the color of the virtual object is changed according to a RGB color scale following the distance between the robot end-effector and the object position. Furthermore the system allows to display virtual objects in the target poses required by each task. The visualized scene is updated at the same frequency of the Kinect (30 Hz), while the added virtual objects are rendered at 60 Hz. The mean AR component runs on a PC with a quad-core Intel i7 CPU (2.3 GHz) and a NVIDIA GeForce GT 650. The latency of the system can be separated into two parts: user's motion to robot motion latency and the sensor to display command latency. To estimate the first component we measured the time from the availability of the desired position to the moment in which the robot's end effector reaches it. This time comprises the control command computation, the network delay, the actual robot's movement and the latency added by the ROS topic queue needed to compute the time required by the robot to complete the movement. To perform a 1 cm linear movement, the system takes 1.9 seconds. The estimated sensor to display command average latency is 89 ms, computed after synchronizing the robot and graphics computers with PTP. The average network latency is taken into account.

[1] http://moveit.ros.org/

[2] http://wg-perception.github.io/tabletop/

4.4 Reference Frames Calibration

The overall structure of the reference systems is shown in Figure 3. The frames are organized in three groups: Remote, Local and Virtual, named from the point of view of the human operator. Remote is the environment that contains the robotic manipulator, the Kinect camera and objects, Local is the location of the human operator with the wearable sensing interface, and Virtual is the environment used for generating the synthetic view displayed to the operator. In particular the Virtual scene comprises the 3D scene obtained from the Kinect camera as point cloud, the point of view of the user virtual head, and visual augmentation (robot model and recognized objects). Calibration procedures have been implemented to compute the fixed transformations represented with dashed lines which links the three frames trees.

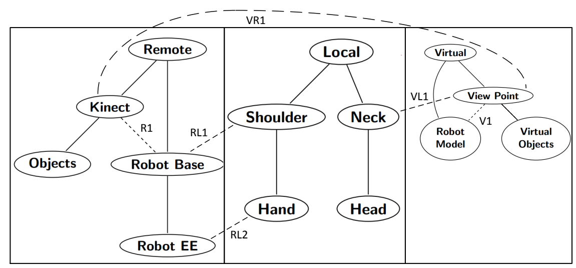

Figure 3: *The transformation graph describing the connection between the relevant reference frames. Local frames are related to the user side, while Remote frames comprises all the reference frames from the remote manipulator side. Virtual frames are the ones represented in the 3D virtual environment. Fixed transformations between the frames (dashed lines) are computed with the calibration procedures.*

5 Preliminary Test and Results

The system usability has been evaluated in a standard pick-and-place task. The task consisted of teleoperating the remote robot to pick a known object from a starting position and placing it in two different target poses T_1 and T_2. The remote robot used for the experiment was a Baxter robot. The robot acted in the remote environment, the operator controlled the robot through his movements. The AR feedback was sent to the operator using the HMD. Four subjects with no previous training with the system participated in the experiment. The complete experiment protocol is the following: 1) the subject signed a consent form which explained the experiment, 2) the wearable device and the HMD were mounted on the operator's body (approximately $10 - 15$ mins), 3) the subject was given 5 minutes to familiarize with the system firstly sharing the operative space with the robot and then with the full AR feedback, 4) the subject was asked to perform the task twice for every target pose, with no augmented information (only the meshified scene was sent to the HMD) and with the AR feedback (target pose, objects and animated robot model) The task performance is based on execution time and final placing error. Subjects 1 and 3 first performed the test without the augmented information and then the test with the full AR feedback, while subjects 2 e 4 first performed the test with the full AR feedback and then without augmented information. The object used for the test was a plastic bowl added to the object recognition database in order to be detected by the ROS object_recognition_tabletop package. The results of the experiment are reported in Figure 4.

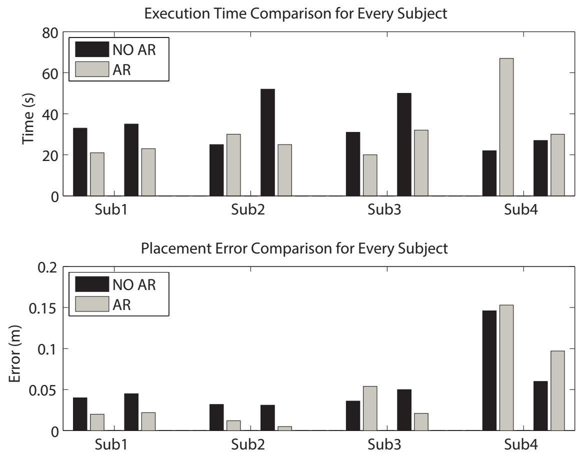

Figure 4: *On top the comparisons between execution times for both the targets for all the subjects is presented, execution time without AR (black) and with AR (grey) are shown. On bottom the comparisons between placement errors for both the targets for all the subjects is presented, execution time without AR (black) and with AR (grey) are shown.*

6 Discussions and Conclusion

The test showed that every subject was capable to complete all the tasks. Table 1 shows the comparisons of final placing errors and execution times between the executions with full AR feedback and no AR feedback. According to the results there is a slight improvement in positioning error and execution time using the full AR but further tests are needed to conclude about the effectiveness of the AR feedback. Anecdotally it has been reported by the subjects that the most useful information provided by the AR feedback is the color mapping of the distance to the target object mesh. Further investigation in this direction could provide a viable solution to the teleoperation problem of judging the relative position of the teleoperated arm to other objects [Nitsch and Farber 2013]. Despite the success rate of the experiment it has to be noted that certain positions assumed autonomously by the remote robot (due to the unconstrained inverse kinematics) can impair the execution of the task. The change of the virtual view point according to the head movement has been helpful in tackling this problem. The preliminary results showed that system could be a valid alternative to the classical haptic teleoperation. The system allows in facts the operator to perform the task with a higher level of embodiment compared to the use of a haptic device. the presented ROS-integrated framework for telepresence in assembly tasks. User motion and muscle strain are captured through a wearable device and reconstructed by a ROS node. The extracted information are then used to control the movements of a remote robot arm during manipulation tasks. A 3D visual feedback of the scenario in which the manipulator is acting is given to the user through a Kinect camera and a HMD. The visual feedback is augmented with information about the pose of the robot arm, objects in the remote environment and target poses. Issues to be further investigated remain about the effectiveness of the use of the augmented reality information and about the degree of autonomy of the teleoperated robot. Further development will regard the addition of a vibrotactile haptic feedback and the possibility to use the system for programming task by demonstration.

Table 1: *Comparisons of mean and standard deviation of the errors and times between AR and no AR executions*

Trail Type	Time [s] μ	Time [s] σ	Error [m] μ	Error [m] σ	Target
AR	34.50	22.13	0.059	0.065	T_1
no AR	27.75	5.12	0.063	0.055	T_1
AR	27.50	4.20	0.036	0.041	T_2
no AR	41	12.03	0.046	0.012	T_2

Acknowledgements

The material is based on work carried out in the TAUM Project co-financed by the European Structural Fund for Regional Development, programme POR CReO FESR 2007-2013 of Regione Toscana.

References

AVIZZANO, C. A., ET AL. 2014. A novel wearable biometric capture system. In *IEEE MED*.

FILIPPESCHI, A., ET AL. 2015. Encountered-type haptic interface for virtual interaction with real objects based on implicit surface haptic rendering for remote palpation. In *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*, IEEE.

FOLGHERAITER, M., ET AL. 2012. Measuring the improvement of the interaction comfort of a wearable exoskeleton. *International Journal of Social Robotics 4*, 3, 285–302.

GEIGER, L., ET AL. 2010. The influence of telemanipulation-systems on fine motor performance. In *IEEE AHCI*, 44–49.

KHASSANOV, Y., , ET AL. 2014. Inertial motion capture based reference trajectory generation for a mobile manipulator. In *IEEE/ACM HRI*, 202–203.

KRON, A., AND SCHMIDT, G. 2003. A bimanual haptic telepresence system-design issues and experimental results. In *Proc. of Int. Workshop on High-Fidelity Telepresence and Teleaction in HUMANOIDS*.

NITSCH, V., AND FARBER, B. 2013. A meta-analysis of the effects of haptic interfaces on task performance with teleoperation systems. *Haptics, IEEE Transactions on 6*, 4, 387–398.

NITSCH, V. 2012. Haptic human-machine interaction in teleoperation systems and its implications for the design and effective use of haptic interfaces.

PEPPOLONI, L., ET AL. 2013. A novel 7 degrees of freedom model for upper limb kinematic reconstruction based on wearable sensors. In *IEEE SISY*, 105–110.

RADI, M., ET AL. 2010. Telepresence technology for production: from manual to automated assembly. In *Haptics: Generating and Perceiving Tangible Sensations*. Springer, 256–261.

ROSENBERG, L. B. 1993. Virtual fixtures: Perceptual tools for telerobotic manipulation. In *IEEE VR*, 76–82.

RUFFALDI, E., ET AL. 2014. A novel approach to motion tracking with wearable sensors based on probabilistic graphical models. In *IEEE ICRA*.

VOGEL, J., ET AL. 2011. EMG-based teleoperation and manipulation with the DLR LWR-III. In *IEEE IROS*, 672–678.

XIA, T., ET AL. 2012. Augmented reality environment with virtual fixtures for robotic telemanipulation in space. In *IEEE IROS*, 5059–5064.