2020 I-RIM Conference
December 10-12, 2020, Rome, Italy (Virtual)
ISBN: 9788894580518
DOI:10.5281/zenodo.4781580

Towards the Estimation of Object Characteristics by Observing Human Manipulation

Linda Lastrico[*1,2], Alessandro Carfì[2], Francesco Rea[1], Fulvio Mastrogiovanni[2], Alessandra Sciutti[1]

***Abstract*—The outstanding ability to detect implicit cues in everyday gestures makes the interaction between humans smooth and seamless. We propose a method to provide a robot with the same capability, studying the scenario of objects handling. The final goal of our on-going work is to enable robots to autonomously infer the properties of an object manipulation action by observing how humans perform it.**

***Index Terms*—Human Motion Understanding, Human-Robot Interaction, Non-Verbal Communication, Deep Learning**

I. Introduction

As humans, we exchange a considerable amount of information with non-verbal signals, through body posture and body movements. People are able to correctly estimate the weight of an object, simply by observing another person lifting it, and the same information can be communicated by the lifting action of a humanoid robot [7]. In human-human interaction, the ability to infer object properties, while observing others manipulating it, is linked to motor resonance between the observer and the performer. The same set of neurons are activated during both action execution and observation, and this provides a common description of our and others' behavior. Since in normal conditions we know the consequences of our own actions, when we observe others we can immediately recognize and understand theirs [5].

In order to achieve a seamless collaboration, robots should understand the relevant characteristics of human actions, also by correctly interpreting the implicit signals concealed in them [1]. If we consider a collaborative scenario, estimating the characteristics of handled objects allows the robot to plan a safe and efficient coordinated motion. For instance being cautious is an optimal robot behaviour when in presence of fragile or slippery objects. The scientific question we address in our research is whether and how the features of an object can be inferred just by looking at the human transporting it [3]. Our approach of estimating objects properties by relying on human kinematics information during manipulation allows us to generalize over previously unseen items. We focused on two features that influence how we handle an object, namely its weight and the carefulness required to move it. The first relates to the velocity we adopt to lift an object; the latter can be influenced by multiple factors, such as the item stiffness, the content about to be spilled, the risk for the object to fall or its fragility. After a preliminary analysis, we found that human motion profiles change depending on object properties. This is confirmed by a recent study on the same topics, where the carefulness in the handling was detected on the basis of wrist position and velocity [2]. Therefore, we used some kinematic features, derived from the observed motion, to train Deep Learning classifiers with the intention to discriminate between different features of handled objects.

Fig. 1: View of the experimental setup with a volunteer in a rest position. The two shelves with the glasses on them, the motion capture markers and the iCub robot are visible

II. Methods

The experimental setup is shown in Figure 1. We acquired the data of 15 participants while performing a series of reaching, lifting and transportation movements of four transparent glasses, identical in shape and appearance. The four glasses were characterized by two weight levels, and by two different levels of carefulness required in their handling, obtained by filling two of them with water till the brim. The data were collected using two different sensors for comparison. The kinematics of human motions was recorded with the Optotrak Certus®, NDI, motion capture (MoCap) system, via active infrared markers placed on the right hand. The other source of information was the left camera of the iCub, which was located opposite to the table. As motion descriptor, from the saved raw images we computed the Optical Flow (OF), estimating the apparent motion vector for each pixel of the image. The same set of motion representations was then extracted from both the motion capture data and the optical flow, i.e., the norm of the velocity, the angular velocity, the radius of curvature and

This paper is supported by the European Commission within the Horizon 2020 research and innovation program, under grant agreement No 870142, project APRIL (multipurpose robotics for mAniPulation of defoRmable materIaLs in manufacturing processes) and CHIST-ERA (2014-2020), project InDex (Robot In-hand Dexterous manipulation).

[1] Istituto Italiano di Tecnologia, Genoa, Italy
[2] The Engine Room, DIBRIS, Università degli Studi di Genova, Genoa, Italy
[*] linda.lastrico@iit.it

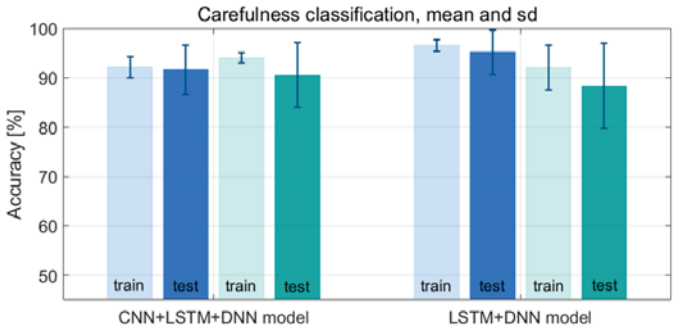
the curvature (see [3], [8] for a detailed description). These features were chosen since they can be easily estimated at every time instant, ideally allowing for classifying the object before the end of the observed action. Weight and carefulness discrimination were approached using two binary classifiers, one for each feature, trained with the four motion features extracted during transportation movements. Two possible deep models for classifying time dependent data were adopted. The former is inspired by [6] and consists of a combination of a Convolutional Neural Network (CNN), Long-Short Term Memory (LSTM), and a Deep Neural Network (DNN). The latter is a simpler LSTM-DNN model. The dataset consisted of 876 total trials. A Leave-One-Out approach was chosen, using for every fold the data of a different participant as test set.

III. Results

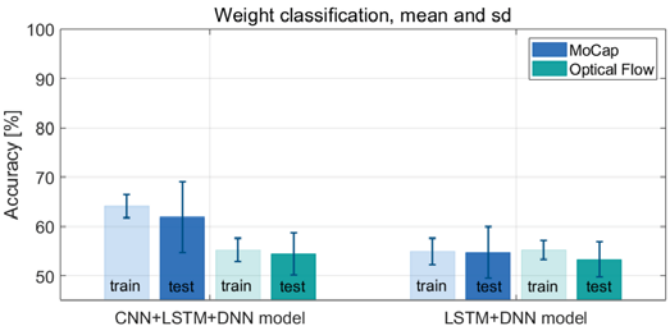
As presented in Figure 2a, the performance of the carefulness classification are good with both the models and the sensing modalities. Interestingly, even the OF from the robot's camera (single point of view) grants an accuracy above 85-90%. The classification of the weight is more difficult, and the results are not as good (Figure 2b). Indeed, we obtained an accuracy around 60% for the first model trained with the MoCap data, while 55% in the other cases. This result may be due to different concurrent factors. The presence of water in some of the glasses may have led the subjects to focus mainly on the carefulness feature, unconsciously overlooking the weight difference. Moreover, previous studies showed how the vertical component of the velocity in a movement is informative about the weight transported [7]. In this dataset, there was a great variability in such dimension, with movements going from the table to the shelves, from top to bottom and vice versa. The first classifier was tested against these two hypotheses, but no significant improvements in accuracy have been achieved. A possibility for future work is focusing on the vertical component of the velocity and exploring these additional hypotheses on reasonably extended datasets to obtain more reliable results.

IV. Conclusions

The proposed classification approach relies exclusively on human body kinematics, overlooking the external appearance of the object, granting the ability of generalizing over previously unseen items. Given the promising results in the carefulness classification, despite the variability in the data, we are currently working on a real-time discrimination while manipulations occur. It should be noted that the movements used in the training phase were acquired in a non-social context. The participants were simply asked to move the glasses from one position to another, but no communicative intention towards observers was required. It is possible, as the signaling theory suggests, that in a collaborative context humans' gestures would become more communicative of the objects features, to facilitate the observer's interpretation of the action [4].



(a) Carefulness classification



(b) Weight classification

Fig. 2: Mean accuracy obtained in the carefulness (2a) and the weight classification (2b), using the features extracted from MoCap data, in shades of blue, and from the Optical Flow, in shades of green

References

[1] Dragan, A.D., Lee, K.C.T., Srinivasa, S.S.: Legibility and predictability of robot motion. In: Proceedings of the 8th ACM/IEEE International Conference on Human-Robot Interaction. Tokyo, Japan (2013, March)

[2] Ferreira Duarte, N., Chatzilygeroudis, K., Santos-Victor, J., Billard, A.: From human action understanding to robot action execution: how the physical properties of handled objects modulate non-verbal cues. In: Proceedings of the 10th International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob). Valparaíso, Chile (2020, October). In press

[3] Lastrico, L., Carfì, A., Vignolo, A., Mastrogiovanni, F., Sciutti, A., Rea, F.: Careful with that! observation of human movements to estimate objects properties. In: Proceedings of the 13th International Workshop of Human-Friendly Robotics (HFR). Innsbruck, Austria (2020, October). In press

[4] Pezzulo, G., Donnarumma, F., Dindo, H.: Human sensorimotor communication: A theory of signaling in online social interactions. PloS One **8**, 1–11 (2013)

[5] Rizzolatti, G., Fadiga, L., Fogassi, L., Gallese, V.: Resonance behaviors and mirror neurons. Archives italiennes de biologie **137 2-3**, 85–100 (1999)

[6] Sainath, T.N., Vinyals, O., Senior, A., Sak, H.: Convolutional, long short-term memory, fully connected deep neural networks. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Brisbane, Australia (2015, April)

[7] Sciutti, A., Patane, L., Nori, F., Sandini, G.: Understanding object weight from human and humanoid lifting actions. Autonomous Mental Development, IEEE Transactions **6** (2014)

[8] Vignolo, A., Noceti, N., Rea, F., Sciutti, A., Odone, F., Sandini, G.: Detecting biological motion for human–robot interaction: A link between perception and action. Frontiers in Robotics and AI **4**, 14 (2017)