Honda Research Institute EU
Honda Research Institute Europe GmbH
https://www.honda-ri.de/

Learning of Object Manipulation Operations from Continuous Multimodal Input

Ulf Großekathöfer, Alexandra Barchunova, Thomas Hermann, Stefan Kopp, Robert Haschke, Mathias Franzius, Helge Ritter

2011

Preprint:

This is an accepted article published in Int. Conf. on Advanced Robotics (*IEEE-RAS International Conference on Humanoids) (ICAR). The final authenticated version is available online at: https://doi.org/[DOI not available]

Learning of Object Manipulation Operations from Continuous Multimodal Input

Ulf Großekathöfer* Alexandra Barchunova* Robert Haschke Thomas Hermann Mathias Franzius Helge Ritter
*Both authors contributed equally to this work.

***Abstract*—In this paper we propose an approach for identification of high-level object manipulation operations within a continuous multimodal time-series. We focus on multimodal approach for robust recognition of action primitive data. Our procedure combines an unsupervised Bayesian multimodal segmentation with a supervised machine learning approach. We briefly outline (1) the unsupervised segmentation and selection of uni- and bi-manual manipulation primitives developed in our previous work. We show (2) an application of the ordered means models to classification of estimated segments. To assess the performance of our approach, we compare the computed labels to the ground truth acquired during the data recording. In our experiments we examined the robustness of the procedure on two different sets of segments: full length ($\approx 95\%$ overlap with the ground truth on average), partial length ($\approx 10\%$ overlap with ground truth on average). We have achieved a cross validation rate of ≈ 0.95 and recognition accuracy of ≈ 0.97 for full length and ≈ 0.84 for partial length test sets.**

I. Introduction

An important objective of today's robotics research is to enable robots to interact and learn from humans. In order to participate in a simple interaction scenario, a robot needs the ability to autonomously single out relevant parts of the movement executed by a human. It also needs a mechanism of representation and identification of these parts.

Object manipulations constitute a large amount of human's interaction with the environment. In this paper we focus on multimodal identification of high-level uni- and bi-manual object manipulations in a continuous sequence. Our multimodal approach is inspired by the fact that during interactions different information sources (e.g. hearing, haptics and proprioception) are available to humans. The modalities of the recorded action sequences are: joint-angles of the hand and palm, force feedback of the fingers for both hands and the audio signal accompanying the object manipulation.

Learning and representation of action primitives and modeling of action sequences has been a popular research topic. Hidden Markov models [1] , finite state automata [2], stochastic context-free grammars [3], [4] and manifold learning [5] constitute a large part of the successfully used methods. A review of different techniques can be found in [6]. Successful approaches to sensor fusion for action and activity recognition have been showed in [7], [8], [9]. A recent overview of action recognition research can be found in [10].

However, in real-world applications, observations are incomplete or fragmented, sensor channels are noisy or completely missing. Thus, such scenario requires an approach with temporal and structural robustness that is able to provide a high degree of flexibility and applicability.

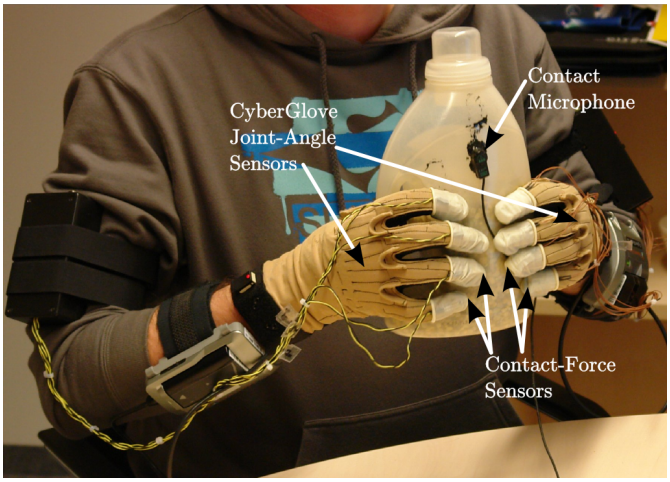

Fig. 1. Experimental setup: a human demonstrator wearing pressure and joint angle sensors performs manipulation operations with a (instrumented) plastic bottle provided with a contact microphone.

Our approach combines unsupervised Bayesian segmentation for multimodal time series [11], [12] and supervised learning with ordered means models (OMMs, [13]), a novel technique to machine learning of time series. An approach towards unsupervised segmentation of the multimodal time series of object manipulations has been developed in our previous work [11]. Our segmentation procedure applies Fearnhead's method for unsupervised changepoint detection for unknown number and location of changepoints [12]. Ordered means models (OMMs) have been successfully used to model multivariate sequential data [14], [15]. In this paper we introduce a two-staged learning procedure. It (1) estimates high-level segmentations and (2) learns OMM representations based on the results of (1). In the evaluation, we assess the performance of OMM-based classification and show the importance of multimodal approach to learning. We examine the following uni- and bi-manual operations: picking up, holding, shaking, putting down, screwing and unscrewing the cap, pouring.

The text is organized as follows: in Section II we describe the experimental setup and the recorded data; in Section III and IV we outline the segmentation method and the classification with OMMs; in Section V we present the data experiments followed by a discussion of the results in Section VI; we give a conclusion as well as a brief outlook of our future work in Section VII.

II. Scenario and Setup

In our scenario, a human demonstrator performs sequences of simple object manipulations with one and both hands. The object of these manipulations is a gravel-filled plastic bottle.[1] The instrumented bottle can be seen in Fig. 1. In the following we provide a brief overview of the data acquisition previously described in [11].

We use the following sensors to record multimodal time-series of the performed action sequences:

- 1 contact microphone attached to the bottle. The contact microphone focuses on in-object generated sound, ignoring most environmental noise.
- 2×24 joint-angles calculated from the measurements of two Immersion Cyberglove devices describe the individual postures of both hands.
- 2×5 FSR pressure sensors attached to the fingertips of each CyberGlove record the contact forces.

This collection of sensors yields a 29-dimensional (24 + 5) representation for each hand in addition to a scalar audio signal. The human demonstrator was told to perform a sequence of basic manipulation actions in the fixed order showed in the following enumeration. To achieve a rich variance of timing between trials, we added Gaussian noise to the nominal duration of the action primitives as specified in parentheses:

1) pick up the bottle with both hands (2 s + η_1)
2) shake the bottle with both hands (.7 s + η_2)
3) put down the bottle (1 s)
4) pause (1 s)
5) unscrew the cap with both hands (1.2 s + η_3)
6) pause (1 s)
7) pick up the bottle with right hand (2 s + η_4)
8) pour with right hand (1 s + η_5 + 1 s + η_5)
9) put down the bottle (1 s)
10) fasten the cap with both hands (1.2 s + η_6)

The random variables $\eta_i \sim \mathcal{N}(0, .5\ s)$ denote the randomized timing of subsequences. The overall length of the time series of a trial accumulates to approximately 30 seconds.

III. Segmentation

The recorded time series of multiple sensor modalities capture complex and high-dimensional descriptions of action sequences. Based on the tactile and audio modalities, our two-stage segmentation identifies and selects relevant multimodal low-level data. In the following paragraphs we briefly outline the two-stage procedure developed in our previous work [11].

The tactile modality is used to obtain a preliminary rough split of the sequence into sub-sequences of "object interaction" and "no object interaction". Sub-sequences that have been recognized as "object interaction" are further analyzed in detail w.r.t. qualitative changes of the **audio** signal in order to refine the rough segmentation. In this subordinate segmentation step, all "object interaction" segments produced in the previous step are sub-segmented. The sub-segmentation is formed by selecting segments that exhibit homogeneous oscillatory properties within the audio modality.

In both stages, the segmentation is performed by applying Fearnhead's method for unsupervised detection of multiple changepoints in time series [12]. Within this probabilistic framework, the optimal segmentation is obtained by maximizing the posterior distribution of the number and location of the changepoints. The segmentation is controlled by a set of local models and a prior for distribution of segment lengths. The estimated changepoints are optimal in the sense that the probability of all sub-sequences to originate from applied data models is maximized.

The resulting segmentation is characterized by constant contact topology in respect to overall hand activity as well as homogeneous characteristics of the audio signal. The outlined approach is robust against noise and delivers a segmentation that has a high degree of temporal and structural accuracy.

Fig. 2 illustrates segmentation determined on the basis of tactile modality. Contact assignments identify parts of the time series that are directly associated with object interactions. The sub-segmentation using the audio channel is showed in the Fig. 3. In this figure the segments "shake", "hold" and "put down" (2,3 and 4 resp.) as well as "grasp", "pour", "hold" (9, 10, 11 resp.) are generated from the "interaction segments" 1 and 5 (see Fig. 2) by subordinate segmentation of the audio modality. In this case six segments have been generated for six semantic descriptions of action primitives. Here, the overlap with ground truth is large. In the other case the recorded audio data provides for finer segmentation within one semantic category. Two segments (6 and 7) have been generated for one semantic description of the action primitive "unscrew". This is due to the change of acoustic signal accompanying the grasping of the bottle lid, being part of the action execution for "unscrew". Fig. 3 shows the corresponding peek in the level of the audio signal within the segment 6. For segment 6 the overlap with ground truth is poor.

Fig. 4 presents a histogram of the generated action primitives i.r.w. their overlap with the ground truth. The right side of each histogram shows the number of generated segments that have a high overlap (e.g. "shake"). These action primitives correspond the first case described above. The left side of each histogram presents the number of generated segments having a low overlap with the ground truth (second case).

As described above, the multimodal data representing an action primitive may contain more than one semantic substructure. Due to this fact, for a semantic action primitive description we receive segments with different degree of ground truth overlap. In order to realize learning of semantic categories ("hold", "shake", "pour" etc.), we use a model that is able to robustly represent partial multimodal sequential data.

IV. Ordered Means Models

In order to classify the segments, we use a specialized generative model, which we refer to as ordered means models (OMMs). To use generative models for classification of

[1] The use of gravel instead of liquid is due the necessity of a distinct audio signal and also safety concerns.

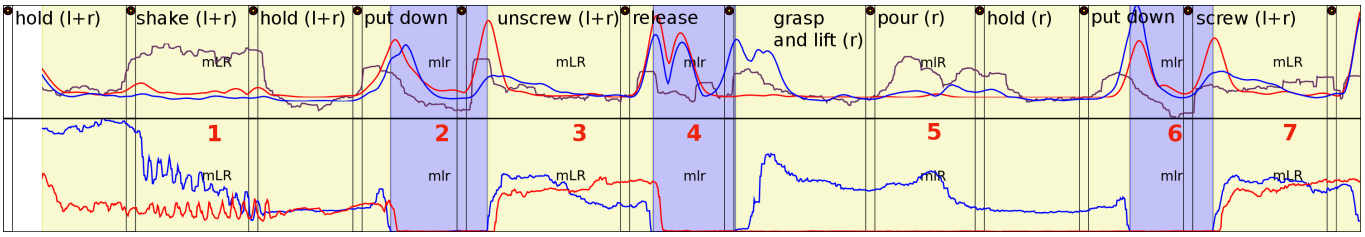

Fig. 2. Initial segmentation of a multimodal time series. First row: preprocessed joint-angle trajectories for both hands and the audio signal; second row: tactile signal for both hands. The black frames indicate the ground truth. The segmentation is estimated by applying Fearnhead's method to the tactile data of both hands.

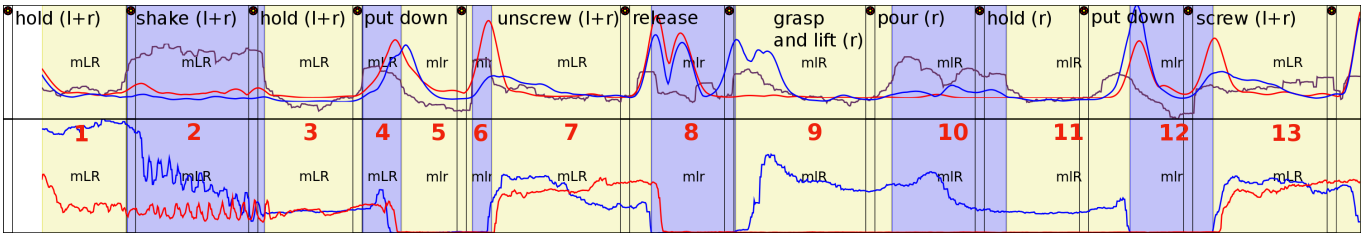

Fig. 3. Sub-segmentation for a multimodal time series. First row: preprocessed joint-angle trajectories for both hands and the audio signal; second row: tactile signal for both hands. The black frames indicate the ground truth. The segmentation is a refinement of the segmentation in the Fig 2, it is estimated by applying Fearnhead's method to the audio signal.

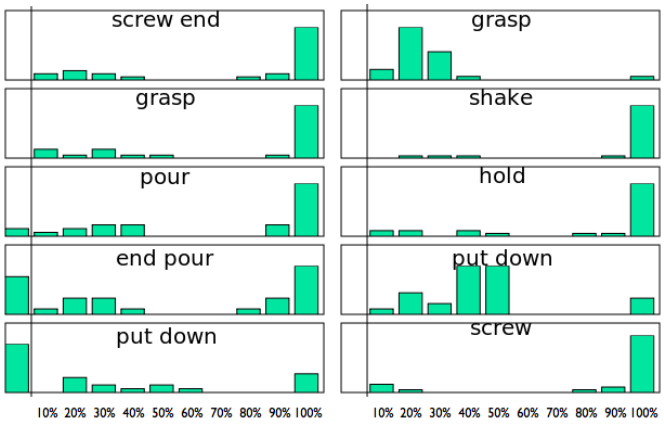

Fig. 4. Histogram of the segment distribution w.r.t. their overlap with ground truth. The leftmost column shows the number of undetected segments per action primitive.

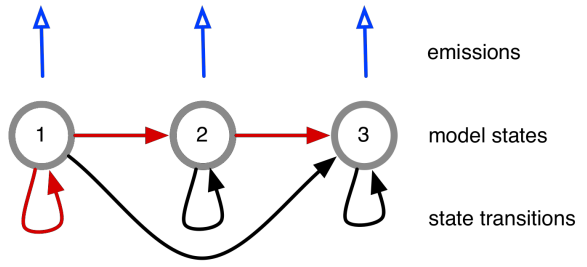

Fig. 5. This figure illustrates the design of an ordered means model with three states. The grey circles represent the model states, the blue arrows represent the emissions from the state, and black and red arrows emphasize the allowed state transitions. The red arrows represents exemplarily one particular path: the state sequence $1, 1, 2, 3$.

unseen time series, one firstly estimate a model for each class by means of labeled examples, i.e. examples that are assigned to a particular class. An unseen example is then classified to the class which model is most likely responsible for generating the example in question.

In case of the generated segment data, especially robust modeling in terms of incomplete and missing data is required. Even though approaches as hidden Markov models reach excellent results for complete data, they might not be the optimal choice for scenarios with unexpected gaps or time series with missing beginnings or endings. In particular, HMMs' implicit modeling of segments length distributions in terms of transition probabilities could lead to an inadequate representation for missing or fragmented data. Here, as a major difference in the overall model design, OMMs do not incorporate any transition probabilities; instead, all paths, i.e. all valid sequences of model states, are equally likely. This yields a model structure that also allows unlikely paths, which, e.g., correspond to time series with unexpected gaps.

In general, ordered means models (OMMs) are, similar to HMMs, generative state space models that emit a sequence of observation vectors $O = [\mathbf{o}_1, \ldots, \mathbf{o}_T]$ out of K hidden states. Figure 5 illustrates a OMM with three states. OMMs incorporate a number of design decisions:

- **Path probabilities:** In OMMs each path, i.e. each valid sequence of states, is equally likely (e.g. the violet marked path in Fig. 5. Note that such a design differs fundamentally from modeling transition probabilities in HMMs. There is no equivalent realization in terms of transition probabilities.
- **Emission densities:** The emissions of each state are modeled as probability distributions $b_k(\cdot)$ and are assumed to

be Gaussian with $b_k(\mathbf{o}_t) = \mathcal{N}(\mathbf{o}_t; \boldsymbol{\mu}_l, \sigma)$. The standard deviation parameter σ is identical for all states and is used as a *global hyperparameter*.

- **Model topology:** OMMs only allow transitions to states with equal or higher indices as compared to the current state, i.e. the network of model states follows a left-to-right topology (cf. 5).
- **Length distribution:** In principle, OMMs require the definition of an explicit length distribution either by domain knowledge or by estimation from the observed lengths in the training data. This, however, may not be possible due to missing knowledge or non-representative lengths of the observations. To circumvent the definition and estimation of a length model we assume a flat distribution in terms of an improper prior according to equally probable lengths.

With regard to these design decisions, an OMM Ω is completely defined by an ordered sequence of reference vectors $\Omega = [\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K]$, i.e. the expectation values of emission densities.

A. Parameter estimation

In order to estimate particular model parameters $\Omega = [\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K]$ by a set of observations $\mathbf{O} = \{O_1, .., O_N\}$ we maximize the log-likelihood $\mathcal{L} = \sum_{i=1}^{N} \ln p(O_i|\Omega)$ with respect to the mean vectors $\boldsymbol{\mu}_k$. To solve this optimization problem, we use an iterative expectation maximization algorithm, similar to the well-known Baum-Welch algorithm from HMMs but without transition probabilities.

V. Experiments

In our evaluation we operate on a data pool containing 30 sequences of object manipulations captured as described in Sec. II. This data was recorded with one human demonstrator during one session. In principle, the structure of all these trials should be identical except the execution timing. During data recording we also acquire ground truth by requesting the subject to start or end subsequences at signalled points in time (cf. [11]). The recorded ground truth timestamps mark the start or the end of the action primitives within the time series. We then selected and labeled the automatically generated segments by means of the recorded ground truth. This set of action primitive segments constitute our data pool for training and testing.

In our experimental work we investigate the following research questions:

1) How does the choice of hyperparameters influences the generalization properties of an OMM classifier?
2) Does the use of multiple modalities improve the recognition performance?
3) Do OMM classifiers provide robustness towards partial action primitives segments?

In order to investigate the first question, we carried out five-fold cross validation training on a set of hyperparameters pairs chosen from

$$H := \{(K, \sigma) | K \in \{2, \ldots, 130\},\ \sigma \in \{0.1, \ldots, 1.5\}\},$$

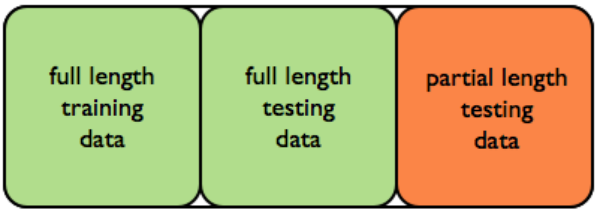

Fig. 6. This figure illustrates the partitions we used for data evaluation. The green color indicates the full length segment data set that was further divided into a training and a testing data set. The red colored square contains partial length data that was used for testing. For cross validation, we used all available segments.

where K denotes the number of OMM states and σ the variance of the corresponding emission distributions.

In addition, to address the second question, we conducted cross validation experiments with all modality combinations: *tactile, joints, audio, joints & tactile, joints & audio, tactile & audio, joints & tactile & audio*[2].

To analyze the robustness properties of OMM classifiers with regard to the generated action primitives, we selected two types of data segments according to overlap with ground truth. For the examples of the first set the overlap with ground truth is on average $\approx 10\%$. This highly partial length segments correspond to a small region in the beginning of an action primitive. In contrast, the second set contains segments with $\approx 95\%$ overlap with ground truth. These segments correspond to almost full length action primitives. In the following, we refer to first selection as *partial length* data and to the second data set as *full length* data. For each modality combination, we randomly divided the full length data in training and test sets. We then trained OMM classifiers with the training data sets. In these experiments, we chose the optimal hyperparameter pairs (K^*, σ^*) yielded in cross validation. We obtained the final test set accuracy values by applying the resulting classifiers to both, the full length and the partial length test data set. See Figure 6 for an illustration of the data partitions.

We applied short-time Fourier transformation to the raw audio signal and used the absolute values of the first ten Fourier coefficients as audio modality feature. The data streams from the tactile and joint angle sensors were used without any feature extraction. All data was normalized to zero mean and unit variance.

VI. Results and Discussion

Figure 7 illustrates the dependency between the hyperparameters $(K,\ \sigma)$ and the cross validation accuracy for the *tactile & joints & audio* and *audio* sets, respectively. This figure clearly demonstrate that with an increasing number of states K and a growing value of emission distribution parameter σ the accuracy remains stable.

[2]In the following we will use the italic font to refer to data set containing the corresponding modalities, e.g. *joints & audio & tactile* stands for all available modality data, *joints & tactile* is the data set in which the audio data channel is omitted.

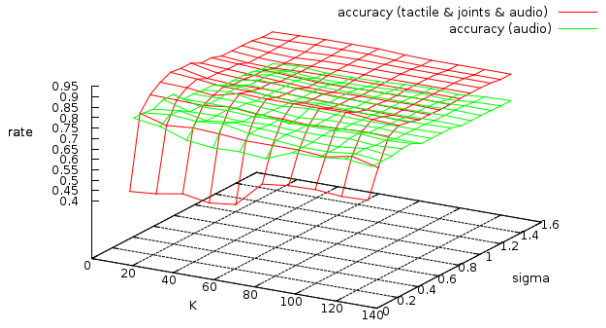

Fig. 7. Dependency between the number of OMM states K, emission distribution parameter σ and the classification rate; *tactile & joints & audio audio* data.

	Hyperparameters	
Sensors	K^*	σ^*
tactile	28	1.2
joints	4	0.9
audio	16	0.4
joints & tactile	4	1.0
joints & audio	8	0.5
tactile & audio	8	0.3
joints & tactile & audio	8	0.7

TABLE I
THIS TABLE SHOWS THE BEST HYPERPARAMETERS FOR ALL SENSOR COMBINATIONS THAT WERE FOUND IN GRID SEARCH.

For $\sigma > 0.3$ OMM classifiers trained with *tactile & joints & audio* reach an average accuracy of ≈ 0.87 with a standard deviation of ≈ 0.02. Classifiers trained with *audio* data only yield an average cross validation performance of ≈ 0.74 with the same standard deviation of ≈ 0.02. These results indicate that OMM classifiers are able to provide good classification results for the evaluated set of action primitives almost independently of the chosen hyperparameters. In particular, a varying number of model states K does not lead to substantial variations in classification performances. Additionally, OMM classifiers are generally robust according to changes in the variance parameter σ. In case of *tactile & joints & audio* small σ values result in loss of cross validation performance. However, for $0.3 < \sigma < 1.5$ the accuracy stabilizes on a high level.

	Accuracy		
Sensors	cross validation	full length	partial length
tactile	≈ 0.64	≈ 0.76	≈ 0.64
joints	≈ 0.69	≈ 0.84	≈ 0.62
audio	≈ 0.79	≈ 0.91	≈ 0.70
joints & tactile	≈ 0.69	≈ 0.89	≈ 0.60
joints & audio	$\approx \mathbf{0.93}$	$\approx \mathbf{0.89}$	$\approx \mathbf{0.79}$
tactile & audio	$\approx \mathbf{0.89}$	$\approx \mathbf{0.97}$	$\approx \mathbf{0.83}$
joints & tactile & audio	$\approx \mathbf{0.95}$	$\approx \mathbf{0.97}$	$\approx \mathbf{0.84}$

TABLE II
THIS TABLE SHOWS THE CLASSIFICATION ACCURACY FOR ALL SENSOR COMBINATIONS. THE SECOND ROW DENOTES THE CROSS VALIDATION ACCURACY FOR THE COMPETE DATA SET, THE THIRD ROW SHOWS THE ACCURACY FOR THE TEST SET THAT CONTAINS THE SUFFICIENT SEGMENTS, AND THE FOURTH ROW SHOWS THE CORRESPONDING CLASSIFICATION ACCURACY FOR THE TEST SET WITH INSUFFICIENT SEGMENTATION EXAMPLES.

The second column of the Table II presents the best results of cross validation accuracy for all evaluated modality combinations with each row containing the highest reached accuracy value. These results indicate that combinations of modalities provide superior representations for action primitives. Two out of three single modalities reach substantially lower classification rates in cross validation as compared to the modality combinations. Only in experiments with separated audio modality OMM classifiers outperform classifiers that make use of *joints & tactile* by ≈ 0.1. The highest cross validation performance of ≈ 0.95 yielded OMM classifiers that used all available data *joints & tactile & audio* with $K = 8$ model states and $\sigma = 0.7$. In addition, these results underline the importance of the audio modality. Classifiers trained with the single audio signal reach the highest cross validation accuracy of ≈ 0.79. The classifier with combinations of audio and other modalities yield substantially higher classification rates of ≈ 0.89 (*joints & audio*) and ≈ 0.93 (*tactile & audio*) as compared to *joints & tactile* with ≈ 69.

The results in the Table II demonstrate that incorporating additional modalities improves classification in most cases. E.g., adding audio modality to the joint or the tactile modality increases the recognition rate by over 20 percentage points. For single modality, the highest rate of 0.79 have been achieved for the audio modality.

The third and fourth rows of the Table II illustrate the test set accuracies that has been achieved in classification experiments for full length and partial length segments, respectively. In these experiments, classifiers with full length segments that use all available modalities reach a recognition rate of ≈ 0.97. Here, the classification performance for full length segments is between ≈ 0.12 (*tactile*) and ≈ 0.29 (*joints&tactile*) higher as compared to results from partial length segments. However, OMM classifiers for partial length segments still reach good accuracy levels of up to ≈ 0.84. These results indicate that OMM classifier provide robustness towards highly partial data segments, in particular if all modalities are used.

VII. Conclusions and Outlook

In this paper we proposed a robust multimodal approach towards learning of object manipulation operation in a continuous sequence. Our approach combines unsupervised segmentation of action sequences and supervised learning with ordered means models.

In our experiments we examined a set of uni- and bimanual operations: picking up, holding, shaking, putting down, screwing and unscrewing the cap, pouring. The recorded modalities were: joint-angles of the hand and palm, force

feedback of the fingers and the audio signal accompanying the object manipulation.

All experiments showed strong benefits of using multiple modalities for supervised recognition with ordered means models. Incorporation of additional modalities improved classification performance in almost all cases. In all experiments the highest recognition has been achieved for the combination of all modalities *tactile & Fourier & audio*. Cross validation on the complete data set yielded the highest recognition rate of ≈ 0.95. Two experiment have been conducted in order to evaluate the robustness of our approach w.r.t. segments overlap with ground truth. For this purpose two sets of segments have been selected: partial length segment set ($\approx 10\%$ overlap with ground truth on average) and full length segments ($\approx 95\%$ overlap with ground truth on average). For the partial length segment set we achieved the highest recognition rate of 0.84; full length set yielded the recognition rate of $\approx 97\%$.

Our future work will be concerned with unsupervised and on-line classification of action primitives based on a larger data pool and a wider range of actions.

REFERENCES

[1] D. Kulić, W. Takano, and Y. Nakamura, "Incremental learning, clustering and hierarchy formation of whole body motion patterns using adaptive hidden markov chains," *The International Journal of Robotics Research*, vol. 27, no. 7, p. 761, 2008.

[2] J. Park, S. Park, and J. Aggarwal, "Model-based human motion tracking and behavior recognition using hierarchical finite state automata," *Computational Science and Its Applications-ICCSA 2004*, pp. 311–320, 2004.

[3] A. Ogale, A. Karapurkar, and Y. Aloimonos, "View-invariant modeling and recognition of human actions using grammars," in *Workshop on Dynamical Vision at ICCV*, vol. 5. Springer, 2005.

[4] Y. A. Ivanov and A. F. Bobick, "Recognition of visual activities and interactions by stochastic parsing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 852–872, 2000.

[5] J. Steffen, M. Pardowitz, and H. Ritter, "A manifold representation as common basis for action production and recognition," in *32nd German Conference on Artificial Intelligence (KI-2009)*, Springer Berlin Heidelberg. Paderborn, Germany: Springer Berlin Heidelberg, 09/2009 2009, pp. 607 – 614.

[6] P. Turaga, R. Chellappa, V. Subrahmanian, and O. Udrea, "Machine recognition of human activities: A survey," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 11, pp. 1473–1488, 2008.

[7] K. Bernardin, K. Ogawara, K. Ikeuchi, and R. Dillmann, "A sensor fusion approach for recognizing continuous human grasping sequences using hidden markov models," *Robotics, IEEE Transactions on*, vol. 21, no. 1, pp. 47–57, 2005.

[8] G. Ogris, T. Stiefmeier, P. Lukowicz, and G. Troster, "Using a complex multi-modal on-body sensor system for activity spotting," *Wearable Computers, IEEE International Symposium*, pp. 55–62, 2008.

[9] T. Stiefmeier, G. Ogris, H. Junker, P. Lukowicz, and G. Troster, "Combining motion sensors and ultrasonic hands tracking for continuous activity recognition in a maintenance scenario," *Wearable Computers, IEEE International Symposium*, pp. 97–104, 2006.

[10] V. Krüger, D. Kragic, A. Ude, and C. Geib, "The Meaning of Action: A Review on action recognition and mapping," *Advanced Robotics*, vol. 21, no. 13, pp. 1473–1501, 2007.

[11] A. Barchunova, J. Moringen, U. Großekathöfer, R. Haschke, S. Wachsmuth, H. Janssen, and H. Ritter, "Unsupervised identification of object manipulation operations from multimodal input (submitted)," in *International Conference on Intelligent Robots and Systems*, 2011.

[12] P. Fearnhead, "Exact and efficient bayesian inference for multiple changepoint problems," *Statistics and Computing*, vol. 16, pp. 203–213, June 2006.

[13] U. Großekathöfer, T. Lingner, H. Ritter, and P. Meinicke, "Ordered means models for analysis and classification of time series," 2011 (submitted).

[14] N. Wöhler, U. Großekathöfer, A. Dierker, M. Hanheide, S. Kopp, and T. Hermann, "A calibration-free head gesture recognition system with online capability," in *International Conference on Pattern Recognition*. IEEE Computer Society, 2010, pp. 3814–3817.

[15] T. Großhauser, U. Großekathöfer, and T. Hermann, "New Sensors and Pattern Recognition Techniques for String Instruments," in *International Conference on New Interfaces for Musical Expression*, 2010.