20th IMEKO TC4 International Symposium and
18th International Workshop on ADC Modelling and Testing
Research on Electric and Electronic Measurement for the Economic Upturn
Benevento, Italy, September 15-17, 2014

A RGB-D Camera-Based Approach for Robot Arm-Hand Teleoperated Control

Francesca Cordella[1], Loredana Zollo[1], Eugenio Guglielmelli[1]

[1]*Laboratory of Biomedical Robotics and Biomicrosystems, Università Campus Bio-Medico, via Alvaro del Portillo 21, 00128 Roma, Italy.,*
Email: {f.cordella, l.zollo, e.guglielmelli}@unicampus.it

***Abstract* – In this paper a master-slave teleoperation system is proposed for reach-and-grasp operations in hazardous environments. A RGB-D camera and a robust marker-based Bayesian estimation method is proposed for tracking the human hand joint positions on the master side. Accuracy and repeatability of the estimation algorithm have been measured through the comparison with the data acquired through optoelectronic system. An anthropomorphic arm-hand robotic system is used on the slave side. The human operator can select the robot working area and the type of grasp action through finger motion. On the other hand, the human wrist motion is mapped into the robot workspace in order to teleoperate the robot end effector during the reaching phase. The optimal robot grasp configuration is automatically performed by the slave robot through a grasp synthesis algorithm previously proposed by the authors. The proposed approach has been experimentally validated on the KUKA LWR 4+ robot arm and the DLR-HIT Hand II.**

I. INTRODUCTION

Telerobotics is the field of robotics that treats operations in which the robotic system is remotely controlled by a human operator working in a safe environment [1]. In a telerobotic system the movements performed by the human operator on the master side can be acquired by different devices and are mapped into movements of the robotic platform, namely the slave, directly working in the hazardous environment. In particular, in telemanipulation different devices for acquiring the human movements can be used, e.g. datagloves, magneto-inertial sensors, vision-based tracking systems, electromyography-based control, and different approaches for mapping human to robot motion have been proposed in the literature [2].

Robot arm motion mapping is typically based on the forward-inverse kinematics in order to replicate, with the robot end effector, the pose of the human wrist. Nonlinear optimization methods have been used in order to maximize the similarity between the human movements and the corresponding robot movements [3]. However, the drawback of nonlinear programming algorithms is that they may reach local minima.

On the other hand, for hand motion mapping it is possible to identify four principal approaches [2]: fingertip mapping, joint-to-joint angle mapping, functional pose mapping and object specific mapping. The limitations of the approaches proposed in the literature regard the high cost of the adopted motion tracking systems and the complexity of the estimation algorithms.

The objective of this work is to propose a fast, cheap and easy method for estimating the master movements and replicating them on the robotic platform. In particular, this paper is focused on the acquisition of the human arm-hand configuration on the master side of a telemanipulation system through a RGB-D camera-based approach. To this purpose a Bayesian marker-based estimation technique has been developed for tracking the human hand joints and wrist position and a human-robot mapping approach has been implemented in order to remotely control an arm-hand robotic platform. The approach could be also used for a slave side composed of an arm-hand robotic system mounted on a mobile platform.

The coordination between the robotic systems could be managed by different combinations of master hand movements. On the master side, the Asus Xtion Prolive motion sensing device and suitable filtering techniques have been used for acquiring the human hand motion. Finger joint positions and wrist trajectories during hand movements have been reconstructed and used as reference for the robotic platform. Human wrist movements have been mapped into the robot workspace through an appropriate scaling factor, thus enabling robot preshaping. On the other hand, the grasping action is selected by the human operator through different finger movements and the corresponding robotic hand configuration is chosen in a grasping database (i.e. specific finger movements correspond to a database element which contains the robotic hand grasping configuration).

The proposed approach has been experimentally validated on an anthropomorphic arm-hand robotic system made of the KUKA LWR 4+ [4] and the DLR-HIT Hand II [5]. The communication between the motion sensing device and the arm-hand robotic platform has been realized via a UDP communication protocol that guarantees high dependability for real-time operations. A geometric

ISBN-14: 978-92-990073-2-7

inverse kinematics method [8] has been used for replicating the human wrist motion on the robot end effector and a position control in the joint space is chosen for controlling the robot arm movements.

The paper is structured as follows: in Section ii. the marker-based algorithm for estimating the hand joint positions is presented; in Section iii. the master-slave motion mapping method is proposed. Experimental results on the whole system are illustrated in Section iv.. Finally, Section v. reports conclusions and future work.

II. ESTIMATION ALGORITHM OF HUMAN HAND JOINT POSITIONS

In order to acquire the human hand finger joint and wrist Cartesian positions, 11 markers (made of blue paper) have been positioned on the human hand, as shown in Figure 3 (left side), and a fast detector based on color histogram and a connected component labeling algorithm have been developed. A marker-based approach has been chosen since it does not require a database of predefined hand configurations. Furthermore, the wrist does not require to be linked to other joints by a kinematic model, thus reducing the complexity of the reconstruction and consequently the computational cost.

The Asus Xtion Prolive motion sensing device working at $30fps$ has been used. It consists of an InfraRed (IR) laser emitter, an IR camera for measuring depth information and a RGB camera, with a resolution of 640×480. Then, a probabilistic approach has been applied in order to robustly track and label the markers attached to the hand. The algorithm, already proposed by the authors in [6], has been significantly improved in order to work in real-time and give information about the flexed finger. The algorithm output is the position of the wrist and finger joints and a binary string of five elements where each value is associated to one finger: value 1 is associated to a flexed finger, whereas value 0 is associated to an extended finger.

The blob detector gives the position of markers and clutters in the image plane. Therefore, it is necessary to make the algorithm robust with respect to the presence of outliers due to clutters and to the possible appearance/disappearance of markers from the scene. Being N the maximum number of markers expected in the scene and assumed that all the markers projections have linear-state dynamic and measurement models, driven by zero-mean white noise, the system state can be written as

$$\begin{cases} x_{k,j} = x_{k-1,j} + w_{k,j} \\ y_{k,j} = x_{k,j} + \nu_{k,j}, \end{cases} \quad (1)$$

where $w_{k,j}$ and $\nu_{k,j}$ are white Gaussian noises with null mean value. The output model is represented by the projection of the visible markers on the image space.

Since the association between a measure, given by the blob detector, and a marker or a clutter is not a priori known, it has been chosen to solve the estimation problem by using a general probabilistic approach, via robust Bayesian filtering [9]. Therefore, the problem can be defined by estimating the filtering (posterior) distribution defined as

$$\begin{aligned} p\left(\mathbf{x}_k|y_{1:k}\right) &= \sum_{a_k} p\left(\mathbf{x}_k, a_k|y_{1:k}\right) \\ &= \sum_{a_k} p\left(\mathbf{x}_k|a_k, y_{1:k}\right)\, p\left(a_k|y_{1:k}\right) \end{aligned} \quad (2)$$

where a_k is the latent variable modeling the measure-to-marker association [7], $p\left(a_k|y_{1:k}\right)$ is the posterior distribution of the data association represented by a set of particles which are recursively updated and reweighted, $p\left(\mathbf{x}_k|a_k, y_{1:k}\right)$ is the posterior (updated) distribution of the markers projections, subject to the association a_k and solved by using a Kalman filter. Therefore, the following relationship holds

$$p\left(\mathbf{x}_k|y_{1:k}\right) \approx \sum_{i=1}^{m} w_k^i\, N\left(\hat{\mathbf{x}}_k\left(a_k^i\right),\, P_k\left(a_k^i\right)\right) \quad (3)$$

where a set of m particles has been defined and contains the augmented state mean $\hat{\mathbf{x}}_k$, the error covariance matrix $P_{\hat{\mathbf{x}}_k}$ of the Kalman filter associated with the i-th sample and the weight w_k^i associated with each particle. The notation $N\left(\cdot,\cdot\right)$ indicates the multivariate normal distribution of order 2. Therefore, the filtering distribution $p\left(\mathbf{x}_k|y_{1:k}\right)$ has been solved by Eqs. (2) and (3). In fact, once generated all the possible hypotheses $a_k = i,\ i = 0, 1, \ldots, N$ and evaluated each of them together with the current observation (by running a Kalman filter), it is possible to evaluate the most likely hypothesis a_k^* which gives the highest score. This will give the most likely marker state according to (3).

The 3D position of each visible marker (with respect to the camera frame) can be reconstructed by multiplying the marker pixel coordinates coming from the estimation algorithm by the RGB camera calibration matrix and by the measured depth value. For the chosen application, the algorithm output are the wrist position during the whole hand movement and the distance, which represents the finger lengths, between the marker on the fingertips and the marker on the corresponding MetaCarpoPhalangeal (MCP) joint. A calibration phase is envisaged at the first frame in order to measure the finger lengths when they are fully extended. Afterwards, a threshold for identifying when a finger is flexed has been defined. To pass the threshold during hand movements implies a flexion of the finger and a corresponding set to 1 in the output binary array.

III. HUMAN TO ROBOT MOTION MAPPING METHOD

The whole system is made of a master side, i.e. the human operator, and a slave side, i.e. the arm-hand robotic platform. The master movements are acquired with the RGB-D camera and reconstructed with the algorithm described in Sect.2, implemented under ROS (Robotic Operating System) in order to ensure a real-time approach. In particular, eleven markers have been applied on the subject hand joints (see Sect.2). The robotic platform (right side of Fig. 1) is composed of the KUKA LWR 4+, which acts as the arm responsible for the reaching task, and an anthropomorphic robotic hand (i.e. the DLR-HIT Hand II) mounted on the KUKA LWR end effector and responsible for preshaping and grasping. The KUKA LWR is a 7 DoFs anthropomorphic robotic arm, which communicates with a remote PC through the Fast Research Interface (FRI) Library. It runs on a remote PC connected to the KUKA Robot Controller via UDP communication protocol that guarantees high dependability for real-time operations.

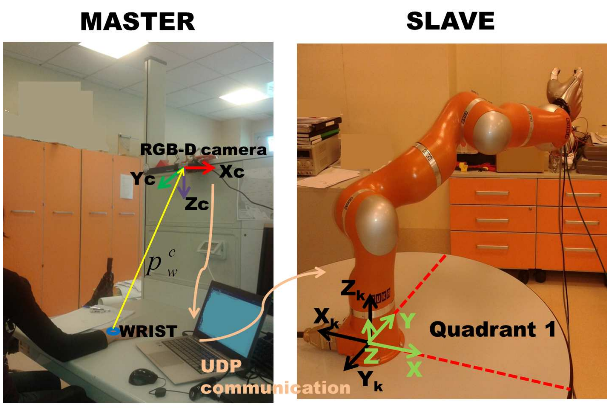

Fig. 1. Experimental Setup adopted for the teleoperated control. The system reference frames are outlined.

The first joint of the KUKA robot has a range of motion (RoM) of $\pm170^o$; therefore it allows dividing the robot workspace in 4 quadrants (Fig. 2). Finger flexion on the master side will identify the working quadrant: index finger is associated to the first quadrant, middle finger to the second quadrant and so on. Once selected the working area of the robot arm, it is necessary to make robot end effector reach the position defined by the human wrist through proper transformation matrices. Suppose for example to select the first quadrant. The wrist movements are mapped as translations along the 3 unit vectors (in green in Fig. 1) referred to the robot arm base frame. The following mapping has been chosen:

- A movement along the $\hat{x}_c$ unit vector of the camera frame corresponds to a movement along the unit vector $\hat{x}$ in the KUKA frame;
- A translation along the y_c axis implies a translation along the $\hat{y}$ unit vector in the KUKA frame;
- A movement along the z_c axis is a movement along the $\hat{z}$ unit vector in the KUKA frame.

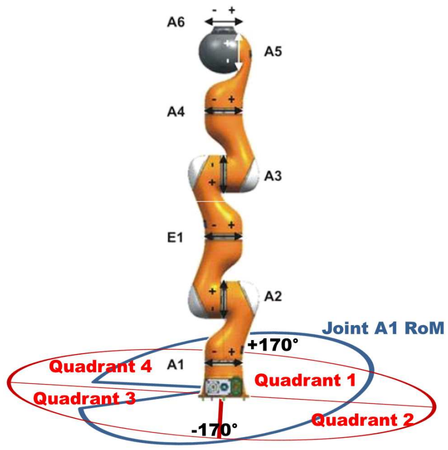

Fig. 2. The KUKA joints are outlined, so as the RoM of the base joint and the quadrants.

The rotation matrix from the camera frame to the KUKA frame can be expressed as $R_c^k = R_x(180^o)R_z(-90^o)$.

Once the movement directions have been defined as shown above, a scaling factor is introduced in order to adapt the robot workspace to the human wrist and RGB-D camera workspace. The wrist motion has been transmitted via UDP to the KUKA controller. Once the reaching phase is completed, the grasping action is selected by the user with appropriate finger movements. In detail: if the user flexes the index and the thumb to simulate a pinch grasp, the robotic hand will perform a pinch grasps; if the user flexes all the finger, the robotic hand will perform a power grasp; finally, if the user simulate a tripod grasp by flexing thumb, index and middle fingers, the task accomplished by the robotic hand is the tripod grasp. The selection of the grasping action by the user implies an automatic selection of finger configuration and robot arm end effector orientation.

IV. EXPERIMENTAL RESULTS

In the following, the results related to a planar movement along the x axis with respect to the camera frame are reported in Fig. 3. A cylindrical object of known geometric characteristics has been positioned in the first quadrant. To reach and grasp it, the subject on the master side flexed the index finger for selecting the robot working quadrant and then moved the hand towards the object. The position of all the marker centers were acquired with the proposed estimation algorithm.

The accuracy of the hand joint motion reconstruction algorithm has been evaluated by using an optoelectronic system (i.e. BTS Smart-D). The BTS Smart-D motion capture

Fig. 3. Simulated grasp configuration (Left) and final grasp configuration performed by the robot (Right).

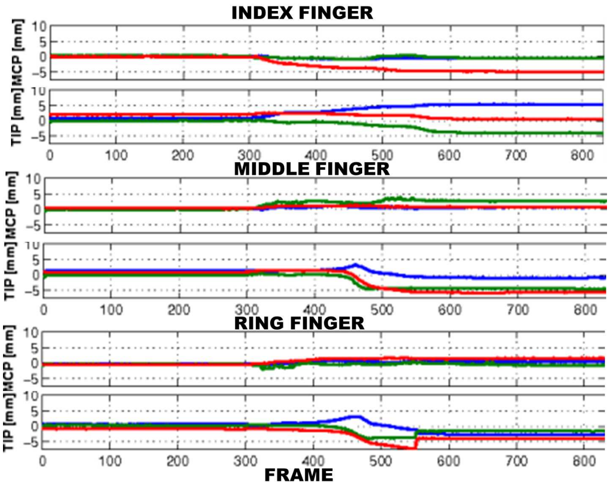

Fig. 4. Error between the marker 3D coordinates obtained with the BTS and the motion estimation algorithm.

system is a 7-camera motion analysis system with an acquisition rate of 60 Hz. Retroreflective markers have been positioned on the subject hand in the same configuration shown in Figure 3 (left side). The subject has been asked to perform the same movements performed during the acquisition with the RGB-D camera. The marker positions were measured with the optoelectronic system (i.e. our ground truth) and compared with the output of the estimation algorithm presented in Sect. ii.. The error of the estimation algorithm for index, middle and ring fingers during a power grasp is reported in Fig. 4. In particular, the position error of the tip and middle joints (y-axis of Fig. 4) is computed during the whole acquisition (x-axis of Fig. 4). The reaching action starts around frame 300 and the grasping action finishes around frame 600. Therefore, before and after these instants, the error is constant. As evident, the error is always less than 5 mm, which is acceptable for the application proposed in this paper. Similar values have been obtained for the other type of grasps and during several repetitions of the same grasp configuration. In particular, the mean error ($\pm SD$) calculated during three repetitions of the same task is around $0.16mm \pm 0.33$ for the MCP joints and 0.41 ± 0.28 for the TIP joints.

The wrist motion has been scaled and real-time replicated by the KUKA end effector. In Figs. 5, 6 the human wrist and the robot end effector trajectories are reported respectively. The results show the estimation algorithm robustness with respect to outliers.

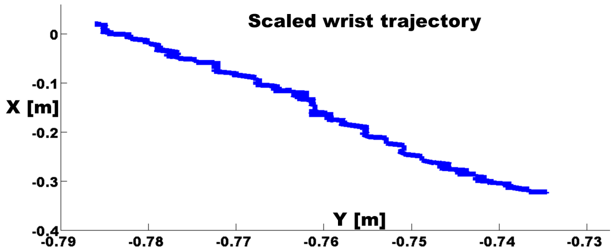

Fig. 5. Scaled human wrist trajectory.

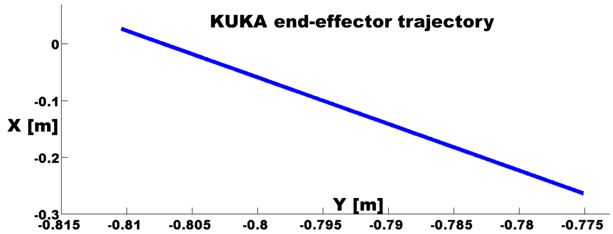

Fig. 6. Robot end effector trajectory.

After reaching, the subject flexed all the fingers to select the type of grasp. An array composed by five elements equal to 1 has been generated in order to select the preshaping (i.e. the wrist orientation) and grasping configurations (i.e. the hand joint angles) resorting to the results of our previous work on grasp synthesis in [10]. The hand finger joint torques during grasping have been measured. They are listed in Table 1. Results show that the thumb and the index finger are the fingers that apply higher forces in the grasp, having higher torques than the others.

Table 1. Finger hand joint torques. Values are expressed in mNm.

CYLINDER					
DoF	*thumb*	*index*	*middle*	*ring*	*little*
F/E PIP	18.7	-2.7	-3.0	-1.7	5.0
F/E MCP	256.3	177.7	-5.0	3.7	94.3
A/A MCP	34.3	98.7	146.0	91.3	74.3

V. CONCLUSIONS

In this paper a RGB-D camera-based approach for teleoperation has been proposed. The master side is composed of a human operator whose hand joint motion (wrist included) is acquired with a RGB-D camera and estimated via a Bayesian approach. Accuracy and repeatability of the estimation algorithm have been measured through the comparison with the data acquired through optoelectronic system (working as ground truth). The slave side of our teleoperated system consists of an arm-hand robotic system made of the KUKA LWR 4+ robot arm and the DLR-HIT Hand II. The human operator can select the working portion of the robot workspace and the type of grasp action through finger motion. On the other hand, the human wrist motion is mapped into the robot workspace and is used to directly teleoperate the robot end effector during the reaching phase. The optimal robot grasp configuration is automatically performed by the slave robot through a grasp synthesis algorithm proposed in [10]. The proposed approach has been experimentally validated. Future works will be addressed to (i) extend the algorithm to wrist orientation in order to teleoperate the complete pose of robot end effector; (ii) analyze the possibility of considering a marker-free approach for the hand pose estimation; (iii) extend the approach to a mobile manipulator on the slave side; (iv) coordinate motion of the mobile platform and with the arm-hand robotic system; (v) provide the operator on the master side with a haptic feedback.

VI. AKNOWLEDGMENT

This work was partly supported by the Italian Ministry of Instruction, University and Research with PRIN Hand-Bot project (CUP: B81J12002680008), partly by the Italian Ministry of Health with NEMESIS project, partly by the Neurohand project (DTB2 Filas) and partly by the PPR2 project supported by INAIL.

VII. *

References

[1] G. Niemeyer, C. Preusche, G. Hirzinger, "Telerobotics", In B. Siciliano, O. Khatib (ed.), Springer Handbook of Robotics, pp. 741–757, 2008.

[2] M.V. Liarokapis, P.K. Artemiadis, C.P. Bechlioulis, K.J. Kyriakopoulos, "Directions, Methods and Metrics for Mapping Human to Robot Motion with Functional Anthropomorphism: A Review", Technical Report, 2013.

[3] M. Do, P. Azad, T. Asfour, R. Dillmann, "Imitation of Human Motion on a Humanoid Robot using Non-Linear Optimization", IEEE-RAS International Conference on Humanoid Robots, 2008.

[4] A. Albu-SchZffer, S. Haddadin, C. Ott, A. Stemmer, T. WWimbock, G. Hirzinger, "The DLR ightweight robot: design and control concepts for robots in human environments", Industrial Robot An international journal, vol. 34, no. 5, pp. 376–385, 2007.

[5] H. Liu, K. Wu, P. Meusel, N. Seitz, G. Hirzinger, M.H. Jin, Y.W. Liu, S.W. Fan, T. Lan, Z.P. Chen, "Multisensory Five-Finger Dexterous Hand: The DLR/HIT Hand II", IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 3692–3697, 2008.

[6] F. Cordella, F. Di Corato, L. Zollo, B. Siciliano, P. van der Smagt, "Patient performance evaluation using Kinect and Monte Carlo-Based fnger tracking", IEEE RAS/EMBS International Conference on Biomedical Robotics and Biomechatronics, pp. 1967–1972, 2012.

[7] F. Cordella, F. Di Corato, G. Loianno, B. Siciliano, L. Zollo, "Robust Pose Estimation Algorithm for Wrist Motion Tracking", IEEE/RSJ International Conference on Intelligent Robots and Systems, 2013.

[8] E. Lopez, L. Zollo, E. Guglielmelli "Teleoperated Control based on Virtual Fixtures for a Redundant Surgical System", IEEE/RSJ International Conference on Intelligent Robots and Systems, 2013.

[9] S. Sarkka, A. Vehtari, J. Lampinen, "Rao-blackwellized Monte Carlo data association for multiple target tracking", International Conference on Information Fusion, vol. 7, pp. 583-590, 2004.

[10] A. Provenzale, F. Cordella, L. Zollo, A. Davalli, R. Sacchetti, E. Guglielmelli, "A grasp synthesis algorithm based on postural synergies for an anthropomorphic arm-hand robotic system", IEEE RAS/EMBS International Conference on Biomedical Robotics and Biomechatronics, 2014.