Third Point of View Augmented Reality for Robot Intentions Visualization

Emanuele Ruffaldi(✉), Filippo Brizzi, Franco Tecchia, and Sandro Bacinelli

TeCiP, Scuola Superiore Sant'Anna, Pisa, Italy
e.ruffaldi@sssup.it
http://www.percro.org

Abstract. Lightweight, head-up displays integrated in industrial helmets allow to provide contextual information for industrial scenarios such as in maintenance. Moving from single display and single camera solutions to stereo perception and display opens new interaction possibilities. In particular this paper addresses the case of information sharing by a Baxter robot displayed to the user overlooking at the real scene. System design and interaction ideas are being presented.

A new generation of robotic systems is being introduced in working environments from small to large factories. These robots, thanks to advancements in actuation and perception, are capable to cooperate with human workers in the execution of task, rather than performing their own task independently inside a highly structured workflow. Examples of such robotic system are the Baxter [1] from Rethink Robotics and ABB Yumi, anticipated in the research world by many projects [2].

With the increased capability of these robots and the expected cooperative interaction, there is a need for the operator to understand the robot intention and current state as much as the robot needs to understand operator intentions. The former for supervision, the latter for safety and proactivity. The nature of the Human-Robot Communication (HRC) between these robots and human workers needs to take into account the specificities of working environment that limits traditional communication channels [3]: possibly over the average sound levels, direct manipulation of touch devices limited by gloves or by the working activity.

A specific need for the operation is the possibility of understanding the intention of the robot contextualized over the working environment, that is to understand if the chosen object to be manipulated is the correct one or the target location. There are several display options for providing this information spanning from the traditional ones, such as display panels placed in the environment, on the robot, wear by the operator or simply in the hand, to projective or presented in eye- or head- mounted displays. In any of this case we are interested in presenting the selection highlighted in the real world by means of the capabilities offered by Augmented Reality (AR).

Industrial plants commonly require specialist maintenance expertise; as a consequence, plants located in remote sites and away from where the compo-

L.T. De Paolis and A. Mongelli (Eds.): AVR 2016, Part I, LNCS 9768, pp. 471–478, 2016.
DOI: 10.1007/978-3-319-40621-3_35

nents were produced can be difficult to service effectively. Addressing major equipment failures often requires specialist on-site intervention, which can result in significant down-time and cost, but, more importantly, some maintenance and corrective procedures are so complicated or site-specific that a local engineer often is not able to proceed without complex instructions. The potential of Augmented Reality and Robotic Assistance in these frequent situations is therefore potentially disruptive, as both can greatly decrease the perceived complexity of the tasks.

The paper presents and discusses a stereo Augmented Reality eye-wear integrated in a working helmet for HRC with a humanoid robot for collaborative applications, a Baxter, discussing the supporting components, system architecture and calibration issues. The long term research question is on which information is better to be displayed, how it can be overlaid on the real scene, and which are the usability challenges. The specific research question of this short paper is on the challenges in calibrating the different point of view, and supporting the augmentation on a low power system such the one proposed.

Section 2 presents the State of the Art, Sect. 3 discusses the helmet and architecture, followed by the augmentation in Sect. 4. Then, follows Discussion in Sect. 5.

1 Background

There is recent strong technological trend in eye-level AR that is moving from research prototypes to high-quality displays for wearable AR such as HoloLens [4], Meta, Canon MREAL, Epson Moverio, Dahiri, followed by more unknown technology such as Magic Leap. All these systems are characterized by some form of pass-through display technology that allows to overlay the information over the field of view of the operator coupled with a wearable computer vision subsystem that acquires information about the external environment and allows to provide gesture recognition, object recognition and more over information for the correct overlay of augmented models over the real world.

The AR helmet presented in this work is not providing a see-through experience, but it is instead based on a display that is located in the upper part of the field-of-view of the operator. The choice of this solution is coming from the experience in tele-maintenance with helmets in which the operator is interested in an augmented view only in certain phases of the work while keeping the clear site of the environment. The motivation is also in the compactness of the solution and integration in a regular safety helmet.

The interaction between human and robots in cooperative environments has been investigated in several projects looking at specific aspects of safety, physical interaction, coping also with cognitive capabilities for the collaboration. The communication part has been typical based on the audio channel or regular display. In the area of Augmented Reality many works have been devoted to the overlay of information coming from a database or a running system such as examples in industrial or automotive maintenance [5]. An example of alternative communication from robots is based on projective augmented reality [6].

An important aspect is the physical co-location of the human operator with respect to the robot. When such co-location is not necessary it is possible to employ more immersive display techniques such as a Head Mounted Displays (HMD) as the authors employed in a Mixed Reality (MR) setup with a Baxter robot [7].

The present work proposes the setup to explore the HRC in collaborative robotics by means of a industrial helmet augmented with a non-see through display that is connected with the robot system, being capable, in this way, to show the results from the perceptual and cognitive subsystems of the robot.

2 Helmet for Industrial AR

In this section we are discussing the design and realization of a video see-trough Augmented Reality helmet as shown in Fig. 1. This device has been explicitly designed to fit over a standard safety helmet without the need of modifications or an adapter. This is a key aspect to comply to safety regulations in many industrial environments. There are two main snap-on parts, kept together by elastic bands: the main part contains a wireless, battery-powered module equipped with a stereo display, a stereo camera subsystem with interchangeable lenses and on-board computing. The display module is constituted by two compact LCD displays with LED retro-illumination. The field of view (FOV) for each eye is 32 degrees, with a 100 % overlap. The display block is structured along a plastic pipe that contains a mechanism to allows the adjustment of the intra-ocular distance. An unique characteristic of the displays is their placement: the design of the unit was carefully chosen to occlude only a small portion of the worker field of view. In this way it is still easy to navigate into the environment without the encumbrance usually associated to video see-trough solutions. The computer vision part is characterized by two 5 M pixel cameras that support Full HD video streaming at 30 Hz. Anyway in this setup they have been scaled down to VGA resolution (640×480) at 30 Hz in order to limit the potential effects of WiFi interferences. The on-board processing is provided by two ARM modules based on the BCM2835 chipset, running at 700 MHz and equipped with 512 MB of RAM, that have been conceived exactly for compact high efficiency processing. This architecture allow to access the cameras using a low-latency Mobile Industry Processor Interface (MIPI). to control the output display via HDMI and to perform compression and decompression of the image streams in real-time. The helmet has been designed for a general purpose of tele-presence and its main software capability is to stream and receive compressed video images, encoded in low-latency h.264, to a target computer or to another helmet over WiFi with a end-to-end latency comparable (or inferior) to what is usually allowed by standard USB cameras. The device is also equipped with a 6-DOF motion sensor based on the Invensense MPU-9150, although it was not used in the setup discussed in this work.

An unique characteristic of the displays is their placement: the design of the unit was carefully chosen to occlude only a small portion of the worker

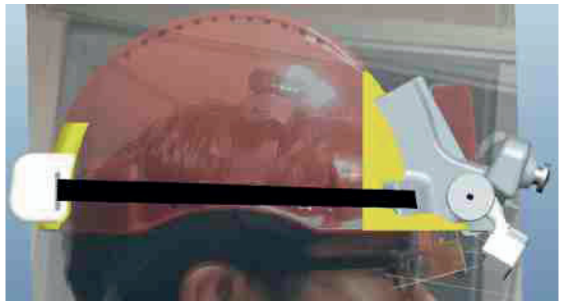

Fig. 1. Helmet concept

field of view, as discussed in [8]. In this way it is still easy to navigate into the environment without the encumbrance usually associated to video see-trough solutions

3 System Setup

The HRC case discussed in this paper has been realized with a Baxter robot that is provided with two arms. The Baxter has natively one VGA camera at each arm end-effector and it has been augmented with an Asus xTion Live Pro RGB-D camera that has characteristics similar to the original Kinect 360. The robot has an internal computer running the Robotic Operating System (ROS), while a separate computer is bridging the robot with the Helmet performing all the necessary augmentation computations. The bridging is performed by using the CoCo framework for Mixed Reality [9,10] that has been extended for supporting the Helmet: the helmet appears as a stereo camera source and as a stereo display output. The low-level driver of the Helmet exposes two shared memory buffers that are polled/updated dealing with compression and decompression of the image stream. System architecture is shown in Fig. 2.

The resulting interaction in the experimental environment is shown in Fig. 3.

4 Augmentation

The objective of the augmentation discussed in this work is to provide information about the status of the robot represented in the field of view of the operator. In particular the robot localizes an object, and if it is also in the field of view of the operator it is highlighted in the head-up display. The head-up shows the images coming from the helmet cameras augmented with the object highlight. The augmentation is performed by transferring the pose of the object from the robot reference systems to the operator ones without the need of performing object recognition in the helmet. The outcome is shown in Fig. 5.

The relevant reference frames and their connections are discussed here for clarifying the proposed approach: a frame is named by an upper case letter,

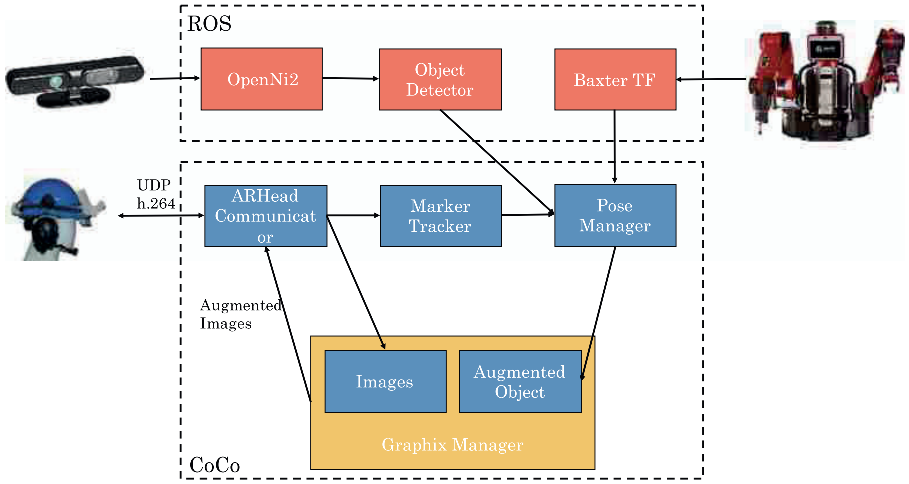

Fig. 2. System Architecture: on the top part the Baxter robot is shown with the ROS nodes supporting object recognition. The lower part shows the CoCo based processing that acquires the camera images, augment them using the information from the robot and send them back to the ARHead.

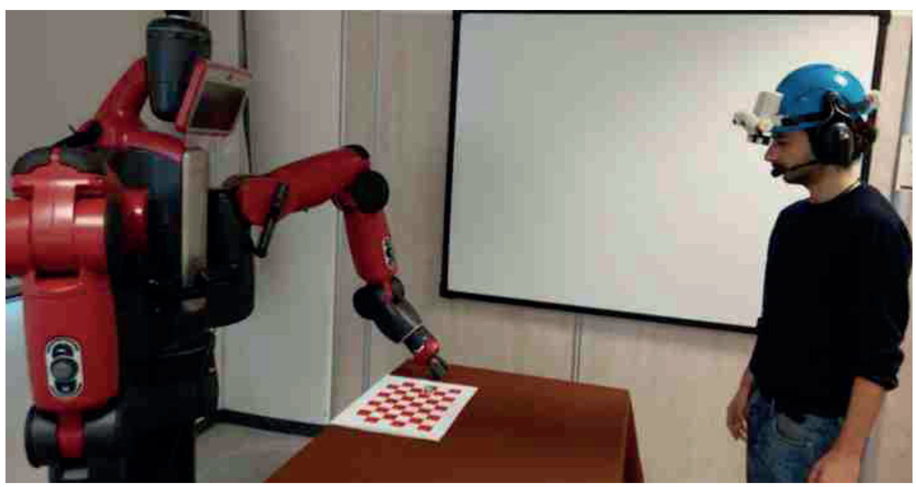

Fig. 3. Experimental environment with Baxter and operator

while the associated transformation is expressed by an arrow, that is simple as in $A \rightarrow B$ for dynamic transformations and double as in $A \Rightarrow B$ for static ones. Let's call (R) the robot root at the hip, (C_i) is the i-th camera of the robot, (O) is the object, (H) is the operator helmet, (Q_i) is the i-th camera of the helmet. For the robot cameras can be fixed $C_i \Rightarrow R$, as for the RGB-D at the torso, or $C_i \rightarrow R$ kinematic-based as for the cameras at end-effectors. The object is localized using the LINEMOD algorithm [11] providing the $O \rightarrow R$ transformation. Sensor fusion has been employed for obtaining a better estimate of the object $O \rightarrow R$ and $H \rightarrow R$ poses. In particular the robot object recognition takes into account uncertainty of the kinematics, distance of objects and motion as reported in a pending publication. For the operator all the transformations in the helmet are fixed and obtained by a registration procedure, namely obtaining $Q_i \Rightarrow H$.

What is missing is the transformation of the operator's head with respect to the robot base $H \rightarrow R$. This information can be obtained by some form of localization, and in this case it is based on an Aruco fiducial marker [12] that is seen both by the robot and the helmet's cameras. Calling F the marker frame then having $F \Rightarrow R$ and $F \rightarrow H$ allows to relate all the transformations. In the specific case discussed in this paper the robot is not moving, while the person is moving with respect to the reference. The overall structure of the reference systems is shown in Fig. 4.

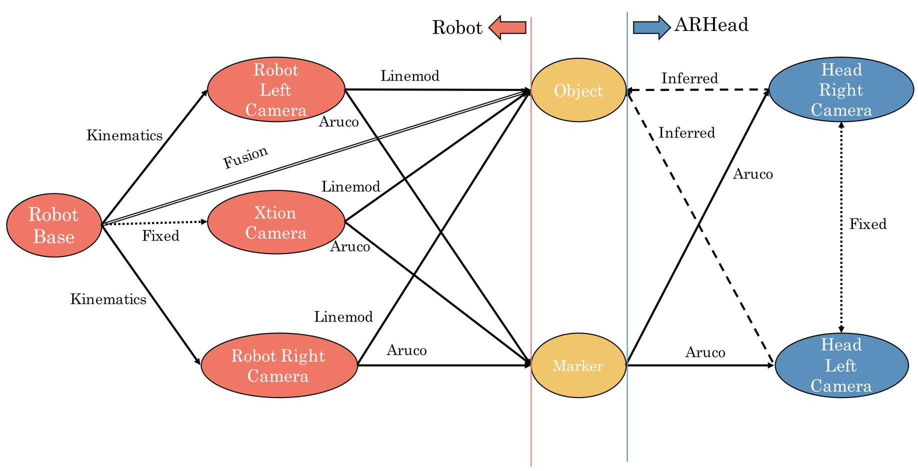

Fig. 4. Reference systems and connections. On the left the ones of the robot, in the middle the working environment, on the right the ARHead. Object position is obtained by sensor fusion and shown as double line.

The helmet pose is obtained by fusing the two estimates $F \rightarrow Q_1$ and $F \rightarrow Q_2$ in SE3 and weighted by the reprojection errors.

At present the resulting augmentation is limited to representing a reference system over the target object as shown in Fig. 5. Several types of augmentations can be investigated as briefly listed in the discussion section.

5 Discussion

Augmentation elements have been only sketched in this work, while most of focus has been on reference systems and architecture. Possible types of feedbacks that

Fig. 5. System Augmentation from two different point of views of the left eye. The robot frames have been highlighted and the object identified has been marked with a red cylinder. The direction of motion of the robot arm is shown with an arrow (Color figure online)

are being investigated are: (1) robot trajectory, (2) target position for moved object, (3) better highlight of the object, (4) robot workspace, (5) information about the object. Robot trajectory is useful for programming by demonstration task or in any case in which the robot is going to execute a motion path and the operator would like to see it in advance [13]. Object highlighting can be used to identify which is the object that is going to be picked by the robot, or that the robot suggests to the operator [7,14].

6 Conclusions

The paper has presented the system setup, the architecture and the reference frame issues that emerge from the possibility of creating a third point of view augmented reality feedback based on robot state. Basic augmentation has been presented as the result of the system design.

The main challenge in present setup is the quality of the tracking to obtain the common integrated reference system with the robot. The VGA resolution of the ARhead cameras could be raised up to Full HD at the cost of higher bandwidth requirements. Anyway, due to the fact that the head tracking information is based on the marker, while most of the computer vision part is done on the robot the tracking could be processed locally.

The next stage of the work is the investigation of effectiveness of the feedbacks, and the understanding on feedbacks that can be adapted depending on the level of uncertainty of the tracking.

Acknolwedgments. This work has been carried out within the framework of the Tuscany Regional Project TAUM. We also acknowledge Erika Di Stefano for having provided the Baxter setup and the computer vision algorithm, and Alessandro di Fava for the robot control.

References

1. Guizzo, E., Ackerman, E.: How rethink robotics built its new baxter robot worker. IEEE Spectrum (2012)
2. Heyer, C.: Human-robot interaction and future industrial robotics applications. In: 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 4749–4754. IEEE (2010)
3. Mavridis, N.: A review of verbal and non-verbal human-robot interactive communication. Robot. Auton. Syst. **63**, 22–35 (2015)
4. Chen, H., Lee, A.S., Swift, M., Tang, J.C.: 3d collaboration method over hololens and skype end points. In: Proceedings of the 3rd International Workshop on Immersive Media Experiences, pp. 27–30. ACM (2015)
5. Regenbrecht, H., Baratoff, G., Wilke, W.: Augmented reality projects in the automotive and aerospace industries. IEEE Comput. Graph. Appl. **25**(6), 48–56 (2005)
6. Chadalavada, R.T., Andreasson, H., Krug, R., Lilienthal, A.J.: Thats on my mind! robot to human intention communication through on-board projection on shared floor space. In: European Conference on Mobile Robots (ECMR) (2015)
7. Peppoloni, L., Brizzi, F., Ruffaldi, E., Avizzano, C.A.: Augmented reality-aided tele-presence system for robot manipulation in industrial manufacturing. In: Proceedings of the 21st ACM Symposium on Virtual Reality Software and Technology, pp. 237–240. ACM (2015)
8. Huang, W., Alem, L., Tecchia, F.: HandsIn3D: supporting remote guidance with immersive virtual environments. In: Winckler, M. (ed.) INTERACT 2013, Part I. LNCS, vol. 8117, pp. 70–77. Springer, Heidelberg (2013)
9. Ruffaldi, E., Brizzi, F., Filippeschi, A., Avizzano, C.A.: Co-located haptic interaction for virtual USG exploration. In: Proceedings of IEEE EMBC, pp. 1548–1551 (2015)
10. Ruffaldi, E., Brizzi, F.: CoCo - a framework for multicore visuo-haptic in mixed reality. In: De Paolis, L.T., Mongelli, A. (eds.) AVR 2016, Part I. LNCS, vol. 9768, pp. 339–357. Springer, Switzerland (2016)
11. Hinterstoisser, S., Holzer, S., Cagniart, C., Ilic, S., Konolige, K., Navab, N., Lepetit, V.: Multimodal templates for real-time detection of texture-less objects in heavily cluttered scenes. In: 2011 IEEE International Conference on Computer Vision (ICCV), pp. 858–865. IEEE (2011)
12. Garrido-Jurado, S., Muñoz-Salinas, R., Madrid-Cuevas, F.J., Marín-Jiménez, M.J.: Automatic generation and detection of highly reliable fiducial markers under occlusion. Pattern Recogn. **47**(6), 2280–2292 (2014)
13. Zaeh, M.F., Vogl, W.: Interactive laser-projection for programming industrial robots. In: IEEE/ACM International Symposium on Mixed and Augmented Reality, ISMAR 2006, pp. 125–128, October 2006
14. Olwal, A., Gustafsson, J., Lindfors, C.: Spatial augmented reality on industrial CNC-machines. In: Electronic Imaging 2008. International Society for Optics and Photonics, p. 680409 (2008)