Simple and Robust Automatic Detection and Recognition of Human Movement Patterns in Tasks of Different Complexity

Lisa Gutzeit(1) Marc Otto(1) Elsa Andrea Kirchner(1,2)

(1)Robotics Research Group, University of Bremen,
Robert-Hooke-Str.1, 28359 Bremen, Germany
lisa.gutzeit@uni-bremen.de
http://www.informatik.uni-bremen.de/robotik
(2) Robotics Innovation Center, German Research Center for Artificial Intelligence (DFKI), Robert-Hooke-Str.1, 28359 Bremen, Germany

Abstract

In many different research areas it is important to understand human behavior, e.g., in robotic learning or human-computer interaction. To learn new robotic behavior from human demonstrations, human movements need to be recognized to select which sequences should be transferred to a robotic system and which are already available to the system and therefore do not need to be learned. In interaction tasks, the current state of a human can be used by the system to react to the human in an appropriate way. Thus, the behavior of the human needs to be analyzed. To apply the identification and recognition of human behavior in different applications, it is of high interest that the used methods work autonomously with minimum user interference. This paper focuses on the analysis of human manipulation behavior in tasks of different complexity while keeping manual efforts low. By identifying characteristic movement patterns in the movement, human behaviors are decomposed into elementary building blocks using a fully automatic segmentation algorithm. With a simple k-Nearest Neighbor classification these identified movement sequences are assigned to known movement classes. To evaluate the presented approach, pick-and-place, ball-throwing, and lever-pulling movements were recorded with a motion tracking system. It is shown that the proposed method outperforms the widely used Hidden Markov Model-based classification. Especially in case of a small number of labeled training examples, which considerably minimizes manual efforts, our approach still has a high accuracy. For simple lever-pulling movements already one training example per class sufficed to achieve a classification accuracy of above 95%.

1 Introduction

In the future, novel approaches in industry, production, personal services, health care, or medical applications, require a close collaboration of humans with robotic systems. To facilitate the requirements of these new approaches, not only the robotic systems must be equipped with enlarged mechanisms and skills that allow intuitive and safe interaction, but also the human intention, behavior and habits have to be better understood [12]. To allow this, novel methods to analyze human behavior are needed, which can easily be applied in different applications.

Understanding human behaviors is one important factor to successfully achieve intuitive human-computer interaction. For example, based on the knowledge of the current state of the human, systems can interact with humans in an appropriate manner. To obtain this knowledge, it is necessary to identify the representative parts of the human behavior and to assign the identified behaviors into categories which induce different reactions of the system. Only if the state of the human and the context which is described by this state are known, the system can follow the working steps that are required in this situation or can support the human if desired.

If robots become part of our everyday life in the future, it becomes important that also non-experts can teach a robotic system new skills. Robotic learning from demonstration is an active research area in robotics that promises to be a powerful tool to reach this goal, see for example [9, 14, 16, 17]. With learning from demonstration approaches, human demonstrations of a task can be transferred to a robotic system and generalized to solve different but similar tasks [9]. This allows also non-experts to demonstrate the system a way to solve a certain task without knowledge about robot control techniques. However, transferring a complex behavior to a system can be very time-consuming or even impossible. In order to learn also complex behaviors, the demonstration should be segmented into its main building blocks to be learned more efficiently [19]. By grouping segments that belong to the same behavior and by recognizing these behaviors, it can be determined which segments are needed to be learned for a certain situation. Beyond that, movements can be identified that can already be executed by the system and thus do not need to be learned.

Behavioral studies indicate that also humans learn complex behavior incrementally, as can be seen, e.g., in a study on infants [2]. The hypothesis is that complex behaviors are learned based on simple individual building blocks that are chunked together to a more complex behavior [8]. The idea in this work is to identify building blocks of human manipulation demonstrations so that they can be learned by the robotic system. In this way the system can learn a repertoire of behavior building blocks based on human demonstration which can easily be combined to different complex manipulation movements. To detect building blocks of human demonstrations, characteristic movement patterns have to be identified. In manipulation behaviors, bell-shaped velocity profiles have been found to be a suitable pattern [15]. In this work, a velocity-based behavior segmentation algorithm, introduced in previous work as velocity-based Multiple Change-point Inference (vMCI) [19], is used to segment recorded human manipulation movements. The applied algorithm detects movement sequences that show a bell-shaped velocity profile and are therefore assumed to be building blocks of human behavior. Furthermore, the vMCI algorithm identifies movement building blocks automatically without need for parameter tuning despite noise in the data [19].

The identified building blocks of human movements have also to be classified according to the actual behavior they belong to. By assigning suitable annotations to the recognized movement classes, the selection as well as the detection of the required behavior becomes intuitive and easy to use in different interaction scenarios. For supervised movement classification approaches the training data needs to be manually labeled. To keep the manual input low, it is desirable that the classification works with small sets of training data. We propose to classify detected building blocks by using simple k-Nearest Neighbor (k-NN) classification which satisfies this condition.

In this paper, our previous work presented in [10] is recapped and extended with an additional experiment and evaluation. Beside the application of our methods on pick-and-place and ball-throwing movements, the proposed methods are additionally applied to segment and recognize lever-pulling movements in a third experiment. The paper is organized as follows: In Section 2, different state-of-the-art approaches for segmentation and recognition of human movements are summarized. Our approach is described in Section 3. Afterwards in Section 4, the approach is evaluated on real human manipulation movements in tasks of different complexity. All results are compared to Hidden Markov Model (HMM)-based approaches which are widely used in the literature to represent and recognize movements. At the end of this paper, a conclusion is given.

2 Related Work

Depending on the modality to record human movements, there are a lot of different methods to recognize human behaviors. In many applications, human actions are recognized in videos, e.g., to find tackles in soccer games, to support elderly in their homes or for gesture recognition in video games [18]. Human action classification is just as important as detecting the human itself in video-based action recognition. Algorithms like Support Vector Machines, or their probabilistic variant the Relevance Vector Machines, Hidden Markov Models, k-Nearest Neighbors or Dynamic Time Warping-based classification are used to classify the observed actions. A more detailed overview is given in [18].

If the human behavior is not observed in a video but using motion tracking, e.g., with markers placed on the body, the segmentation of the recorded movements is next to the classification of high interest. For example in [5], human arm movements were tracked and segmented into so-called movement primitives at time points where the angular velocity of a certain number of degrees of freedom crosses zero. After a PCA-based dimensionality reduction, the identified movements were clustered using k-Means. Even though this approach promises to identify the primitive units of human movements, it requires the selection of thresholds to determine the segment borders. This is very sensitive to noise in the input data which results in over-segmentation of the data and requires adaption of the parameters for different applications. Gong et al., on the other hand, propose Kernelized Temporal Cut to segment full body motions, which is based on Hilbert space embedding of distributions [6]. In their work, different actions are recognized using Dynamic Manifold Warping as similarity measure. In contrast to the analysis of full body motions, we focus on the identification and recognition of manipulation movements which show special patterns in the velocity which should be considered for segmentation.

Beyond that, HMM-based approaches are often used in the literature, both for movement segmentation as well as for movement recognition. For example, Kulic et al. stochastically determine motion segments which are then represented using HMMs [13]. The derived segments are incrementally clustered using a tree structure and the Kullback-Leibler distance as segment distance measure. In a similar fashion, Gräve and Behnke represent probabilistically derived segments with HMMs, where segments that belong to the same movement are simultaneously classified into the same class if they can be represented by the same HMM [7]. Besides these approaches, solely training-based movement classification with HMMs is widely used, e.g. in [20, 1]. Because HMMs are expected to perform not well when few training data is available, we propose to use k-NN instead and compare it with the HMM approach.

3 Methods

This section describes a velocity-based movement segmentation algorithm which automatically identifies building blocks in human manipulation movements without the need of parameter tuning. In the second part of this section, an approach to recognize different known movement segments in an observed behavior is described.

3.1 Segmentation of Human Movement into Building Blocks

The purpose of this work it to find sequences in human manipulation movements that correspond to elementary building blocks which are characterized by bell-shaped velocity profiles as shown in [15]. Therefore, a segmentation algorithm is needed that identifies these building blocks. A second important property of the algorithm is the ability to handle variations in the movements. Human movement shows a lot of variations both during the execution by different persons as well as by the same person. For this reason, it is important that the algorithm for human movement segmentation finds sequences that correspond to the same behavior despite differences in their execution. Furthermore, the algorithms should be applicable to different tasks with low efforts. This can be accomplished with an algorithm that does not require parameter tuning if different types of movements are analyzed.

In previous work, we introduced the velocity-based Multiple Change-point Inference (vMCI) algorithm which tackles these issues [19]. The algorithm detects building blocks in human manipulation movements fully automatic. It is an probabilistic method, which can handle variations in the movement and the direct parameters of the data model are inferred from the data. It is based on the Multiple Change-point Inference (MCI) algorithm [4] in which segments are found in time series data using Bayesian Inference. Each segment $y_{i+1:j}$, starting at time point i and ending at j, is represented with a linear regression model (LRM) with q predefined basis functions ϕ_k:

$$y_{i+1:j} = \sum_{k=1}^{q} \beta_k \phi_k + \varepsilon, \tag{1}$$

where ε models the noise that is assumed in the data and $\beta = (\beta_1, ..., \beta_q)$ are the model parameters [19]. It is assumed that a new segment starts if the underlying LRM changes. This modeling of the observed data allows to handle technical noise in the data as well as variation in the execution of the same movement. To determine the segments online, the segmentation points are modeled via a Markov process in order that an online Viterbi algorithm can be used to determine their positions [4].

We expanded the MCI algorithm in our previous work to detect movement sequences that correspond to building blocks characterized by a bell-shaped velocity profile [19]. To accomplish this, the LRM of Equation 1 is split to model the velocity of the hand independent from its position with different basis functions, where the basis function for the velocity dimension is chosen in a way that it has a bell-shaped profile. In detail this means that the velocity y^v of the observed data sequence is modeled by

$$y^v = \alpha_1 \phi_v + \alpha_2 + \varepsilon, \tag{2}$$

with weights $\alpha = (\alpha_1, \alpha_2)$ and noise ε [19]. The model has two basis functions. First, the bell-shaped velocity curve is modeled using a single radial basis function [19]:

$$\phi_v(x_t) = \exp\left\{ -\frac{(c - x_t)^2}{r^2} \right\}. \tag{3}$$

If half of the segment length is chosen for the width parameter r, the basis function can cover the whole segment. The center c is determined automatically by the algorithm and regulates the alignment to velocity curves with peaks at different positions. Additionally, the basis function 1 weighted with α_2 accounts for velocities unequal to zero at start or end of the segment. As in the original MCI method, an online Viterbi algorithm can be used to detect the segment borders.

Figure 1 shows an exemplary result of the segmentation of artificial data using the vMCI algorithm. At the top, a one-dimensional simulated movement can be seen. The lower figure shows the corresponding velocity. To simulate two different behavior segments, the movement is slowed down at time point

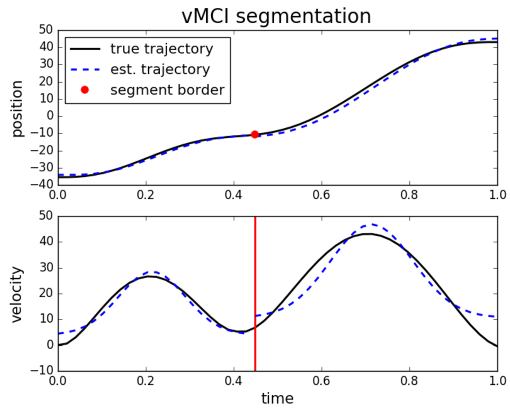


Figure 1: Artificial movement consisting of two sequences with a bell-shaped velocity. The vMCI segmentation successfully detected the transition point. Extracted from [10].

0.4. For the position dimension, the algorithm fits LRMs to the data according to Equation 1 with pre-defined basis functions. In this case, autoregressive basis functions are chosen. The velocity dimension is simultaneously fit with a LRM as introduced in Equation 2. The algorithm automatically selects the models which best fits parts of the data. In this case, it is most likely that the data arises from two different underlying models. This results in a single segmentation point which matches the true segmentation point within an acceptable margin. In contrast to other segmentation algorithms, for example a segmentation based on the detection of local minima, vMCI is very robust against noise in the data, as shown in [19]. Furthermore, the method is not sensitive to the choice of its hyper-parameters [19], hence, no parameter tuning is needed if it is applied to different data.

3.2 Recognition of Human Movement

There are many different possibilities to classify human movements, as reviewed in Section 2. The goal in this work is to choose a simple and robust classification method. To make the algorithm easily applicable on different manipulation data, minimal need for parameter tuning is of high interest. Furthermore, manual efforts can be minimized if the algorithm reliably classifies movement segments even if only a small training set is available. For this reasons we use a k-NN classifier for movement recognition. It has only one parameter, k, and is able to classify manipulation movements with a high accuracy given a small training set, as shown in our experiments.

3.2.1 Feature Extraction

Movement trajectories of markers placed at certain positions on the demonstrator are used in this work as features for the classification. The movements are recorded in Cartesian coordinates which results in different time series if the same movement is executed at a different position. Thus, the data is transposed into a coordinate system which is not global but relative to the human demonstrator. The position of the back is used as reference point (see Figure 2a) at the first time point of a segment, i.e. the data is transformed into a coordinate system centered at this point. Additionally, variances in the execution of the same movement are reduced by normalizing each movement segment to zero mean.

To successfully classify movement segments, additional features may be relevant. In manipulation movements where objects are involved, the positions of the objects as well as their spatial relation to the demonstrator are important features to distinguish between movement classes. Thus, the distance of the human hand to the manipulated object as well as the object speed are used in the pick-and-place experiment described in Section 4.2 to classify manipulation segments into distinct movements. Depending on the recognition task additional features, like the rotation of the hand to distinguish between different grasping positions, can be relevant.

3.2.2 Movement Classification

Due to its simplicity, we propose to use a k-NN classifier to distinguish between different movements. In the k-NN classification, an observed movement sequence is assigned to the movement class, which is the most common among its k closest neighbors of the training examples. To determine the closest neighbors, we use the standard Euclidean distance metric. All segments are interpolated to the mean segment length in order to account for segments of unequal length. Alternatively, dynamic time warping (DTW) could be used as a distance measure. However, in a preliminary analysis of k-NN classification on manipulation behaviors the presented approach outperformed a DTW-based k-NN. The number of neighbors k is set to 1. That means just the closest neighbor is considered for classification which leads to a good accuracy in case of small number of training examples. A bigger k could result in more classification errors due to the very low number of examples of each class.

4 Experiments

The proposed segmentation and classification methods are evaluated in this section on real human manipulation movements tracked using a motion capturing system. The experimental setup including the evaluation technique used in three different experiments are described in Section 4.1. Afterwards, the application and evaluation of the presented approaches on several demonstrations of three different manipulation movements are described. The evaluation on a

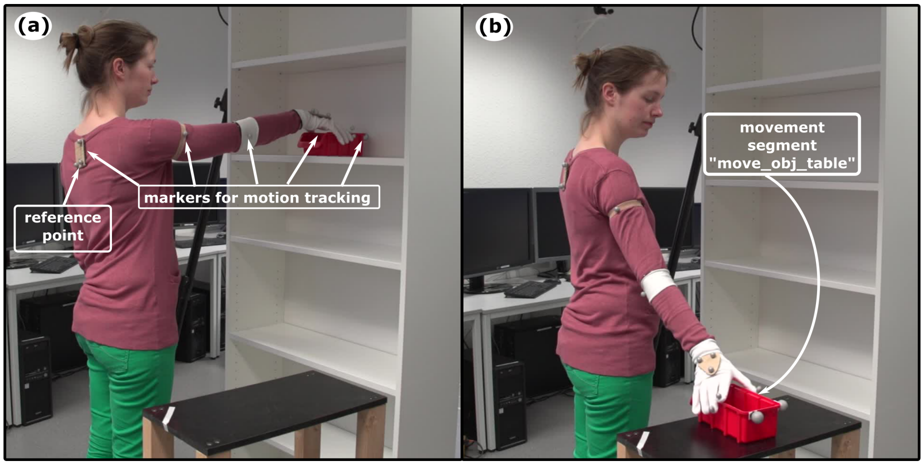


Figure 2: Snapshots of the pick-and-place task analyzed in this work. The images show the grasping of the object from the shelf (a) which is then placed on a table standing on the right hand side (b) which corresponds to the movement segment `move_obj_table`. Extracted from [10].

pick-and-place and ball-throwing tasks were already part of our previous publication [10] and are recapped here. Additionally, we evaluated the approaches on a lever-pulling task in Section 4.4. For all experiments it is shown that the vMCI algorithm correctly detects segments in the recorded demonstrations which correspond to behavior building blocks with a bell-shaped velocity pattern. Furthermore, the classification with k-NN using small number of training data is evaluated and compared to the results with an HMM-based classification.

4.1 Experimental Setup

The demonstrations of all manipulation movements were tracked using a marker-based motion tracking system. The 3D positions of visual markers placed on the subject were measured with 7 motion capture cameras at a frequency of $500\,Hz$. In a pre-processing step this data was down-sampled to $25\,Hz$. The positions of the markers can be seen in Figure 2 and Figure 3. Three markers were placed on the back of the demonstrator to determine the position of the back and its orientation. This was used to transform the recorded data into the coordinate system relative to the back, as described in Section 3.2. To track the movement of the manipulating arm, markers were placed at the shoulder, the elbow, and the back of the hand. The orientation of the hand was determined by placing three markers instead of one on it. Grasping movements in the pick-and-place demonstrations were recorded using additional markers which were placed at thumb, index, and middle finger. Furthermore, two more markers were placed on the manipulated object in this experiment to determine its position and orientation. However, the tasks in our experiments required

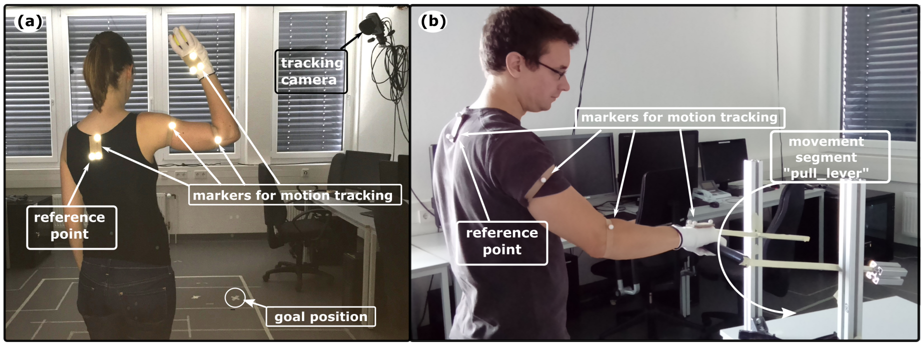


Figure 3: Snapshot of the ball-throwing (a), extracted from [10], and the lever-pulling task (b).

only basic manipulation movements, e.g., approaching the object or moving the object. Thus, just the position of the hand and the manipulated object were used for segmentation and recognition. However, the orientations are needed if the demonstrated movements should be transferred in a further step to a robotic system using learning from demonstration techniques [9].

Movement building blocks were identified in the demonstrations using the vMCI algorithm described in Section 3.1. The segmentation algorithm was applied on the position and the velocity of the recorded hand movements. As proposed in [19], the recorded positions of each demonstration were pre-processed to a zero mean and such that the variance of the first order differences of each dimension is equal to one.

To evaluate the proposed classification method, the resulting movement segments were manually labeled into one of the movement classes defined for each experiment. However, some of the obtained segments could not be assigned to one of these classes because they contain only parts of the movement. This could result from errors in the segmentation as well as from demonstrations where a movement was slowed down before the movement class ends. A case would be when the subject thought about the exact position to grasp the object. An example can be seen in the top plot of Figure 4. The concatenation of the first two detected segments belong to the class `approach_forward`. Nonetheless, the vMCI algorithm detected two segments, both with a bell-shaped velocity curve, because the subject slowed down the movement right before reaching the object. These incomplete movement segments were discarded for the evaluation of the classification approach. Furthermore, some of the identified movement segments did not belong to one of the pre-defined movement classes of the experiment. Usually, these nonassignable segments belonged to small extra movements, that were not part of the main movement task and thus were not considered in the defined movement classes. These movement segments were as well not used for the evaluation of the classification.

Before classification, the original recorded marker positions of each obtained segment were pre-processed as described in Section 3.2. Depending on the manipulation task, additional features were calculated. As proposed in Section 3.2, the obtained segments were classified using the 1-NN algorithm. For each of the two experiments, the accuracy of the 1-NN classification was evaluated using a stratified 2-fold cross-validation with a fixed number of examples per class in the training data. The training set sizes were varied from 1 example per class to 20 examples per class and the remaining data was used for testing. Since we want to show the performance of the classification with small training set sizes, the maximal number of training examples per class was kept low. For each number of examples per class in the training data, the cross-validation was performed with 100 iterations.

For comparison, the data was also classified using a HMM-based approach, which is a standard representation method for movements in the literature, see Section 2. In the HMM-based classification, one single HMM was trained for each movement class. To classify a test segment, the probability of the segment to be generated by each of the trained HMMs was calculated. The label of the most likely underlying HMM was assigned to the segment. The number of states in the HMMs was determined with a stratified 2-fold cross-validation repeated 50 times with equally sized training and test sets. As a result, we trained each HMM with one hidden state. The accuracy of the HMM-based classification with 1 hidden state per trained HMM was evaluated like the 1-NN classification with a stratified 2-fold cross-validation with fixed numbers of training examples for each class.

4.2 Segmentation and Recognition of Pick-and-Place Movements

In the first experiment, the presented approach was evaluated on pick-and-place movements. The task of the human demonstrator, partly shown in Figure 2, was to grasp a box from a shelf, move it to a table standing on the right side of the subject, and move the box back to the shelf. After placing the box on the table or the shelf, the subject should move the arm into a rest position in which it loosely hangs down. This task resulted in 6 different main movement classes: `approach_forward`, `move_obj_table`, `move_to_rest_right`, `approach_right`, `move_obj_shelf`, and `move_to_rest_down`. Short periods of time in which the demonstrator did not move his arm were assigned to the class `idle`.

The pick-and-place task was performed by three different subjects, repeated 6 times by each. Two of these subjects performed the task again with 4 repetitions while their movements were recorded with slightly different camera positions and a different global coordinate system. This resulted in different positions of the person and the manipulating object in the scene which should be handled by the presented movement segmentation and recognition methods. A total of 26 different demonstrations from different subjects and with varying coordinate systems were available to evaluate the proposed approaches.

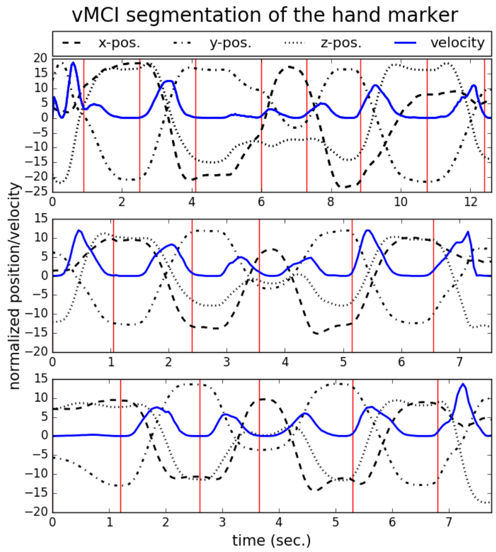


Figure 4: Segmentation results of three different demonstrations of the pick-and-place task. The x-, y- and z-position of the hand are visualized with black lines. The blue line corresponds to its velocity and the red vertical lines are the segment borders determined by the vMCI algorithm. Extracted from [10].

4.2.1 Results

The demonstrations of the pick-and-place task could be successfully segmented into movement parts with a bell-shaped velocity profile using the vMCI algorithm. Three examples of the segmentation results can be seen in Figure 4. The resulting movement segments were manually labeled into one of the 7 movement classes described above. This resulted in 155 labeled movement segments with different occurrences of each class, as summarized in Table 1.

Table 1: Occurences of each class in the recorded pick-and-place data. [10]

movement class	num. examples
`approach_forward`	20
`move_obj_table`	26
`move_to_rest_right`	25
`approach_right`	23
`move_obj_shelf`	26
`move_to_rest_down`	24
`idle`	11

As described in Section 4.1, next to the positions of the markers attached on the subject, the distance from the hand to the object and the object velocity were

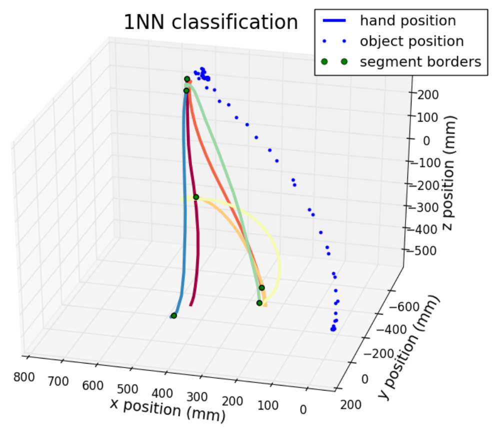


Figure 5: Classification result of a demonstration of the pick-and-place task with 1-NN. Different colors along the color spectrum starting with red for `approach_forward` and ending with blue for `move_to_rest_down` mark the different movement classes. Extracted from [10].

calculated as additional features in this experiment. An example result of the classification using 1-NN is shown in Figure 5. For this example demonstration of the pick-and-place task, all segments have been labeled with the correct annotation using a training set with 5 examples for each class.

The results of the cross-validation using 1-NN and HMM-based classification are shown in Figure 6. Because the data contains 7 different classes, an accuracy of 14.3% can be achieved by guessing. The 1-NN classification clearly outperforms the HMM-based classification using training sets with occurrences of each class smaller or equal to 20. Already with 1 example per class an accuracy of nearly 80% can be achieved using 1-NN. With 10 examples per class, the accuracy is 97.5% and with 20 examples per class 99.2%. In contrast, 14 examples per class are needed in the HMM-based classification to achieve an accuracy of 90% in this evaluation. With not more than 10 examples per class, the accuracy of the HMM-based classification is considerably below the achieved accuracy using 1-NN.

These results show that with the proposed 1-NN classification, manipulation movements can be assigned to known movement classes with a very small number of training examples. This means that with minimal need for manual training data labeling and no parameter tuning, very good classification results can be achieved using the proposed approach. Furthermore, the 1-NN classification considerably outperforms the widely used HMM-based classification in case that only a small number of training examples is available.

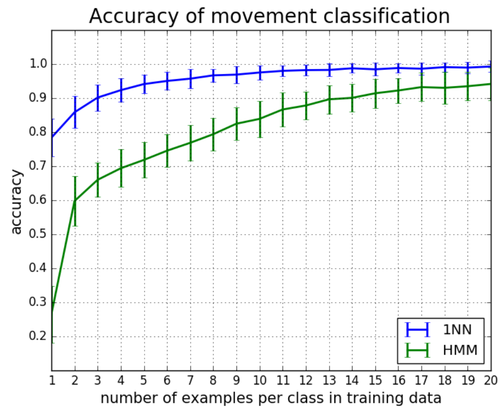


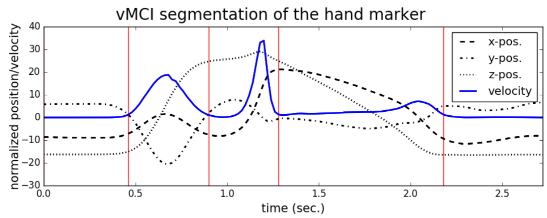
Figure 6: Comparison of the accuracy of the classification of pick-and-place movement segments using 1-NN and HMM-based classification. Extracted from [10].

4.3 Segmentation and Recognition of Ball-Throwing Movements

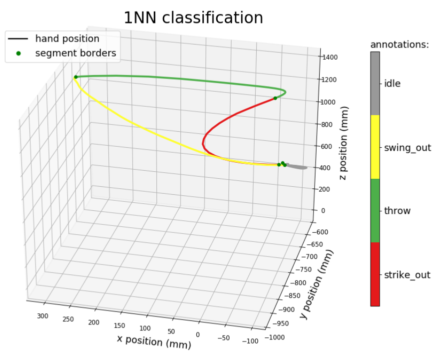
The vMCI segmentation and the 1-NN classification were evaluated in a second experiment on ball-throwing demonstrations. Compared to the pick-and-place experiment, this task is more challenging because no fixed objects are involved resulting in more possibilities of movement execution. The task of the subject was to throw a ball to a goal position on the ground located approximately 1.5 m away. The numerous possibilities to throw the ball were limited by the restriction that the ball should be thrown from above, i.e. the hand has a position higher than the shoulder before the ball leaves the hand, see Figure 3a. Nonetheless, the recorded throws show high variations in the demonstrations compared to the pick-and-place task. This could stem from different experiences in ball-throwing of the different subjects and training effects.

Before and after the throw, the subject had to move into a rest position, in which the arm loosely hangs down. The individual movement parts of each throw could be divided into four different main classes: `strike_out`, `throw`, `swing_out` and `idle`. In contrast to the pick-and-place task, only the movement of the arm was tracked in this task and not the position of the involved object, the ball. This is because in this experiment, the spatial distance of the ball to the demonstrator plays only a minor role and the movement of the arm has a much higher relevance to distinguish between movement classes. Furthermore, it was not recorded if the goal position was actually hit by the ball.

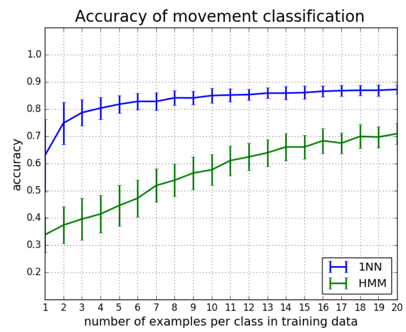
The ball-throwing task was demonstrated by 10 different subjects, each performing 24 throws.



(a)



(b)



(c)

Figure 7: Segmentation (a) and classification (b) result of one demonstration of the ball-throwing task. The presented methods successfully identified the segment borders and recognized the different movement classes. (c) Comparison of the accuracy of the classification of ball-throwing segments of all demonstrations using 1-NN and HMM-based classification. Extracted from [10].

4.3.1 Results

In a previous evaluation of the vMCI method on ball-throwing movements, it was already shown that the algorithm is able to identify the individual throws based on the position and velocity of the hand [19]. This result was confirmed by the evaluation of the demonstrations conducted for this work. A representative example of the segmentation result is shown in Figure 7a. Segment borders were correctly identified at positions were bell-shaped curves of the velocity profile end.

The resulting segments of all 240 ball-throw demonstration were manually assigned to one of the four movement classes to evaluate the classification. Again, each class had a different occurrence in the available data, as summarized in Table 2.

The positions of the markers attached to the subject, see Figure 3a, were used as features for the movement classification in this experiment. Figure 7b shows an example classification result using 1-NN and 5 examples per class in the training data. The 5 movement segments were correctly classified into one

Table 2: Occurences of each class in the ball-throwing data. [10]

movement class	num. examples
`strike_out`	221
`throw`	227
`swing_out`	339
`idle`	208

of the predefined classes.

The results of the cross-validation comparing 1-NN with HMM-based classification are visualized in Figure 7c. Like in the pick-and-place experiment, 1-NN outperforms HMM-based classification in the case of small training data sets. This experiment contains considerably more demonstrated movements and higher variance along demonstrations compared to the pick-and-place task. In here, the difference between classification algorithms is more contrasting. With one example per class in the training data, an accuracy of 62.9% using 1-NN can be achieved and only 33.8% by using HMM-based classification. This experiment contains 4 different classes, i.e. an accuracy of 25% can be achieved by guessing. Using 1-NN, a classification accuracy of 80% is accomplished using 4 examples per class during training. In contrast to this, this accuracy is not reached using HMM-based classification in this evaluation. For comparison, the evaluation was additionally conducted using 100 examples per class during training. This resulted in an accuracy of 91.5% using 1-NN, and 77.8% using HMM-based classification. This shows that even if more training data is available, the 1-NN classification outperforms the HMM-based approach.

4.4 Segmentation and Recognition of Lever-Pulling Movements

In a third experiment, the presented methods were evaluated on lever-pulling demonstrations. The task of the subject was to pull a lever down. The lever was fixed to a table and thus movement execution was in comparison to the other two experiments strongly predetermined, see Figure 3b. At the beginning of each demonstration the subject was in a rest position with the arm hanging down at the side of the body. Next, the subject reached for the lever with the right arm and pulled down the lever. Finally, the subject turned the arm back to the rest position (arm hanging down). After returning to the rest position, the lever had to be pulled up again, which was done with the left arm and had not been recorded by motion tracking. We chose this very simple behavior to show that for simple movements only very few demonstrations are needed for classification.

The individual movement parts of each movement could be divided into 4 different main classes: `idle`, `approach_forward`, `move_lever`, `move_to_rest`. As for the ball-throwing experiment, only the movement of the arm was tracked and not the position of the involved object, the lever. This is because in this experiment, the spatial distance of the lever to the demonstrators hand is fixed and plays no role. Only the movement of the arm can be used to distinguish between movement classes.

The lever-pulling task was demonstrated by two different subjects, performing 32 and 36 pulls, respectively.

4.4.1 Results

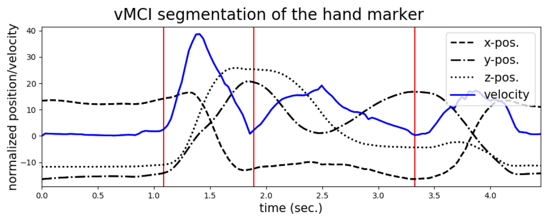
In this task the velocity of the movements did not always show smooth bell-shaped curves like in the previous experiments. This is because the positions of the hand was more predetermined. The subjects could not move their hand free, resulting in some cases in a slowed down movement with a more noisy velocity profile. This effect may be minimized by more demonstration trails generating a training effect on the subjects. Nonetheless, the vMCI algorithm successfully segmented the trajectories of the lever-pulling demonstrations without any adaptions of (hyper-)parameters or an additional preprocessing of the data. An example of the segmentation results can be seen in Figure 8a. The resulting movement segments were manually labeled into one of the 4 movement classes that are present in the lever-pulling task. The occurrences of each class can be found in Table 3.

Table 3: Occurences of each class in the lever-pulling data.

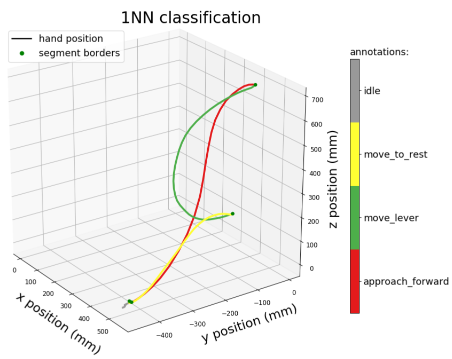
movement class	num. examples
`move_lever`	62
`approach_forward`	76
`move_to_rest`	72
`idle`	72

The positions of the markers attached to the subject, see Figure 3b, were used as features for the automatic movement classification. Figure 8b shows an example classification result using 1-NN and only one example per class in the training data. The 4 movement segments were correctly classified into one of the predefined classes.

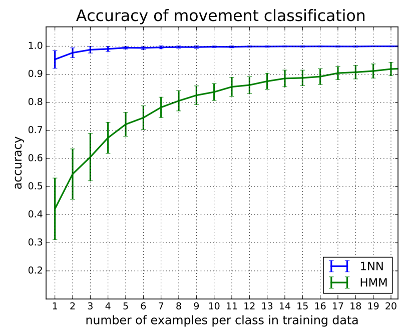
The results of the cross-validation comparing 1-NN with HMM-based classification are visualized in Figure 8c. As in the other two experiments, 1-NN outperforms HMM-based classification in the case of small training data sets. In this experiment, which contains simpler movements compared to the ball-throwing and the pick-and-place examples, the difference between the classification algorithms is very vivid in the area of very few training examples. Indeed, very high accuracy can already be achieved with one training example for each class, i.e., 95.3% using 1-NN but only 42.1% by using HMM-based classification. Using 1-NN, a classification accuracy of 99.0% is accomplished using 4 examples per class during training. In contrast to this, this accuracy is not reached using HMM-based classification in this evaluation.



(a)



(b)



(c)

Figure 8: Segmentation (a) and classification (b) result of one demonstration of the lever-pulling task. (c) Comparison of the accuracy of the classification of lever-pulling segments of all demonstrations using 1-NN and HMM-based classification.

This experiment shows that even if very few training data is available, training is possible resulting in a high accuracy using 1-NN classification in case that the movements are very simple.

5 Conclusions

We presented in this paper an approach to segment and classify human manipulation behavior. The segmentation was done using the unsupervised vMCI segmentation, formerly introduced in [19], which identifies building blocks of manipulation movements based on the velocity profile of the hand. For classification, we applied a simple 1-NN classifier using the Euclidean distance measure. Both algorithms were applied on pick-and-place, ball-throwing and lever-pulling movements. All these manipulation movements of different complexity could be successfully segmented and classified without the need of manual adaptions of the algorithms like, e.g., parameter tuning. Although a supervised classification method like 1-NN always needs manually labeled training data, we showed that the recognition of the movements can be done using a small set of training data, which considerably minimizes manual efforts. For the lever-pulling task, which is the simplest of the considered movements, a high classification accuracy could be achieved with just one training example per class. In comparison to widely used HMM-based movement classification, the accuracy was a considerably higher with small training sets in all experiments. Furthermore, the good classification results were achieved without any sophisticated feature selection methods.

For the development of embedded multimodal interfaces [11], simple approaches as the one presented here allow to use miniaturized processing units with relatively low processing power and energy consumption. This is, e.g., relevant in robotics, since the integration of interfaces into a robotic system is limited. But also wearable assisting devices have limitations regarding size, energy and computing power. For these applications not only accurate but also simple methods are needed. With the evaluation of our approaches we show that both, accuracy and simplicity, can be accomplished.

For future work, an integrated algorithm for segmentation and classification should be considered. Especially when extra segments are generated, e.g., caused by not fluently executed movements, an integrated algorithm where segmentation and classification influence each other becomes relevant. Such segments could be merged by identifying that only their concatenation can be assigned to one of the known movement classes.

In addition, manual effort needed for classification should be further minimized by recognizing the movement segments using an unsupervised approach. Annotations, like `move_object`, which are needed in many applications, e.g. to select segments that should be imitated by a robot, are ideally done without manual interference. These movement annotations can, e.g., be derived by analyzing features of the movement arising from different modalities. Psychological data, such as eye-tracking or electroencephalographic-data, could be used for this.

References

[1] Aarno, D., Kragic, D.: Motion intention recognition in robot assisted applications. Robotics and Autonomous Systems **56**, 692–705 (2008)

[2] Adi-Japha, E., Karni, A., Parnes, A., Loewenschuss, I., Vakil, E.: A shift in task routines during the learning of a motor skill: Group-averaged data may mask critical phases in the individuals' acquisition of skilled performance. Journal of Experimental Psychology: Learning, Memory, and Cognition **24**, 1544–1551 (2008)

[3] Argall, B.D., Chernova, S., Veloso, M., Browning, B.: A survey of robot learning from demonstration. Robotics and Autonomous Systems **57**(5), 469–483 (2009)

[4] Fearnhead, P., Liu, Z.: On-line inference for multiple change point models. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **69**, 589–605 (2007)

[5] Fod, A., Matrić, M., Jenkins, O.: Automated derivation of primitives for movement classification. Autonomous Robots **12**, 39–54 (2002)

[6] Gong, D., Medioni, G., Zhao, X.: Structured time series analysis for human action segmentation and recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence **36**(7), 1414–1427 (dec 2013). https://doi.org/4B458234-5E6D-453D-B6A0-C9F3A51683BB, http://www.ncbi.nlm.nih.gov/pubmed/24344075

[7] Gräve, K., Behnke, S.: Incremental action recognition and generalizing motion generation based on goal-directed features. In: International Conference on Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ. pp. 751–757 (2012), http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6386116

[8] Graybiel, A.: The basal ganglia and chunking of action repertoires. Neurobiology of Learning and Memory **70**, 119–136 (1998)

[9] Gutzeit, L., Fabisch, A., Otto, M., Metzen, J.H., Hansen, J., Kirchner, F., Kirchner, E.A.: The besman learning platform for automated robot skill learning. Frontiers in Robotics and AI **5**, 43 (2018). https://doi.org/10.3389/frobt.2018.00043, https://www.frontiersin.org/article/10.3389/frobt.2018.00043

[10] Gutzeit, L., Kirchner, E.A.: Automatic detection and recognition of human movement patterns in manipulation tasks. In: Proceedings of the 3rd International Conference on Physiological Computing Systems (2016)

[11] Kirchner, E.A., Fairclough, S., Kirchner, F.: Embedded multimodal interfaces in robotics: Applications, future trends and societal implications. In: Oviatt, S., Schuller, B., Cohen, P., Sonntag, D. (eds.) Handbook of Multimodal-Multisensor Interfaces, vol. 3, chap. IX, p. n.A. ACM Books, Morgan Claypool, forthcoming (2018)

[12] Kirchner, E.A., de Gea Fernandéz, J., Kampmann, P., Schröer, M., Metzen, J.H., Kirchner, F.: Intuitive Interaction with Robots - Technical Approaches and Challenges, pp. 224–248. Springer Verlag GmbH Heidelberg (2015)

[13] Kulić, D., Ott, C., Lee, D., Ishikawa, J., Nakamura, Y.: Incremental learning of full body motion primitives and their sequencing through human motion observation. The International Journal of Robotics Research **31**(3), 330–345 (2012)

[14] Metzen, J.H., Fabisch, A., Senger, L., Gea Fernández, J., Kirchner, E.A.: Towards learning of generic skills for robotic manipulation. KI - Künstliche Intelligenz **28**(1), 15–20 (2013). https://doi.org/10.1007/s13218-013-0280-1, http://dx.doi.org/10.1007/s13218-013-0280-1

[15] Morasso, P.: Spatial control of arm movements. Experimental Brain Research **42**, 223–227 (1981)

[16] Mülling, K., Kober, J., Koemer, O., J.Peters: Learning to select and generalize striking movements in robot table tennis. The International Journal of Robotics Research **32**, 263–279 (2013)

[17] Pastor, P., Hoffmann, H., Asfour, T., Schaal, S.: Learning and generalization of motor skills by learning from demonstration. In: 2009 IEEE International Conference on Robotics and Automation. pp. 763–768. Ieee (may 2009). https://doi.org/10.1109/ROBOT.2009.5152385, http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5152385

[18] Poppe, R.: A survey on vision-based human action recognition. Image and Vision Computing **28**(6), 976–990 (Jun 2010). https://doi.org/10.1016/j.imavis.2009.11.014, http://linkinghub.elsevier.com/retrieve/pii/S0262885609002704

[19] Senger, L., Schröer, M., Metzen, J.H., Kirchner, E.A.: Velocity-based multiple change-point inference for unsupervised segmentation of human movement behavior. In: Proccedings of the 22th International Conference on Pattern Recognition (ICPR2014). pp. 4564–4569 (2014). https://doi.org/10.1109/ICPR.2014.781

[20] Stefanov, N., Peer, A., Buss, M.: Online intention recognition in computer-assisted teleoperation systems. In: Haptics: Generating and Perceiving Tangible ensations, pp. 233–239. Springer Berlin Heidelberg (2010), http://link.springer.com/chapter/10.1007/978-3-642-14064-8_34