Original approaches to interpretation learning, and modelling, from the observation of human manipulation

Francesco Corato*, Pietro Falco*, Martin Lösh†, Emilio Maggio‡, Jäkel Rainer†, and Luigi Villani§

*Dip. di Ingegneria dell'Informazione, Seconda Università di Napoli, Aversa, Italy
Email: pietro.falco@unina2.it
Email: fcorato@unina.it
†FZI, Karlsruhe, Germany
Email: loesch@ira.uka.de
Email: jaekel@ira.uka.de
‡OMG plc, Oxford, UK
Email: Emilio.Maggio@vicon.com
§Dip. di Informatica e Sistemistica, Università di Napoli Federico II, Napoli, Italy
Email: lvillani@unina.it

I. Introduction

The DEXMART project is focused on artificial systems reproducing smart sensory-motor human skills, which operate in unstructured real-world environments. The emphasis is on manipulation capabilities achieved by dexterous and autonomous, and also human aware dual-arm/hand robotic systems.

Manipulation cannot be fully planned without human interaction and intervention. Even then, human demonstration, advice and correction are important parts of robot manipulation learning and execution. Research will target methods and mechanisms to observe human manipulations in a way that the observed actions enhance the robots skills (e.g. by new or optimized skills) and tasks (e.g. due to the optimized, context and goal depending combination of skills).

A number of challenges have to be tackled about the observation and interpretation of manipulation activities performed by humans, first of all setting up a suitable sensory system and reconstructing motion with high accuracy. Furthermore, the question of what is the best kinematic model of the human hand has to be answered.

Subsequently to the fundamental questions of observation and modelling of the human hand, the challenge is to build systems which are able to learn from the observation of manipulations demonstrated by a human. Different robot learning tasks in this context reach from simple recording and replaying of observed trajectories to the abstraction and generalization of the results of observed manipulations. Often the observation of the hand(s) only is not sufficient, but also the context, e.g. manipulated objects, have to be considered depending on the complexity of the learning goal: more advanced sensor fusion algorithms which integrate joint angular positions, marker positions and contact force measurements will be developed. In the following, we will limit the presentation to learning goals which can be learned using only observations of the human hand. In this case, two main questions remain:

1) What can be learned from the observation and how does the actual learning take place?
2) What are relevant features for the learning and how can they be selected?

In the *DEXMART* project, learning is pursued on two different levels of abstraction, the so-called *action level* and *activity level*. While the activity level is more abstract, the action level covers the interfacing between subsymbolic representations and atomic symbolic operations.

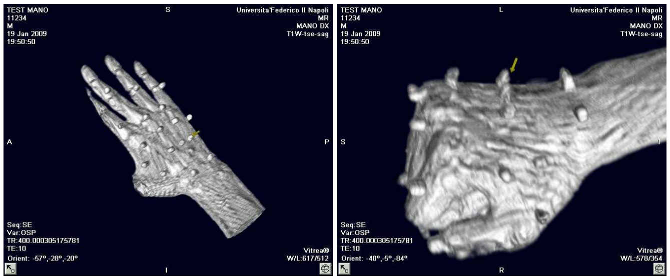
II. Observation of the Human Hand

A. Motion Capture

A number of problems must be tackled to carry out highly-accurate observation of the human hand:

- marker movements due to skin motion with respect to the bones
- marker occlusions
- positioning of the marker-set

Marker movements due to skin motion are a real problem for hand motion capture as the error can have the same order of magnitude than the bone lengths. Existing models treat the marker positional error as a residual covariance. This assumption simplifies the mathematical formulation. However, residual error analysis conducted within DEXMART showed that it may be possible to improve the predictive power of the kinematic model by explicitly accounting for marker movements. To this extent, we developed a novel kinematic calibration procedure that accounts for soft tissue artifacts



(a) (b)

Fig. 1. MRI measurements for the marker RMM4 on open (left column) and closed (right column) hand. (a)-(b): marker spatial position (yellow arrow)

by allowing the markers to move according to polynomial functions of the joints angles. To validate the motion marker model and to increase the knowledge about marker slipping, an MRI system has been used to capture the human hand in two different static poses, shown in Fig. 1. The MR images have been used to reconstruct a three-dimensional model of the hand bones according to a co-registration procedure whose details can be found in [4]. The measurement results show that the marker slipping over the bones is quite significant ranging between 2.5 mm and 10 mm along the hand axial direction.

Missing marker measurements due to occlusions can heavily affect the reliability of the joint angle measurements. This problem was tackled from two directions: (i) investigation about the optimal number of markers and their positioning; (ii) study about the fusion of data-glove angular information with the marker measurements. Regarding the marker-set, two possible configurations were tested: a minimal marker set with approximately one marker per segment and a redundant marker configuration with three or more markers per segment. To handle the occlusion problem using the minimal marker set, in [1] a sensor fusion algorithm is presented. It ”fuses”, through a Kalman-like algorithm, positions in the space of a set of reflective markers and measurements from low-cost angular sensors disposed on the finger joints. The algorithm allows real-time tracking of complex hand movements.

B. Kinematic model of the human hand

Another key topic is the question of what is the best kinematic model for hand data. There is no generally accepted methodology for making such a choice. Ideally an objective comparison between different kinematic models would require the actual joint angle values measured with a procedure that guarantees higher accuracy than motion capture measurements. Unfortunately, acquiring accurate ground truth data almost always involves invasive procedures. Many have used intra-cortical (bone) pin mounted markers or capitalized on external bone fixation already in place [2]. As direct measurements are impractical, researchers have evaluate other desirable model qualities such as repeatability or have used synthetic or semi synthetic data [3]. Although repeatability is a well established technique for model validation, it does not tell us how well the model explains the data. To this extent, OMG studied statistical model selection for hand kinematics and tested four different algorithms, namely: Bayesian Information Criterion (BIC), Akaike Information Criterion (AIC), and Consistent AIC (CAIC) and a procedure based on Bootstrapping the data. Among these methods the results show that BIC, CAIC and Bootstrapping can provide a good insight on the number of DoF to assign to each articulation of the hand.

After the selection of the number of DOF’s, a quantitative analysis of joint interdependencies has been done aimed at studying kinematic synergies during reaching and hand grasping activity. Four male subjects have been asked to perform two tasks: a cylinder-grasp with right hand and a voluntary flexion and extension of each individual finger. From acquired data, the computation of the correlation coefficient matrix for all the DOF’s has been carried out, aimed at quantifying the degree of correlation of each DOF with all the others. The results obtained in this study will be used both for reducing the complexity of trajectory planning and for improving the sensor fusion algorithms of kinetostatic data owing to the increased a priori knowledge about the hand kinematics.

III. Learning from observation

To find the most relevant features for the learning, an approach in two steps is investigated. In the first step, the available sensors are exploited to extract and derive as much features as possible. These features include, but are not limited to joint angles, trajectories of all joints and finger tips, velocities of them, statistical features of these points. To accommodate for two-handed manipulations, features like correlations of movements of both hands, temporal and spatial synchronization points are also investigated. Using training data from fundamentally different actions, ranging from simple to complex, all features are analyzed and ranked according to their utility for separating between different action classes. In the second stage, from the complete set of features only the most relevant for a current learning task are selected using an information content-based method for the actual training process. This minimizes the noise in the training data. A complementary method for data reduction is to abstract from sets of fingers, which serve the same purpose in a grasp or manipulation action, to a single finger which combines force and contact information of the whole set.

Acknowledgment

The research leading to these results has received funding from the European Community’s Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 216239 (DEXMART project).

References

[1] P. Falco. Sensor Fusion based on the EKF for human and robotic hand tracking application. Master Degree Thesis, 2008

[2] J. Fuller, L. Liu, M.C. Murphy, and R.W. Mann. A comparison of lower-extremity skeletal kinematics measured using skin- and pin-mounted markers. In *3-D Analysis of Human Movement*, volume 16, pages 219–242, 1997.

[3] I.W. Charlton, P. Smyth, and L. Roren. Repeatability of an optimized lower body model. *Gait and Posture*, 20:213221, 2004.

[4] The DEXMART consortium, Kinematic Modelling of the Human Hand. *Dexmart Deliverable D1.1*, 2008.