"Elbows Out" - Predictive Tracking of Partially Occluded Pose for Robot-Assisted Dressing

Greg Chance, Aleksandar Jevtić, Praminda Caleb-Solly, Guillem Alenyà, *Member, IEEE*, Carme Torras, *Senior Member, IEEE*, Sanja Dogramadzi

***Abstract*—Robots that can assist in the Activities of Daily Living (ADL), such as dressing, may support older adults, addressing the needs of an aging population in the face of a growing shortage of care professionals. Using depth cameras during robot-assisted dressing can lead to occlusions and loss of user tracking which may result in unsafe trajectory planning or prevent the planning task proceeding altogether. For the dressing task of putting on a jacket, which is addressed in this work, tracking of the arm is lost when the user's hand enters the jacket which may lead to unsafe situations for the user and a poor interaction experience. Using motion tracking data, free from occlusions, gathered from a human-human interaction (HHI) study on an assisted dressing task, recurrent neural network models were built to predict the elbow position of a single arm based on other features of the user pose. The best features for predicting the elbow position were explored by using regression trees indicating the hips and shoulder as possible predictors. Engineered features were also created based on observations of real dressing scenarios and their effectiveness explored. Comparison between position and orientation based datasets was also included in this study. A 12-fold cross-validation was performed for each feature set and repeated 20 times to improve statistical power. Using position-based data the elbow position could be predicted with a 4.1cm error but adding engineered features reduced the error to 2.4cm. Adding orientation information to the data did not improve the accuracy and aggregating univariate response models failed to make significant improvements. The model was evaluated on Kinect data for a robot dressing task and although not without issues, demonstrates potential for this application. Although this has been demonstrated for jacket dressing, the technique could be applied to a number of different situations during occluded tracking.**

***Index Terms*—HRI, Safety, Motion Tracking, Pose Prediction, Deep Learning**

I. Introduction

ROBOTS that are capable of assisting humans in activities of daily living (ADL) may become a valuable resource in an aging population. The implications of an aging population are a lower proportion of people working and able to provide care for the demographic of those in need. Support with dressing for those with aging-related impairments, is a key area where assistive technology could help.

This work is done as part of the I-DRESS[1] project which aims to provide multi-modal interactive dressing assistance to users with limited upper-body mobility. In this paper we explore the problem of arm tracking when a robot equipped with a depth camera is assisting a person with jacket dressing, resulting in *garment-occlusion* and how this can be restored using predictive neural-network models. Garment-occlusion happens when the item of clothing is in close proximity to the user and the edges of the user become less well defined and tracking is lost, see Fig. 1. *Self-occlusion* can also occur when the line-of-sight of the depth camera is blocked by the user, usually from a body rotation or bringing a limb in front of or behind the torso or another limb. Self-occlusion may occur during robot-assisted dressing if the user turns away from the camera preventing line-of-sight of the camera to the tracked limb. *Robot-occlusion* typically occurs when the robot intersects the line-of-sight of the depth camera to the user, and is another important area for dressing but is not explored in this work. This is particularly important if the camera is located on the robot and the robot arms may move in front of the camera. If tracking of the arm is lost as the hand enters the jacket, the robot will not be able to plan the trajectory to the elbow, so this paper focuses on restoring the elbow marker.

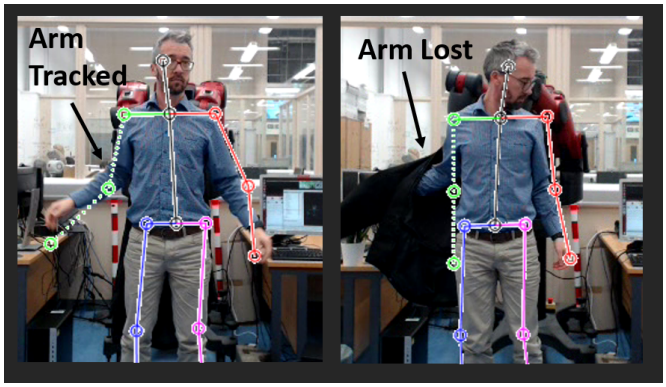

Fig. 1. Showing *garment-occlusion* where the tracking is lost when the hand enters the jacket.

The main contributions of this paper are the identification of features that can be used to predict the user's elbow position, a neural network topology that can be used for prediction and a method on how this might be achieved in a real dressing scenario with a Kinect camera.

II. Problem and Hypothesis

Tracking the user in any HRI task is necessary for trajectory planning, collision avoidance and ensuring user safety. During

G. Chance, P. Caleb-Solly & S. Dogramadzi are with the Bristol Robotics Laboratory, University of the West of England, Bristol, UK e-mail: (greg.chance@brl.ac.uk, praminda.caleb-solly@brl.ac.uk, sanja.dogramadzi@brl.ac.uk).

A. Jevtić, G. Alenya & Carme Torras are with the Institut de Robòtica i Informàtica Industrial, CSIC-UPC, Barcelona, Spain email: (ajevtic@iri.upc.edu, galenya@iri.upc.edu, torras@iri.upc.edu)

[1]https://i-dress-project.eu/

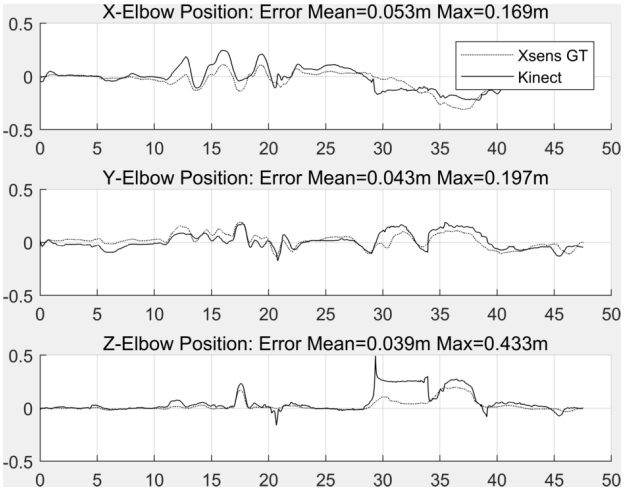

Fig. 2. Showing elbow tracking during a dressing task for Kinect and Xsens data sources simultaneously in the x, y and z axes (*top to bottom*).

a dressing task the user may be in close or physical contact with the robot leading to occlusions. The Kinect camera is widely used in the robotics community for user tracking but may suffer from occlusions in such ADL tasks.

This is the case when, for example, the hand of the user enters the sleeve during dressing of a jacket. A test to illustrate this problem was undertaken using both the Kinect camera and an Xsens motion tracking suit used for ground truth, where the joint markers from both data sources could be simultaneously compared. Fig. 2 shows the elbow position in a time series during a dressing task for all axes indicating a mean error in the Kinect data of 3.9-5.3cm with a maximum of around 43cm in the z-axis. The data shows that around time sample 30, occlusion of the user occurs and the Kinect data drifts away from the Xsens ground truth by a significant margin. It is at this point in the dressing sequence that restoration of the elbow marker would be especially beneficial.

A strategy for dealing with this issue could be using several cameras and using an algorithm to provide the skeleton data which has the highest probability [18]. Data persistence could be used to predict the elbow from the previous known coordinates but this may become unreliable especially if the user moves significantly in the period when tracking is lost.

A. The Elbow Marker

The important markers for jacket dressing are; hand, elbow and shoulder. During the initial stages of dressing the hand is an important target for the robot but is generally not occluded until the hand enters the jacket at which point the location of the hand is unimportant as the next target is the elbow. The movement of the hand once inside the jacket is relatively unconstrained as the jacket is only gripped at the opening of the sleeve and movement of the hand will be unconstrained by the robot. Any movement of the hand inside the jacket is therefore of little concern from a safety or comfort perspective. From experimentation it is observed that the shoulder is a relatively stable marker even when occlusion happens. Therefore the choice to restore the elbow marker is the motivation behind this work as it is the marker most affected by occlusion and will be most beneficial for trajectory planning.

In this paper we hypothesize that a data centric approach can be used to restore the elbow position by using other features of the user's pose. An assessment of the features important to predicting the elbow position was undertaken. Using these features, machine learning techniques were used to predict the elbow position and determine the accuracy to a set of validation data. Following this, a k-fold cross-validation method was used for determining generalization of the model. The model was then tested on real data in a robot-assisted dressing task using the Kinect camera.

III. Related Work

Typically, assistive robotics use some form of vision-based method for localization, user tacking and trajectory planning and any form of occlusion of the user can result in failure of this method. Existing techniques for tracking people in situations where occlusion can occur is well documented, Wang et al. [25] modified a sliding window approach to estimate where in a 2D image the occlusion is and implementing a part-based detection on the un-occluded area. Cucchiara et al. [6] use probabilistic functions and appearance models designed to track users in indoor situations for detecting people falling or lying motionless [7] claiming the system works well with self-occlusions and user shape changing. Considering the more specific case of skeletal tracking of people the issue of occlusions has been explored by using motion information implemented using optical flow between intensity images [22]. Although Schwarz et. al. [22] deal with restoration of skeletal markers during *self-occlusion*, for dressing related tasks *garment-occlusion* will be a bigger issue that is addressed in this paper.

The technologies associated with robotics implemented for assisted living are being investigated for the range of ADL including walking [13] and guidance [8], nursing [20], sit-to-stand [23][3], bathing [21], and feeding [2][28]. There are several groups pursuing research interests specifically in the area of dressing, some of whom deal especially with occlusions. Gao et al. show how adapting to user requirements during dressing for path optimization [11] and minimizing end effector force [27] could be useful in dressing especially for vulnerable users. The same group also proposed user modeling for path planning [12] using a top-view depth camera on a sleeveless jacket which minimizes the issues with garment-occlusion due to the lack of a garment sleeve. Erickson et al. [9] argue that vulnerabilities of the user could be inferred from physics-based simulations used to train predictive models, such as the Long Short-Term Memory (LSTM) network, on the end effector data (force, torque, velocity) which has the benefit of not suffering from occlusion issues if relying solely on these inputs. They validated this approach with a hospital gown dressing task on 10 participants [10]. Twardon et al. [24] also acknowledge the issues with occlusion and self-occlusion, proposing the use of a human model and the trajectories planned around a body-centric policy space, demonstrated

with anthropomorphic hands dressing a hat onto a mannequin head. Their planning algorithm was tested in various head orientations to vary the amount of self-occlusion and found reasonable results. Vision free, and hence occlusion free, dressing using a data driven approach to dressing a hospital gown has shown good results [15]. Although limited to a single axis HMM models that classify dressing have demonstrated $>98\%$ accuracy trained on time-series force data of the end effector. Many other scenarios may experience problems with occlusions in real operating conditions such as in t-shirt [17] or trouser [26] dressing tasks.

IV. Method

The dataset used in this work was collected during a dressing task [4] where 12 users were given assistance from another human posing as a robot, see Fig. 3. The users were wearing a motion tracking suit (Xsens) to interpret the position and orientation of 23 points on the body. In the task, the user was asked to put on a jacket with the help of the 'robot' several times. Each user put the coat on 3 times over 4 conditions; sitting and standing combined with normal mobility and restricted elbow movement, totaling 12 times per person. The height of the users varied in the range of 145-185cm. Each dressing task took approximately 40s and the motion capture system was recording data at 50Hz resulting in 320k time samples. The position and orientation were set to Euclidean and quaternion (7 channels per skeleton node) totaling 161 variables. The dataset size was therefore 50.5x10^6 samples. This dataset was a particularly useful reference as the motion tracking suit does not suffer from any type of occlusion.

A. Data Preparation & Notation

Data preparation was as follows: the training data was tagged with each user's unique identifier number so that individual dressing patterns could be observed. For each user, the position data for each marker were calculated referenced to the pelvis, removing any absolute offset issues associated with observed drift in the Xsens data. The elbow was removed from the training set and used as the response variable. In the data, each skeleton marker was given a unique identifier prefixed with **S** for sensor followed by the sensor number listed in Table I. The ID is suffixed with two further letters, either *p* for position (xyz) or *o* for quaternion orientation ($wxyz$).

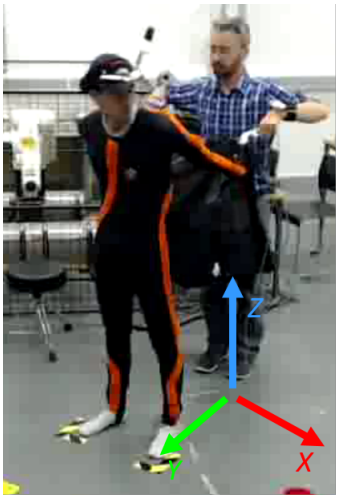

Fig. 3. Training data for this work was gathered from a human-human interaction study where 12 users were given assistance with putting on a jacket whilst wearing an Xsens motion tracking suit.

B. Feature Selection - Multivariate Response

Feature selection is a useful exercise to reduce large datasets to more manageable ones or simply to identify features in the data that best represent the underlying pattern one wishes to represent. For this jacket dressing scenario, the 161 features from the Xsens motion tracking suit would be difficult to attain online using a depth camera. Therefore a regression tree analysis was undertaken to find the pertinent features to this problem. Considering the application of this technology in dressing, we align the features of the training data to those that might be available at deployment. In this example the Microsoft Kinect camera (version 1) is considered in combination with a standard ROS interfaced skeleton tracker, e.g. openni_tracker[2]. The Xsens based training data has a 23 point skeletal representation compared to the Kinect's 15 point, and many of the markers are co-located and directly interchangeable apart from the pelvis which is assumed to lie at the midpoint between the Left and Right Hip markers. Table I gives a breakdown of all the Xsens markers and the Kinect counterpart where appropriate.

A regression tree model was used to find the dominant features for predicting the elbow position (multivariate response: xyz) using only the markers available to the Kinect. Kinect version 1 does not have orientation data available, so all orientation vectors were removed from the training data. The normalized feature importance from the regression tree analysis is shown in Fig. 4 for two cases; when the left hand is included as a feature (*upper graph*) and when it is excluded from the training data (*lower graph*). The mannequin icon

[2]http://wiki.ros.org/openni_tracker/

TABLE I
Skeleton Tracker Markers

Xsens ID	Xsens Description	Kinect Equivalent
1	Pelvis	none[a]
2-4	Spine	Torso
5	Sternum	none
6	Neck	Neck
7	Head	Head
12/8	Collar Bone (L/R)	none[b]
13/9	Shoulder (L/R)	Shoulder
14/10	Elbow (L/R)	Elbow
15/11	Hand (L/R)	Hand
20/16	Hip (L/R)	Hip
21/17	Knee (L/R)	Knee
22/18	Heel (L/R)	Foot
23/19	Toe (L/R)	none

[a] Can be taken from the mid point between hips.
[b] Neck will be slightly higher than collar bone.

(inset Fig. 4) highlights the dominant features with 5% or more importance.

The dominant features for prediction of the left elbow are the position vectors of the left hand x, y and z respectively. This is perhaps somewhat unsurprising given the proximity to the elbow and suggests that movement of the hand is usually linked with movement of the elbow. The right hip marker and left shoulder are in rank 4^{th} and 5^{th}. The hip-shoulder combination may be attributed to a twist or stretch in the torso that accompanies the elbow movement, changing the relative distance between these markers when the elbow is moved. In a real dressing task implemented using a depth camera the hand will likely be occluded and this data will not be available. However, the hand could be estimated from the position of the end effector at the time the occlusion first happened and even estimated as the end effector moves with limited accuracy.

In the second case we assume that no information about the hand is available, Fig. 4 *lower graph*, the right hip then becomes the dominant feature, followed by the left shoulder and the right hand in rank 4. The right hand is an interesting outcome to this analysis, and does actually appear in the 'with-hand' data above in rank 8. This suggests that the right hand, although only having ∼5% normalized importance, has some predictive power for the elbow position. The inclusion of the hand may come from using the other arm as a balance during dressing or possibly due to a repeated sequence that all users go through when dressing. This may be important to consider if tracking of both arms is lost simultaneously.

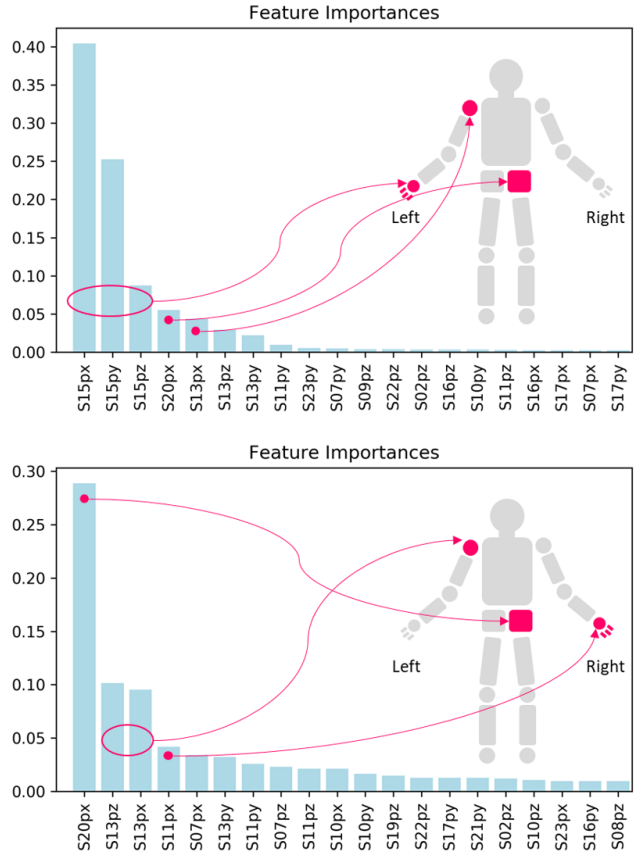

Fig. 4. Normalized feature importance for prediction of left elbow position (xyz), when considering the left hand as an input (*upper*) and when excluded (*lower*).

C. Features for Univariate Response - Single Axis Position

Of further interest might be the features that are important to predict a single position component of the elbow (univariate response) rather than for predicting all three axes simultaneously as above (multivariate response). The regression tree analysis was repeated for predicting the elbow position for x, y and z independently, i.e. a single output rather than three. For consistency the same restrictions around the inputs are maintained, no orientation data and hand and elbow removed. The univariate response is explored to understand the contribution and confidence interval to the elbow estimation for each position component and also to determine if a multivariate model can be trained to the same skill level as three univariate response models. There may also be advantages during deployment of the univariate approach, as missing data in a multivariate response model may lead to complete prediction failure whereas individual models may not suffer the same fate.

Table II shows the 5 most important features in rank order for predicting the elbow position, Φ, independently (univariate response) for x, y and z in columns 2-4 and for all axes (multivariate response) in column 5. For prediction of the elbow in the x-axis, $\Phi(x)$, the tree regression indicates that five features comprise 68% of the total normalized importance, coming from the shoulder, hip, hand and head. It is noted that these features are all in the top 5 for the multivariate response. The distribution of the left elbow x-position is very broad, see Fig. 5 *upper* overlapping with values for both y and z axes, which may explain why predictors of the elbow for $\Phi(x)$ are seen in $\Phi(xyz)$. For $\Phi(y)$, the head marker is dropped in favor of the right elbow and interestingly contains several markers from the x-axis and some markers not seen in either such as the right elbow. $\Phi(z)$ has the neck in the top 5 not being seen

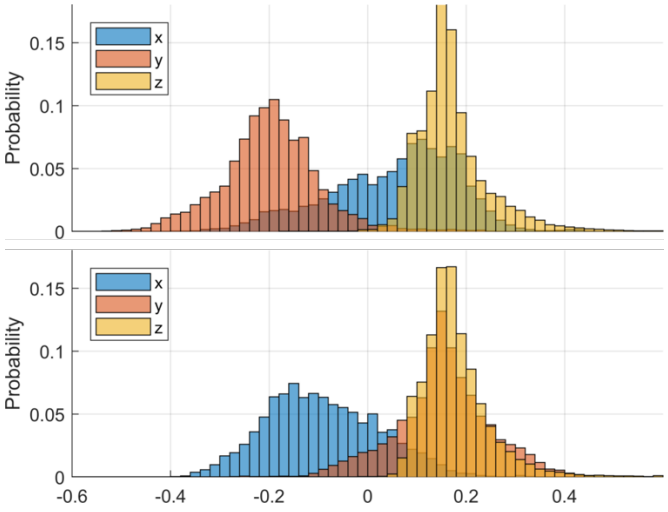

Fig. 5. Showing the normalized probability distribution of position (m) relative to the pelvis for the right (*upper*) and left (*lower*) elbow.

TABLE II
FEATURE IMPORTANCE GIVEN IN RANK ORDER FOR PREDICTION OF ELBOW POSITION (Φ) FOR UNIVARIATE AND MULTIVARIATE RESPONSE. $\Sigma(NI)$ INDICATES SUM OF NORMALIZED IMPORTANCE FOR THE TOP 5 FEATURES.

Rank	$\Phi(x)$	$\Phi(y)$	$\Phi(z)$	$\Phi(xyz)$
1	left shoulder ($13P_x$)	left shoulder ($13P_y$)	left shoulder ($13P_z$)	right hip ($20P_x$)
2	right hip ($20P_x$)	right hip ($20P_x$)	head ($7P_z$)	left shoulder ($13P_z$)
3	head ($7P_x$)	right elbow ($10P_y$)	right hand ($11P_z$)	left shoulder ($13P_x$)
4	right hand ($11P_x$)	right hand ($11P_x$)	head ($7P_x$)	right hand ($11P_x$)
5	left shoulder ($13P_y$)	right elbow ($10P_x$)	neck ($6P_z$)	head ($7P_x$)
$\Sigma(NI)$	0.683	0.684	0.723	0.570

in any other category so far.

D. Engineered Features

Following observations of the video footage of the assisted dressing experiment, a number of additional features were noted as possibly being relevant to key stages in the dressing task. The user's hand would often start low down, near the waist at the start of the task and finish quite high, near the head so the ratio of the hand position relative to the user's head was included. It was also noted that the user would often lean or twist their torso during dressing so the shoulder position relative to the head was also included. If orientation data is available then adding in the user's look direction (head orientation) is a valuable indicator of which arm of the jacket the user wishes to dress. Adding torso orientation would also contribute to the torso twist and lean mentioned above.

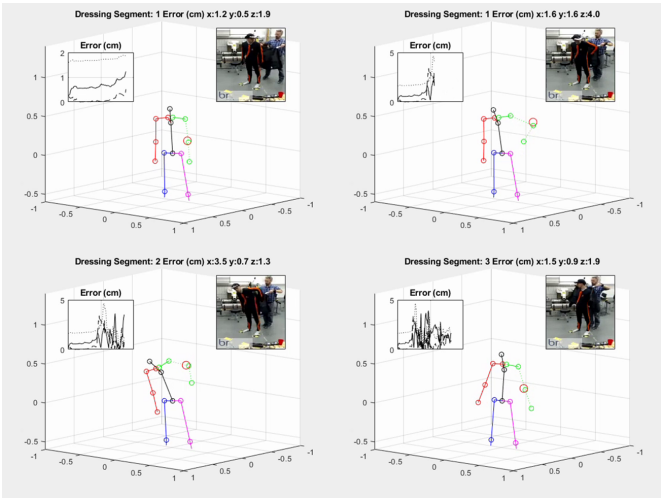

Fig. 6. Results of elbow tracking on unseen data.

E. Feature Sets

To easily assess the predictive power of the different features, they are divided into two main groups; *position based* data and *position and orientation* based data. This is done to align with the different technology levels in tracking hardware, i.e. Kinect I is position only but Kinect II has some orientation information. Within these two groups we have the option of adding the engineered features explained above and ensure that only position based engineered features are included in *position based* feature set. The feature sets are labeled: P - for position only features, PF - for position only with the engineered features added, PO - for position and orientation features, POF - position and orientation with engineered features added. P and PF features sets are suitable for the Kinect I with PO and POF for the Kinect II.

V. PREDICTIVE MODELLING

Using a Python environment in Jupyter Notebooks, Scikit-learn [19] was used alongside Keras [5] using TensorFlow [1] as a backend to train a Long Short-Term Memory (LSTM) network for predicting the left elbow position given the feature sets explored above. This network was chosen for its ability to work with time series data. A number of fully connected LSTM layers were fed into an output layer with 1 or 3 outputs depending on the response variable; 3 for the multivariate response $\Phi(xyz)$ and 1 for the univariate responses, $\Phi(x)$, etc. The standard rectified linear unit (relu) was used for activation of the hidden layers but linear activation for the output layer. The Adam optimizer was used with the mean-squared error loss function. Hyperparameter tuning was done via grid search for batch size, hidden layers, hidden units as well as L1/L2 and rnn_dropout regularization. For most cases the model parameters favored 2 LSTM layers with 10-200 hidden units with 50-60% dropout and a batch size of 5000 and a sample window of 4ms. The root mean square error (RMSE) was used as the main performance metric.

Using unseen data from the position and orientation feature set, the elbow position (multivariate) was estimated and overlaid on the original data for comparison, see Fig. 6. These results show good tracking of the elbow position, with an RMSE around $2cm$ (upper left inset in each graph) apart from when the elbow is at very acute angles where it approaches $5cm$. The test was repeated using the top 10 features predicted from the regression tree analysis and the difference in error was less than $0.5cm$ validating the feature importance method.

The error in the elbow position might be reduced further with a more complex network or more training data or it may be due to the particular unseen user data who may have had a particular style of dressing. This second hypothesis can be tested with cross-validation.

A. Cross-Validation

Cross-validation is a proven technique for assessing model performance over a given dataset. Kohavi argues for the use of 10-fold stratified (class balanced) cross-validation and claims no further improvement is necessarily made by going

beyond this value, depending on certain model constraints [16]. However, since the dataset used in this study consists of 12 individual users, there is an opportunity to observe the generalization of the model to different user's dressing patterns.

For each fold, one set of the user's data was reserved for testing, leaving approximately 92% for training a multivariate response model, $\Phi(xyz)$. In each case the training data was shuffled, the model was trained and the statistics on the RMSE are calculated for 20 repetitions per fold resulting in 240 training cycles in total per feature set, see Fig. 7. The results are quite reassuring, as regardless of the test set the RMSE value is relatively consistent, indicating no significant difference in error rate between users. This is with the exception of users 1 and 8 whose confidence intervals do not overlap but the resulting difference in error is approximately 2mm. This might suggest that people in our study all dressed in a similar fashion or that the network was able to learn these differences. The analysis indicates the RMSE in predicting the elbow position is on average 0.041m with a minimum of 0.038m and a maximum of 0.044m.

The cross-validation analysis was repeated for the remaining feature sets, again using 12-fold holdout with 20 repetitions per user, see Fig 8. The addition of the engineered features to the positional data (PF) significantly improves the error to an average of 0.024m compared to 0.041m when using the position only data (P). The position and orientation feature set (PO) indicates an average error of 0.028m, lower than P but not significantly better than PF. Adding the engineered features to the PO set resulted in an average error of 0.023m but not a statistically significant improvement over PF. This analysis would suggest that adding the engineered features to the position only feature set gives the same predictive power as including orientation information and possibly with greater consistency given the small confidence intervals.

It would appear that the addition of the orientation features does not improve prediction performance. However, adding the orientation data more than doubles the feature vectors which may require additional training and hyperparameter tuning. In addition to exploring the different feature sets, the effect of response type on RMSE was analyzed for the PF dataset using the same cross-validation technique. Fig. 9 shows the RMSE of the elbow position against the 4 response types; the multivariate case RMSE(xyz) 0.024m, and the univariate cases RMSE(x) 0.025m, RMSE(y) 0.027m and RMSE(z) 0.016m. This shows that the error in the y axis (0.027m) is higher than in the x (0.025m) axis and that the error in the z axis (0.016m) is much lower than both x and y. Overall the improvement in RMSE between the multivariate and univariate case is not statistically significant with a reduction in error of 1mm in favor of the aggregated univariate model.

B. Operation with Kinect Camera

To understand if the model could work with data from a Kinect camera in a robot dressing task, a small sample of data was collected. The Baxter robot was used to help dress a user with a rain jacket and the skeleton tracking data was recorded from the Kinect and Xsens motion tracker for a ground truth reference, see Fig. 10. For the Y and Z axes the maximum tracking error is significantly reduced. However for

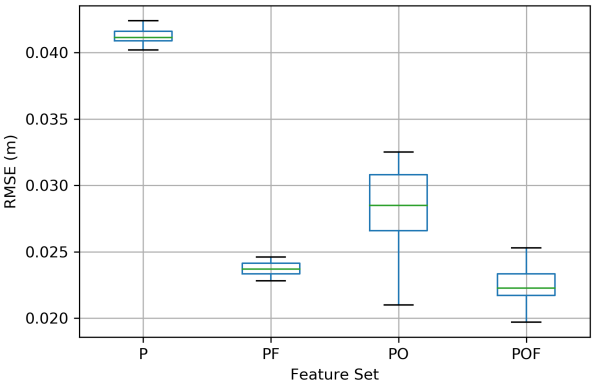

Fig. 8. RMSE of elbow position for the different feature sets: P Position only, PF position and engineered features, PO position and orientation and POF position, orientation and engineered features.

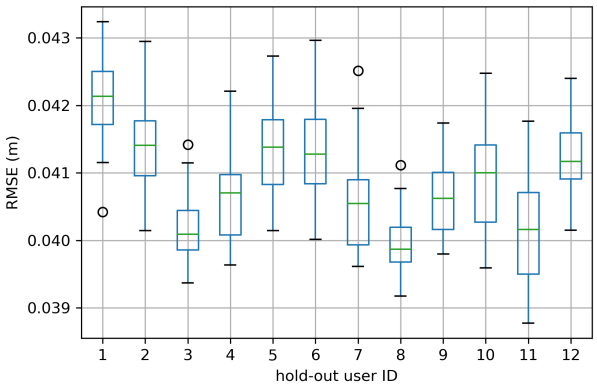

Fig. 7. RMSE of predicted elbow position using the PO (*position only*) feature set and *multivariate* response. Whiskers extend to 1.5 IQR.

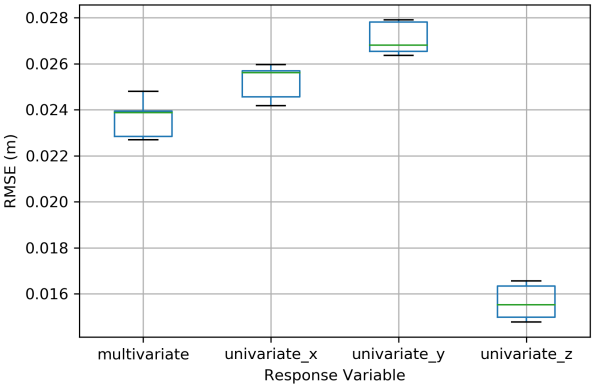

Fig. 9. RMSE of elbow position for multivariate and univariate responses for the PF feature set.

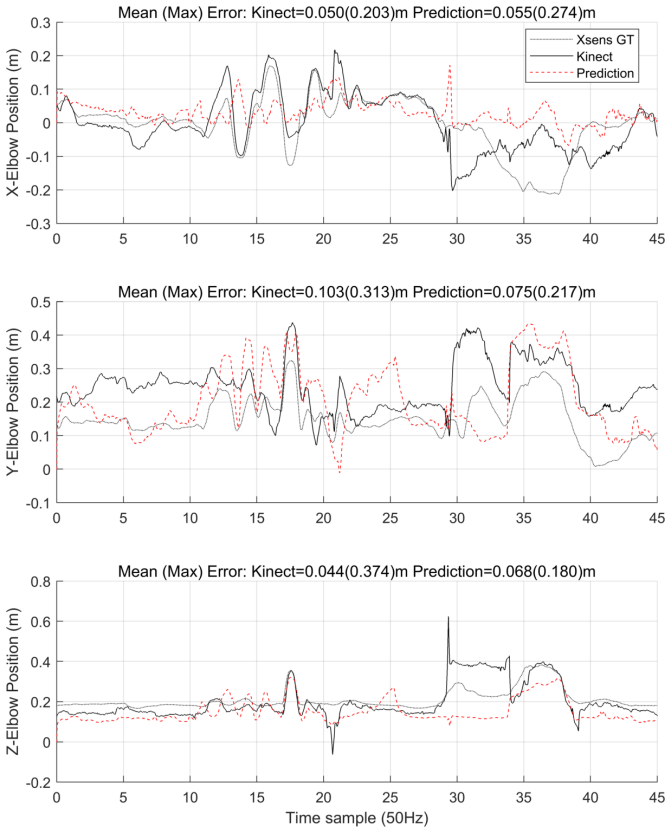

Fig. 10. Assessing model performance on Kinect data.

the X axis the prediction does not do as well which may have been due to calibration of the system prior to recording. The Kinect data also suffers from noise which could be improved with some filtering. The user's legs were static during the test but the Kinect markers were occasionally seen moving up to 10cm either side of the pelvis and this would impact on the prediction accuracy.

Overall, the results show good generalization for predicting the elbow position across different users (Fig. 7). Dependent on the type of garment being dressed, there may be a certain amount of free movement allowing a degree of error in the position. However, this margin of allowable error would diminish when fitting tighter clothing. In this case, adding additional inputs to the system, such as force or end effector position, could be investigated. In addition, no account of the user size and body shape was used for data scaling which could further improve accuracy. Also, implementing a biomechanical model of the arm to limit the solution space, see [14], and prevent unfeasible solutions would also improve prediction accuracy.

VI. Conclusion

The application of robotics to assist people for dressing may be valuable for a society with an aging population. Safe robot interaction requires tracking of the user which is often occluded during dressing and in our example of jacket dressing the elbow is of particular importance. Features to predict where the elbow is, using regression trees indicated a strong correlation with the shoulder, hip and opposite hand. A recurrent neural network model was used to make predictions of the elbow position and cross-validated across all 12 users and repeated 20 times to improve statistical power. Using position-based data the error in elbow position was 4.1cm. Adding engineered features to the data reduced the error to 2.4cm. Including orientation information to the data did not improve the elbow position accuracy. Aggregating the output of three univariate response models also failed to make significant improvements in prediction accuracy. Testing with a Kinect camera was partially successful at tracking the elbow, reducing the maximum error in the Y and Z axes. Overall this has been shown to be a valid method to deal with occlusion in our dataset and may be adapted to other tasks where occlusions occur.

Acknowledgment

This work was performed under the framework of the I-DRESS project (EU CHIST-ERA 2014), partially funded by the programme Acciones de Programacion Conjunta Internacional 2015 of the Spanish Ministry of Economy, Industry and Competitiveness (project ref. num. PCIN-2015-147) and the UK EPSRC (project ref. num. EP/N021703/1). We gratefully acknowledge the support of NVIDIA Corporation with the donation of a Titan Xp GPU used for this research.

References

[1] M. Abadi et al. "TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems". In: (Mar. 2016). arXiv: 1603.04467.

[2] G. Canal, G. Alenyà, and C. Torras. "Personalization Framework for Adaptive Robotic Feeding Assistance". In: *Social Robotics*. Ed. by A. Agah, J.-J. Cabibihan, A. M. Howard, M. A. Salichs, and H. He. Springer International Publishing, 2016, pp. 22–31.

[3] A. S. Chamnikar, G. Patil, M. Radmanesh, and M. Kumar. *Trajectory Generation for a Lower Limb Exoskeleton for Sit-to-Stand Transition Using a Genetic Algorithm*. 2017.

[4] G. Chance, P. Caleb-Solly, A. Jevtić, and S. Dogramadzi. "What's up? Resolving interaction ambiguity through non-visual cues for a robotic dressing assistant". In: *Robot and Human Interactive Communication (RO-MAN), 2017 26th IEEE International Symposium on*. IEEE. 2017, pp. 284–291.

[5] F. Chollet. *Keras*. 2015. URL: https://github.com/fchollet/keras.

[6] R. Cucchiara, C. Grana, G. Tardini, and R. Vezzani. "Probabilistic people tracking for occlusion handling". In: *Proceedings - International Conference on Pattern Recognition*. Vol. 1. 2004, pp. 132–135.

[7] R. Cucchiara, C. Grana, A. Prati, and R. Vezzani. "Computer vision techniques for PDA accessibility of in-house video surveillance". In: *First ACM SIGMM international workshop on Video surveillance*. ACM. 2003, pp. 87–97.

[8] S. Dubowsky et al. “PAMM - a robotic aid to the elderly for mobility assistance and monitoring: a ”helping-hand” for the elderly”. In: *Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings*. Vol. 1, pp. 570–576.

[9] Z. Erickson, A. Clegg, W. Yu, G. Turk, C. K. Liu, and C. C. Kemp. “What does the person feel? learning to infer applied forces during robot-assisted dressing”. In: *Robotics and Automation (ICRA), 2017 IEEE International Conference on*. IEEE. 2017, pp. 6058–6065.

[10] Z. Erickson, H. M. Clever, G. Turk, C. K. Liu, and C. C. Kemp. “Deep Haptic Model Predictive Control for Robot-Assisted Dressing”. In: (Sept. 2017). arXiv: 1709.09735.

[11] Y. Gao, H. J. Chang, and Y. Demiris. “Iterative path optimisation for personalised dressing assistance using vision and force information”. In: *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, Oct. 2016, pp. 4398–4403.

[12] Y. Gao, H. J. Chang, and Y. Demiris. “User modelling for personalised dressing assistance by humanoid robots”. In: *IEEE International Conference on Intelligent Robots and Systems*. Vol. 2015-Decem. 2015, pp. 1840–1845.

[13] B. Graf. “An Adaptive Guidance System for Robotic Walking Aids”. In: *Journal of Computing and Information Technology* 17.1 (2009), p. 109.

[14] Y. Jiang and C. K. Liu. “Data-Driven Approach to Simulating Realistic Human Joint Constraints”. In: (Sept. 2017). arXiv: 1709.08685.

[15] A. Kapusta, W. Yu, T. Bhattacharjee, C. K. Liu, G. Turk, and C. C. Kemp. “Data-driven haptic perception for robot-assisted dressing”. In: *Robot and Human Interactive Communication (RO-MAN), 2016 25th IEEE International Symposium on* (2016), pp. 451–458.

[16] R. Kohavi. “A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection”. In: IJCAI’95 (1995), pp. 1137–1143.

[17] T. Matsubara, D. Shinohara, and M. Kidode. “Reinforcement Learning of Motor Skills with Non-rigid Materials using Topology Coordinates”. In: *Advanced Robotics* 27.7 (2013), pp. 513–524.

[18] S. Moon, Y. Park, D. W. Ko, and I. H. Suh. “Multiple kinect sensor fusion for human skeleton tracking using Kalman filtering”. In: *International Journal of Advanced Robotic Systems* 13.2 (2016), p. 65.

[19] F. Pedregosa et al. “Scikit-learn: Machine Learning in Python”. In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[20] J. Pineau, M. Montemerlo, M. Pollack, N. Roy, and S. Thrun. “Towards robotic assistants in nursing homes: Challenges and results”. In: *Robotics and Autonomous Systems* 42 (2003), pp. 271–281.

[21] H. Satoh, T. Kawabata, and Y. Sankai. “Bathing care assistance with robot suit HAL”. In: *2009 IEEE International Conference on Robotics and Biomimetics, ROBIO 2009* (2009), pp. 498–503.

[22] L. A. Schwarz, A. Mkhitaryan, D. Mateus, and N. Navab. “Human skeleton tracking from depth data using geodesic distances and optical flow”. In: *Image and Vision Computing* 30.3 (2012), pp. 217–226.

[23] M. Shomin, J. Forlizzi, and R. Hollis. “Sit-to-stand assistance with a balancing mobile robot”. In: *Robotics and Automation (ICRA), 2015 IEEE International Conference on* (2015), pp. 3795–3800.

[24] L. Twardon and H. J. Ritter. “Learning to Put On a Knit Cap in a Head-centric Policy Space”. In: *IEEE Robotics and Automation Letters* (2018), pp. 1–1.

[25] X. Wang, T. X. Han, and S. Yan. “An HOG-LBP human detector with partial occlusion handling”. In: *2009 IEEE 12th International Conference on Computer Vision*. 2009, pp. 32–39.

[26] K. Yamazaki, R. Oya, K. Nagahama, and M. Inaba. “A method of state recognition of dressing clothes based on dynamic state matching”. In: *System Integration (SII), 2013 IEEE/SICE International Symposium on* (2013), pp. 406–411.

[27] F. Zhang, A. Cully, and Y. Demiris. “Personalized Robot-assisted Dressing using User Modeling in Latent Spaces”. In: *Proceedings of the International Conference on Intelligent Robots and Systems* October (2017), pp. 3603–3610.

[28] X. Zhang, X. Wang, B. Wang, T. Sugi, and M. Nakamura. “Real-time control strategy for EMG-drive meal assistance robot - My spoon”. In: *2008 International Conference on Control, Automation and Systems, ICCAS 2008* 1 (2008), pp. 800–803.