

Chapter 1

Predicting Object Weights from Giver's Kinematics in Handover Actions

Lena Kopnarski*, Laura Lippert*, Claudia Voelcker-Rehage, Daniel Potts, and Julian Rudisch

Abstract Handover actions describe the action when an object is handed over from one actor (human/robot) to another. A requirement for a smooth handover action is precise coordination between the two actors in space and time. Part of a handover action are reach and grasp movements. In order to be able to perform adequate reach and grasp movements, precise models regarding the object properties are necessary, only then anticipatory grip force scaling can take place. It is possible that receivers in handover actions observe the giver during object manipulation in order to estimate the object weight more accurately. Knowledge about the change in kinematics due to object weight in handover actions can be used to improve human-robot interactions by providing robots with better weight estimation through prediction based on human kinematics. The aim of this study was to investigate whether predictions about the object weight can be achieved from the kinematics of the giver in a handover action. Furthermore, the aim was to analyze which joint angles are particularly suitable for classifying the object weight (i.e., are most influenced by the object weight).

Keywords classification, handover, kinematics, object weight, pattern recognition

* contributed equally.

Lena Kopnarski, Claudia Voelcker-Rehage, Julian Rudisch
Institute of Sport and Exercise Sciences, University of Münster, D-48149 Münster, Germany, e-mail: lena.kopnarski@uni-muenster.de

Laura Lippert, Daniel Potts
Faculty of Mathematics, Chemnitz University of Technology, D-09107 Chemnitz, Germany, e-mail: laura.lippert@math.tu-chemnitz.de

1.1 Introduction

Joint handover actions are performed almost every day. Whether it is handing over the stapler to a colleague at work, the glass of wine to the partner at home or to exchange cash in the supermarket. In order to carry out a smooth handover action, both actors need to coordinate precisely in space and time [Sebanz et al., 2006]. Handover actions are a joint action that can be divided into individual sub-actions. Thus, a handover action includes reaching and grasping as well as carrying/transporting an object comparable to a replacement task.

For reach, grasp and manipulation tasks, precise scaling of grip and load forces is necessary. Motor control research has shown that grip and load forces are already planned in anticipation [Hermsdörfer et al., 2011]. A suitable grip force must be large enough to overcome the load force and prevent the object from slipping out of the fingers, but must also avoid being too large, so that the object is not crushed or the person is not getting fatigue. The necessary grip and load forces depend on both the intended action and the object properties. Accordingly, the successful scaling of the grip and load forces depend a lot on the accuracy of the estimates of the object properties. A handover action is a time-critical action for both actors at the moment of object transfer. While both actors are in physical contact with the object, the giver reduces and the receiver increases grip forces rapidly [Mason and Mackenzie, 2005]. Therefore, precise anticipatory grip-force scaling is of particular importance. Estimation of object properties (such as weight) is usually based on previous experience and knowledge [van Polanen and Davare, 2015]. Furthermore, it is also possible that information about the weight of an object is additionally obtained from the observed kinematics of another person lifting or moving the object [Hamilton et al., 2007]. This means that heavy objects can influence the kinematics of movement differently than light objects (e.g., the joint angle configuration). When an object is handed over from one person to another, the receiver can observe this movement and obtain information to create an accurate forward model on the receiver's side.

An accurate estimation of the object weight also plays a major role for the grip force scaling of robots [Copot et al., 2016]. If robots are confronted with the task of grasping and transporting different, unknown objects, it is necessary that they can estimate the object weight in order to produce a suitable grip force. So far, predicting the object weight before the robot has physical contact with the object is a major challenge that has already been addressed by different approaches such as image recognition [Standley et al., 2017] or thermography [Aujeszky et al., 2019]. In the context of hybrid societies, the human-robot handover scenario plays a central role. As a the receiver, predicting the weight of the object through the kinematics of the giver could allow the robot to anticipate appropriate grip forces, even if it is an unknown object. This would give a big advantage over approaches like image recog-

dition, where object classes have to be learned first and therefore can never cover the variety of everyday objects.

In this study, we therefore want to investigate how changes in the weight of an object can be identified from the motion kinematics of a giver during a handover task. Furthermore, we investigate which joint angle contributes most to the classification of object weights in order to find out which kinematic characteristic of the giver should be considered most in a handover movement in order to provide the most reliable prediction. Hence, we recorded kinematics in handover actions in which the weight of the object to be handed over was varied. The aim was to classify the kinematics (time-profiles of joint angles) of the giver using a support vector machine (SVM) and thus to predict the object weight.

1.2 Methods

1.2.1 *Participants*

Forty healthy subjects (31 female) aged 22.0 ± 4.3 years participated in the experiment, thus data were collected from a total of 20 dyads in the handover experiment. All subjects had normal or corrected to normal vision, no psychiatric or neurological disorders, and no orthopedic upper limb impairments. According to the Edinburgh Handedness Inventory [Oldfield, 1971], 39 subjects were right-handed and one subject was ambidextrous. This study was approved by the Ethics Committee of the Chemnitz University of Technology, Faculty of Behavioral and Social Sciences, on July 12, 2019 – number V-343-17-CVR-SFB_A01-24062019.

1.2.2 *Materials*

A passive-marker based optical motion capture system (Vicon Motion Systems Ltd, Oxford, UK) with 10 cameras (5 Vantage, 5 Vero) was used to record subject motions at a sampling frequency of 100 Hz. Sixteen spherical reflective markers with a diameter of 6.4 mm were used for the upper body (head, trunk, shoulders, right arm). We used a marker set based on the PlugIn Gait Model [Vicon, 2020]. The following joint angles were extracted for the right arm: shoulder (flexion/extension, abduction/adduction, internal/external rotation), elbow (flexion/extension), and wrist (flexion/extension, internal/external rotation, ulnar/radial deviation).

Two different self-constructed, 3D printed test objects were used. Test objects included transducers for the measurement of grip forces (not used

for this study) and 5 infrared LEDs which enabled tracking in the Vicon system. Two different test objects, differing in size, were used. Both objects had an identical base body ($8\text{ cm} \times 8\text{ cm} \times 8\text{ cm}$), which contained both the LEDs and the possibility of attaching weights inside the object. The grasping surfaces, which differ in size and distance from each other ($5\text{ cm} \times 5\text{ cm} \times 5\text{ cm}$; $8\text{ cm} \times 8\text{ cm} \times 8\text{ cm}$) between the two objects, were located above the base body and were arranged one above the other (see Figure 1.1). The lower (blue) grasping surfaces were used by the giver, the upper (yellow) by the receiver. Three different object weights were used so that the handover object weighed 400 g in the light condition, 700 g in the medium condition and 1000 g in the heavy condition (weight conditions were the same for both object sizes).

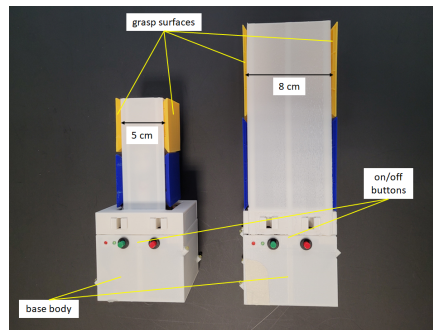


Fig. 1.1: Small (left) and big (right) object. Weights can be embedded in the base body.

1.2.3 Procedure

Participation in the study consisted of two test sessions with about 7 to 14 days in between. The first session consisted of different motor and sensory tests of the right hand, a questionnaire and a replacement task (no joint action, not considered further here). In the second session, two subjects sat opposite each other at a table. At the start of a trial, the object was placed on a foam pad ($17\text{ cm} \times 20.5\text{ cm}$) fixed centrally to the table on the right-hand side of the giver (see Figure 1.2). The subjects were instructed to perform a handover action as natural as possible. After an acoustic signal, the giver grasped the object at the lower grasping surfaces (blue) and handed it over to the receiver, who grasped it at the upper grasping surfaces (yellow). The receiver then placed the object on a foam pad on the other side of the table (see Figure 1.2), which ended the trial. Subjects performed handover actions with all six object configurations, with object size blocked and object weight

presented in a pseudo-randomized order. Each subject was assigned either the role of the giver or the receiver, with the assignment being swapped halfway through the trials. Across all trials, each condition was performed 10 times (2 roles \times 2 object sizes \times 3 object weights), resulting in four blocks and 120 trials. In total, our data set has $M = 2400$ trials (40 givers, with 60 trials each).

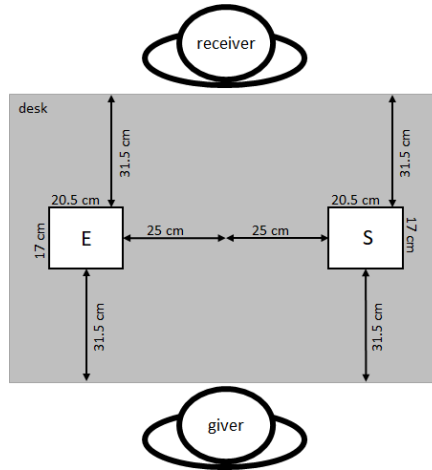


Fig. 1.2: Experimental set up.

1.2.4 Analyses

1.2.4.1 Data Preprocessing

The used data for our approach are times series, denoted by $\alpha_i(t)$ for the joint angle $i = 1, \dots, 7$ belonging to the three wrist, three shoulder and one elbow angles. In order to compare the time series, they were cut to their individual start and end time stamp. The starting point is the time after which the giver has grasped the object. This is defined by the moment when the velocity of the object exceeds a threshold of 0.02 m/s for the first time. The end point is reached when givers and receivers wrist have minimal distance. Consequently, due to differences in the movement, the time series have different lengths for different trials. This can be fixed by a time normalization, where we use for every trial instead of the given time stamps t_1, \dots, t_e the normalized time stamps

$$0, \frac{t_2 - t_1}{t_e - t_1}, \frac{t_3 - t_1}{t_e - t_1}, \dots, \frac{t_{e-1} - t_1}{t_e - t_1}, 1.$$

Hence, we modeled every angle by a function depending on the time,

$$\alpha_i(t), \quad t \in [0, 1], \quad i = 1, \dots, 7, \quad (1.1)$$

which we have given at discrete time points and has length 1. Missing data was interpolated linearly and some trials were discarded due to too many missing values, which reduced the total amount of 2400 trials. After filtering due to data recording errors we received $M = 2256$ trials. The time series $\alpha_i(t)$ are plotted in Figure 1.3 for some random trials from one person.

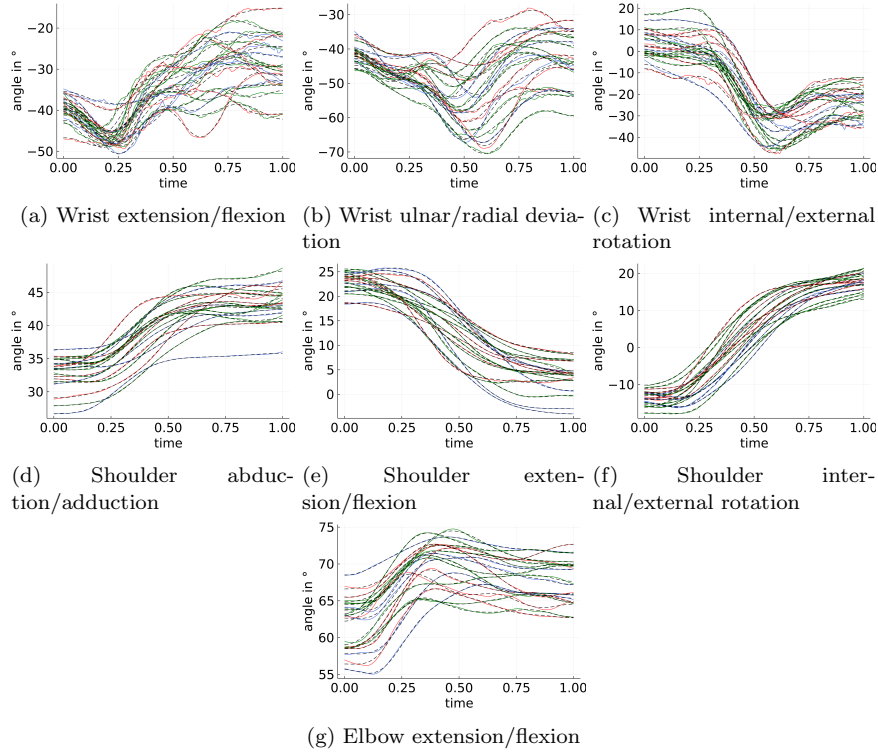


Fig. 1.3: A subset of the time-series $\alpha_i(t)$ (solid) of one person for the 7 joint angles, together with the approximated time series $\tilde{\alpha}_i(t)$ (black, dashed). The colors belong to the light (blue), medium (red) and heavy (green) object.

1.2.4.2 Prediction procedure

The aim was to extract the important information from the time series $\alpha_i(t)$ which classify the objects weight and to find out which joint angles are impor-

tant for the classification. Each time series $\alpha_i(t)$ can be viewed as a smooth function defined on $[0, 1]$, which allows for a decomposition in basis functions. A useful basis in this case is the half-period-cosine basis, which allows us for every angle i to approximate the time series well by

$$\alpha_i(t) \approx \sum_{k=0}^{n-1} a_k^{(i)} \cos(\pi kt) =: \tilde{\alpha}_i(t), \quad i = 1, \dots, 7. \quad (1.2)$$

The half-period cosine basis is a good choice for the approximation of non-periodic functions, since in this case the decay rate is $\mathcal{O}(n^{3/2})$ and the coefficients $a_k^{(i)}$ can be calculated easily and fast from the time series at discrete points by using the discrete cosine transform (DCT), see [Plonka et al., 2018, Chapter 6]. Hence, we described the time series $\alpha_i(t)$ by $n = 8$ coefficients $a_k^{(i)}$ with $k = 0, \dots, 7$, which is a very compressed expression in contrast to the full time series. Furthermore, $n = 8$ is a reasonable choice between over- and underfitting. For smaller n the error between $\alpha_i(t)$ and $\tilde{\alpha}_i(t)$ is too big, whereas for bigger n the approximation fits to much the noise in the measured data. The approximated time series $\tilde{\alpha}_i(t)$ are plotted as dashed lines in Figure 1.3. An other advantage of this procedure is that the sum of cosine functions smooths out the measurement inaccuracies, which led to noisy data in the original time series. Each trial has $n = 8$ Cosine-coefficients for each of the considered 7 joint angles. This can be seen as data compression in comparison the raw time series, since there are only 56 degrees of freedom for each trial.

We have the label vector $\mathbf{y} \in \{1, 2, 3\}^M$, which assigns 1, 2 and 3 to the trials with light, medium and heavy object, respectively. When we use all 7 joint angles for the prediction, we receive a matrix $\mathbf{X} \in \mathbb{R}^{56 \times M}$ containing the coefficients $a_k^{(i)}$ for every trial. For the classification we use Julia’s SVM, which is contained in LIBSVM in the Machine Learning package. Different strategies for a train/test split and cross-validation (CV) are possible in order to give a measure how good our classification is. For cross-validation we first split all M trials randomly 80/20 in train- and test data (CV all trials). And in a second variant we choose randomly the trials from 8 persons as test set and the trials from all other persons as training set, which is also an approximately 80/20-splitting (CV person-wise). In all cases we did the CV 10 times, resulting in 50 classification tasks in total, from which we average the classification rate. In general, we standardize the values of \mathbf{X} belonging to the test trials by a Z-transform, which transforms the mean and the variance of every column of \mathbf{X} to zero and one, respectively. This is necessary, since the coefficients are scaled differently. For the prediction of the trials in the test set we have to transform the values in \mathbf{X} belonging to the test trials by the same transformation like the training set.

We want to study the influence of the 7 different joint angles for the classification. Therefore we do the previously described classification using

only the joint angles in the subset $\mathbf{u} \subset \{1, \dots, 7\}$. This means, we only use the coefficients $a_k^{(i)}$ with $i \in \mathbf{u}$ for the classification. We denote the classification rate by $\text{cv}(\mathbf{u})$. The *Shapley values* are a common tool for describing feature contributions. They were introduced in [Shapley, 1952] for game theory and more recently also used for approximation theory, [Owen, 2014, Sundararajan and Najmi, 2020].

Our variables are the 7 different joint angles. Given any subset $\mathbf{u} \subset \{1, \dots, 7\}$ of the joint angles, the value that subset creates on its own is its explanatory power. We use here the classification rate as a way to measure explanatory power. Shapley showed that there is a unique valuation ϕ , that satisfies some reasonable axioms. Using our classification rate $\text{cv}(\mathbf{u})$, these values are defined by

$$\phi_i = \frac{1}{7} \sum_{\mathbf{u} \subseteq \{1, \dots, 7\} \setminus \{i\}} \binom{6}{|\mathbf{u}|}^{-1} (\text{cv}(\mathbf{u} \cup \{i\}) - \text{cv}(\mathbf{u})), \quad (1.3)$$

where $\text{cv}(\emptyset) := 0$. The values ϕ_i give some notion for importance of the joint angles for classification task.

1.3 Results

Our aim was to predict the object weight classification from the cosine coefficients $a_k^{(i)}$ of the time series belonging to the time series of the joint angles. Considering all joint angles, we achieved a classification rate of 0.683 (CV all trials) and 0.567 (CV person-wise). We plotted in Figure 1.4 some coefficients $a_k^{(i)}$ for $k = 0, 1$. One can see that there are some person specific behaviors, which means that some coefficients $a_k^{(i)}$ slightly differ for the persons, independent of the objects weight.

The mean classification rate which we reached for different subsets \mathbf{u} and the different CV strategies are summarized in Table 1.1. There we present the subsets \mathbf{u} , for which the highest classification rates are possible. The best classification rate 0.68 is reached involving all joint angles. Additionally, using for instance only one of the joint angles results in a low classification rate: Predicting the weight only from wrist ulnar/radial deviation gives prediction rate 0.354, since 3 different classes have to be predicted, this is no meaningful prediction.

One further question was the influence of the different joint angles to the classification rate. We calculated the Shapley values (1.3) for the two different CV strategies and plot the results in Figure 1.5. The wrist ulnar/radial deviation is least necessary for the classification. Whereas, the shoulder in-

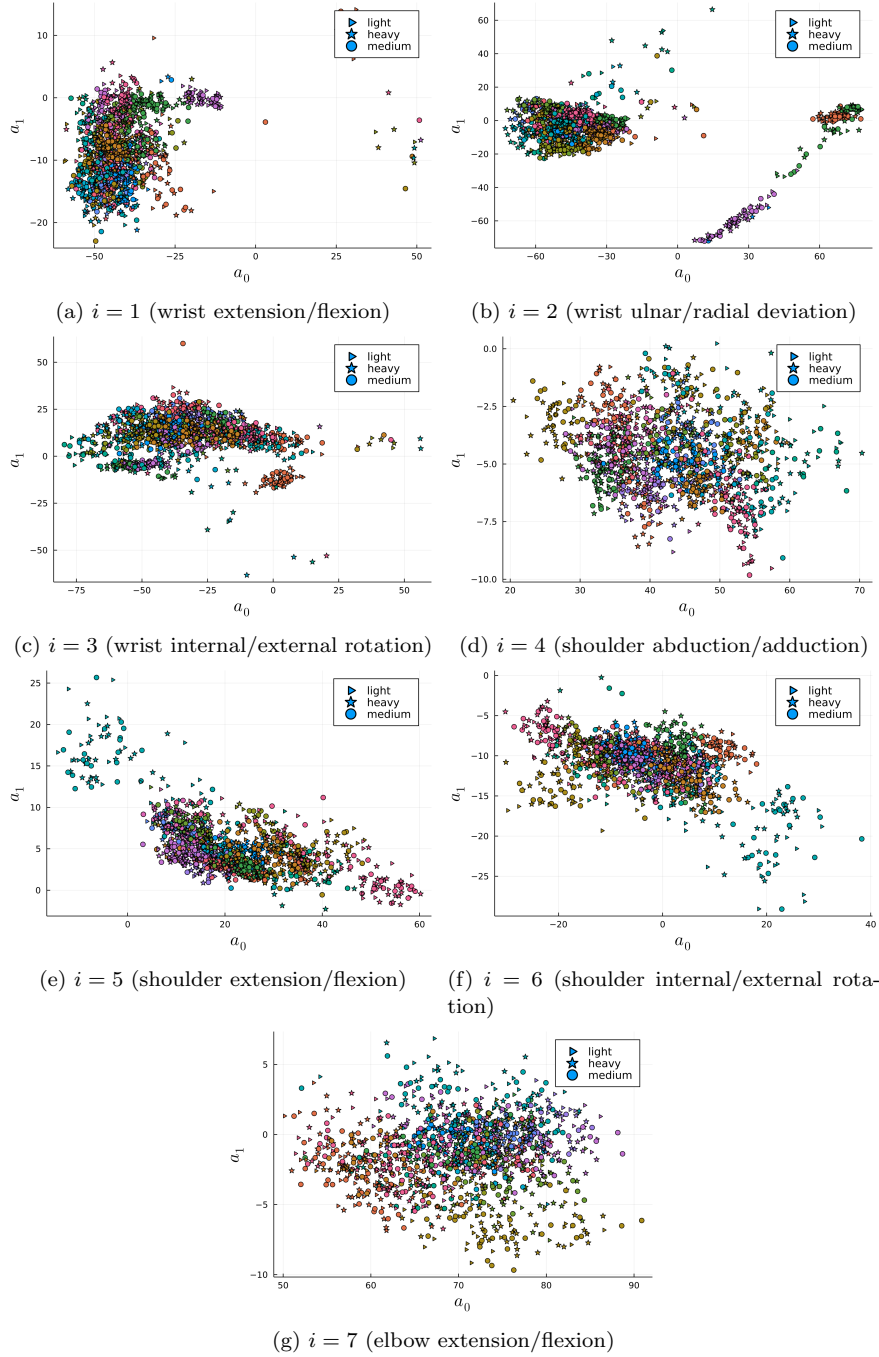


Fig. 1.4: The coefficients $a_k^{(i)}$ for $k = 0, 1$ for the different angles and all trials. The different colors belong to trials from different persons. The shape belongs to the object's weight.

ternal/external rotation and elbow extension/flexion are import angles for the classification.

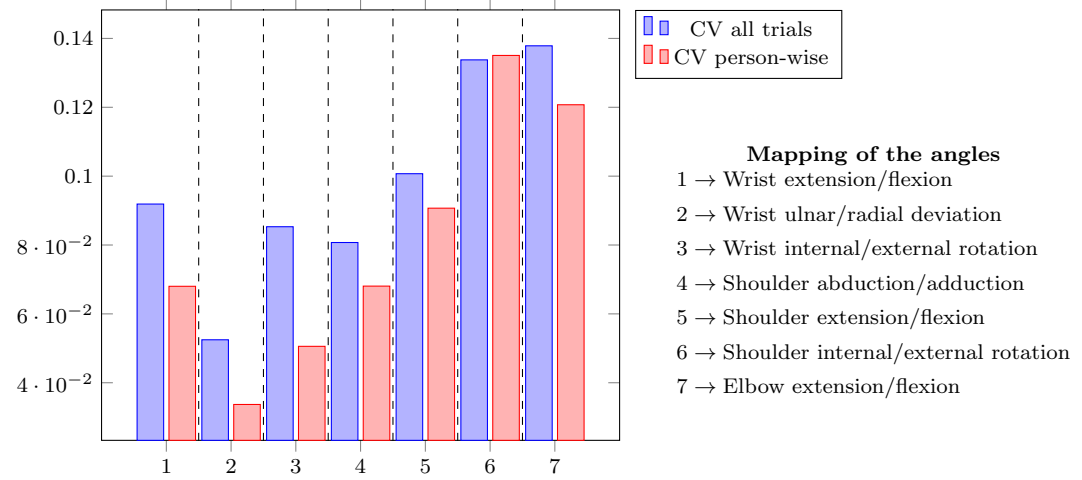


Fig. 1.5: The Shapley values with the classification rate as explanatory power and two different cross-validation strategies.

Table 1.1: Classification rates of predicting the objects weight from subset u of joint angles using different cross-validation strategies.

$u \setminus CV$	all trials	person-wise
{1, 2, 3, 4, 5, 6, 7}	0.683	0.567
{1, 3, 4, 6, 7}	0.680	0.578
{1, 3, 4, 5, 6, 7}	0.679	0.579
{3, 4, 5, 7}	0.678	0.578
{2, 4, 5, 7}	0.676	0.573
{2, 3, 4, 5, 6, 7}	0.676	0.603
{3, 4, 6, 7}	0.675	0.613
{1, 4, 5, 6}	0.675	0.575
{1, 2, 4, 5, 6, 7}	0.674	0.587
{4, 6, 7}	0.664	0.619
{4, 5, 6, 7}	0.670	0.614
{5, 6, 7}	0.659	0.606
⋮	⋮	⋮
{3}	0.482	0.421
{4}	0.447	0.437
{2}	0.391	0.354

Mapping of the angles
 1 → Wrist extension/flexion
 2 → Wrist ulnar/radial deviation
 3 → Wrist internal/external rotation
 4 → Shoulder abduction/adduction
 5 → Shoulder extension/flexion
 6 → Shoulder internal/external rotation
 7 → Elbow extension/flexion

1.4 Discussion

In this study we investigated how the weight of an object in a handover task can be predicted by the joint angles of the giver’s active arm. For this purpose, a discrete cosine transform and a support vector machine (SVM) were used to classify the different weights.

The results of the cross-validation (CV) show that a prediction of the object weight is possible, whereby a higher classification rate is achieved when the data of a subject is included in both the training and the test data set (CV all trials). We attribute this to the individuality of the movement. Looking at Figure 1.4, we notice that the data are clustered by person. This individual influence factor on movement has also already been shown in several studies [Bekemeier et al., 2019, Girges et al., 2015, Cunado et al., 1997, Bednarik et al., 2005]. The influence factor of individuality possibly also affects the classification performance. In other words, if an SVM is trained with the kinematic data of one person, a more reliable prediction of the kinematics of the same person is obtained than the prediction with kinematic data of another person.

Furthermore, by determining the Shapley value, it could be shown that especially the shoulder rotation and elbow movement is influenced by the object weight and therefore makes an important contribution to the prediction of the object weight. Wrist ulnar/radial deviation provides the least amount of explanation in object weight prediction.

It has already been shown that people have the ability to estimate the weight of an object by observation while another person is grasping, transporting or manipulating the object [Sciutti et al., 2014, Rizzolatti et al., 1999]. Efforts to find out on which kinematic characteristics these judgements depended yielded the result that mainly the duration of the lifting movement is used to make such judgements about the object weight [Hamilton et al., 2007]. The results of our study extend these findings and allow the use of kinematic data for weight prediction by using the joint angles alone without explicitly determining the lifting duration.

A limitation of our experimental setup is that we recorded the giver kinematics in a very controlled environment, as the start position of the object was always the same and the grasping position was only varied by the two different object sizes. This results in a relatively low variance of the giver movements. This contrasts with real handover actions in everyday life, where the object can always be at a different starting position and orientation, resulting in a greater kinematic variance of the giver.

Therefore, we suggest that in further experiments the starting position could be varied. Moreover, a variation of the weight classes would be interesting. In this study we investigated the three weight classes 400 g, 700 g, 1000 g, which could be varied both in distance and in range, possibly resulting in different classification accuracies. By changing the weight classes, it is conceivable that in the future not only classifications but also weight estimations will be pos-

sible through the analysis of kinematic data.

The SVM is widely used for classification tasks. Therefore, also here good results were expected. Furthermore, one has to choose specific parameters from the time-series for the classification, since the raw time series are noisy. And, furthermore, the joint angles at specific times are not robust against changes in the absolute position: The same movement from a slightly different starting position can lead to completely different joint angles.

In joint handover actions, observation and prediction (about the intentions of the other but also about the object properties) play an important role. This is why it is necessary that robots in hybrid societies are also able to achieve this, so that human actors can interact intuitively and smoothly. The approach described in this study can contribute to improving human-robot interactions in hybrid societies by having the robot predict object properties through observation.

Acknowledgment

This research has been supported by the German Research Foundation (DFG, 416228727) - SFB 1410 Hybrid Societies.

References

- [Aujeszky et al., 2019] Aujeszky, T., Korres, G., Eid, M., and Khorrani, F. (2019). Estimating Weight of Unknown Objects Using Active Thermography. *Robotics*, 8(4):92. Number: 4 Publisher: Multidisciplinary Digital Publishing Institute.
- [Bednarik et al., 2005] Bednarik, R., Kinnunen, T., Mihaila, A., and Fränti, P. (2005). Eye-Movements as a biometric. In *Proceedings of the 14th Scandinavian conference on Image Analysis, SCIA'05*, pages 780–789, Berlin, Heidelberg. Springer-Verlag.
- [Bekemeier et al., 2019] Bekemeier, H., Maycock, J., and Ritter, H. (2019). What Does a Hand-Over Tell?—Individuality of Short Motion Sequences. *Biomimetics*, 4(3).
- [Copot et al., 2016] Copot, D., Ionescu, C., Nascu, I., and De Keyser, R. (2016). Online weight estimation in a robotic gripper arm. In *2016 IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR)*, pages 1–6.
- [Cunado et al., 1997] Cunado, D., Nixon, M. S., and Carter, J. N. (1997). Using gait as a biometric, via phase-weighted magnitude spectra. In Bigün, J., Chollet, G., and Borgefors, G., editors, *Audio- and Video-based Biometric Person Authentication*, Lecture Notes in Computer Science, pages 93–102, Berlin, Heidelberg. Springer.
- [Girges et al., 2015] Girges, C., Spencer, J., and O’Brien, J. (2015). Categorizing identity from facial motion. *Quarterly Journal of Experimental Psychology*, 68(9):1832–1843.
- [Hamilton et al., 2007] Hamilton, A., Joyce, D. W., Flanagan, R., Frith, C. D., and Wolpert, D. M. (2007). Kinematic cues in perceptual weight judgement and their origins in box lifting. *Psychological research*, 71(1):13–21.
- [Hermsdörfer et al., 2011] Hermsdörfer, J., Li, Y., Randerath, J., Goldenberg, G., and Eidenmüller, S. (2011). Anticipatory scaling of grip forces when lifting objects of everyday life. *Experimental Brain Research*, 212(1):19–31.

- [Mason and Mackenzie, 2005] Mason, A. H. and Mackenzie, C. L. (2005). Grip forces when passing an object to a partner. *Experimental Brain Research*, 163(2):173–187.
- [Oldfield, 1971] Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9(1):97–113.
- [Owen, 2014] Owen, A. B. (2014). Sobol' indices and shapley value. *SIAM/ASA Journal on Uncertainty Quantification*, 2(1):245–251.
- [Plonka et al., 2018] Plonka, G., Potts, D., Steidl, G., and Tasche, M. (2018). *Numerical Fourier Analysis*. Applied and Numerical Harmonic Analysis. Birkhäuser.
- [Rizzolatti et al., 1999] Rizzolatti, G., Fadiga, L., Fogassi, L., and Gallese, V. (1999). Resonance behaviors and mirror neurons. *Archives Italiennes De Biologie*, 137(2-3):85–100.
- [Sciutti et al., 2014] Sciutti, A., Patane, L., Nori, F., and Sandini, G. (2014). Understanding Object Weight from Human and Humanoid Lifting Actions. *IEEE Transactions on Autonomous Mental Development*, 6(2):80–92.
- [Sebanz et al., 2006] Sebanz, N., Bekkering, H., and Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends in Cognitive Sciences*, 10(2):70–76.
- [Shapley, 1952] Shapley, L. (1952). A value for n-person games. *Technical report*. DTIC Document.
- [Standley et al., 2017] Standley, T., Chen, D., Sener, O., and Savarese, S. (2017). image2mass: Estimating the Mass of an Object from Its Image. page 10.
- [Sundararajan and Najmi, 2020] Sundararajan, M. and Najmi, A. (2020). The many shapley values for model explanation. In III, H. D. and Singh, A., editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 9269–9278. PMLR.
- [van Polanen and Davare, 2015] van Polanen, V. and Davare, M. (2015). Sensorimotor Memory Biases Weight Perception During Object Lifting. *Frontiers in Human Neuroscience*, 9.
- [Vicon, 2020] Vicon, M. S. (2020). Vicon Nexus Plug-in Gait Reference Guide.