IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.

Unsupervised human motion segmentation based on characteristic force signals of contact events

Keito Sugawara[1], Sho Sakaino[2], and Toshiaki Tsuji[1]

***Abstract*—Humans perform complex tasks involving force interactions daily. Learning from demonstration, a method for transferring such human manipulation skills to robots, requires techniques for segmenting the demonstrations into movement primitives. Therefore, we propose an unsupervised motion segmentation method that utilizes small characteristic fluctuations of 6-axis force/torque signals as features for motion segmentation. This method includes a feature extraction using a time derivative process and detects segmentation points based on the time derivative of 6-axis force/torque signals obtained during the task demonstrations. The segmentation method was evaluated using a peg-in-hole task and bottle-lid opening task. The experimental results demonstrate the validity of using time derivative of forces and torques for motion segmentation.**

***Index Terms*—Learning from Demonstration, Task and Motion Planning, Force and Tactile Sensing.**

I. Introduction

The automation of contact-rich tasks, such as assembly and tool manipulation, will expand the field of robotic applications. For the motion planning of contact-rich tasks, it is necessary to model multiple contact states and determine the robotic actions according to these states. Programming these tasks is difficult, although humans perform them daily. Therefore, learning from demonstration (LfD) has been studied, which is a method for teaching robots to use human motions [1], [2]. This enables teaching tasks that are difficult to program in a more intuitive manner. Although certain LfD approaches learn the sequence of the overall demonstrations [3], [4], others divide the demonstration into several simple subtasks. In this approach, the task consists of several simple movements called movement primitives. Learning the movement primitives and their sequences from demonstrations is essential for automating the manipulation skills in contact-rich tasks.

Techniques for segmenting demonstrations into primitives are critical in addressing this issue. The supervised approach for segmentation is significantly labor intensive because the demonstrator must manually label the segmentation points at each demonstration; therefore, an unsupervised approach is desirable. Niekum *et al.* proposed an unsupervised segmentation method based on the Gaussian mixture model [5]. Lee *et al.* proposed a segmentation method based on a hidden Markov model (HMM) and applied it to robots to acquire skills from demonstrations [6]. Examples of other segmentation methods include zero-velocity crossings (ZVC) [7], using motion features and object distances [8], the Bayesian changepoint inference [9], [10], and a deep learning method based on variational autoencoders [11]. These methods perform motion segmentation based on the position and velocity information.

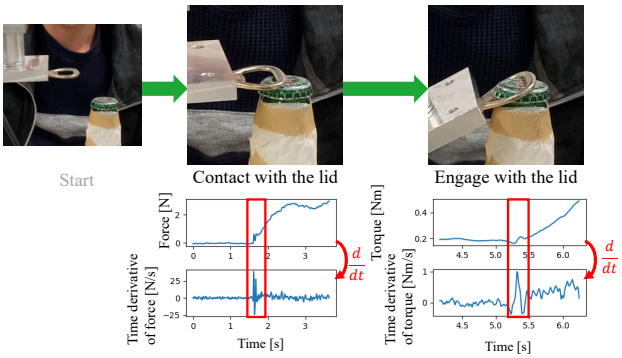


Fig. 1. Overview of feature extraction.

In contact-rich tasks, the contact state between objects changes several times. Humans perform tasks incrementally using sensorimotor control points represented by contact events [12]. In LfD studies, phasing based on contact states has been proposed to be effective for acquiring skills in contact-rich tasks [13]. Therefore, unsupervised motion segmentation based on contact-state switching is desirable for learning contact-rich tasks. The position information used for motion segmentation in [5]–[11] is not a physical quantity that directly represents a contact. Therefore, studies have been conducted to perform motion segmentation by actively using the contact information obtained from force or tactile sensors. The simplest method involves performing segmentation when the contact force exceeds a certain threshold [14]. Another example is a supervised method that combines a support vector machine and HMM [15]. Certain probabilistic segment point detection approaches also use force information [16]–[18] to detect contact state transitions and have demonstrated that the performance of contact-rich tasks improves with the planning of contact state transitions. Motion segmentation by applying ZVC for force information [19] is an effective method; however, ZVC tends to oversegment when the data contains a significant amount of noise or have a high degree of freedom [20]. Because the reproducibility of the contact force of human motion is low, segmentation based on threshold

Manuscript received: March, 21st, 2023; Revised June, 5th, 2023; Accepted July, 8th, 2023.

This paper was recommended for publication by Editor Aleksandra Faust upon evaluation of the Associate Editor and Reviewers' comments. This work was supported in part by JSPS KAKENHI, Japan Grant Number 21H01280.

[1]Keito Sugawara and Toshiaki Tsuji are with the Graduate School of Science and Engineering, Saitama University, 255 Shimo-okubo, Sakura, Saitama, 338-8570 (e-mail: tsuji@ees.saitama-u.ac.jp)

[2]Sho Sakaino is with the Graduate School of Systems and Information Engineering, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8577 (e-mail: sakaino@iit.tsukuba.ac.jp)

Digital Object Identifier (DOI): see top of this page.

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.

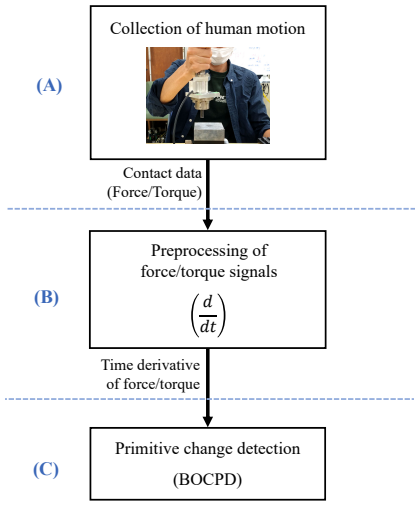


Fig. 2. Overview of the proposed method.

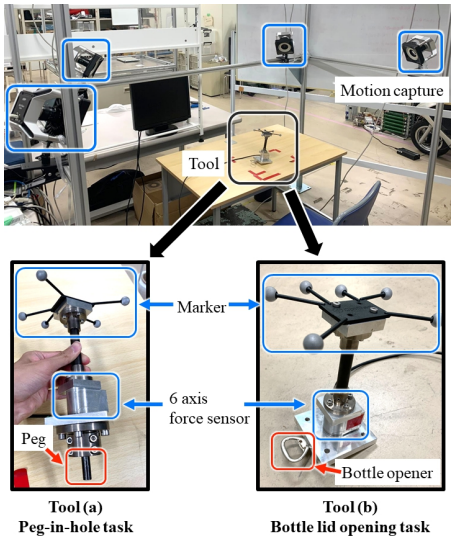


Fig. 3. Setup and tools for data collection.

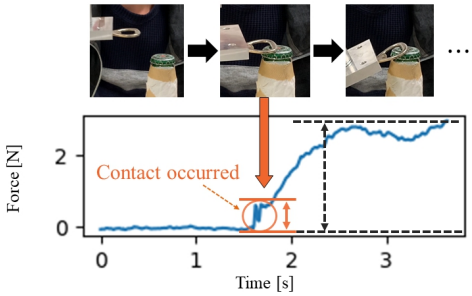


Fig. 4. Force response and range when the bottle opener contacts the lid.

values is generally unreliable. At this type of segmentation point, a change occurs in the force response as the contact state switches, which can be detected [21], [22]. However, this idea has never been utilized for motion segmentation, and previous studies have only considered the absolute value of the force.

Therefore, we propose an unsupervised motion-segmentation technique that considers the changes in the force to reliably segment unrepeatable human motions. To perform segmentation based on the amount of change, features were obtained by differentiating the force and torque obtained from the force sensor, as shown in Fig. 1.

To perform segmentation using the noisy force sensor signals and their derivatives, we used an unsupervised segmentation method called Bayesian online change-point detection (BOCPD) [23], which considers the effect of noise in a probabilistic model. The experiments verified whether the aforementioned process can detect the segmentation point of an object-manipulation task without prior labeling.

The remainder of this paper is organized as follows. Section II describes the proposed segmentation method. Section III describes the experiments and their results. Finally, Section IV discusses the conclusions and proposes future research directions.

II. Method

The proposed unsupervised motion segmentation method is presented herein, for which an overview is demonstrated in Fig. 2. The force information of human motion is first obtained, which is followed by obtaining its derivative value through preprocessing. Finally, motion segmentation based on BOCPD is performed. The steps involved in this process are presented below.

A. Data collection of human motion

In this study, data collection of the human motion was achieved by embedding sensors in tools that can be regarded as rigid bodies. Fig. 3 presents the setup and tools used for data acquisition. Four motion capture cameras (Prime41, NaturalPoint) and a 6-axis force sensor (CFS034CA301U, Leptrino) were synchronized to acquire the motion and contact information during the task demonstration. Markers of the motion capture and 6-axis force sensor were attached to the tool, and the position (p_x, p_y, p_z) [m], posture $(\mathrm{roll}, \mathrm{pitch}, \mathrm{yaw})$ [rad], force (F_x, F_y, F_z) [N], and torque (M_x, M_y, M_z) [Nm] were measured. The sampling time for the motion capture and 6-axis force sensor was 8 ms. We applied a first-order low-pass filter with an 8 Hz cut-off frequency to remove noise after data acquisition.

B. Preprocessing of force/torque signals

Characteristic fluctuations were extracted using the obtained time-series force/torque signals. As shown in Fig. 1, we focused on the instantaneous fluctuations of the force/torque signals and used this feature for unsupervised motion segmentation. Fig. 4 presents the force response when the bottle opener contacted the top of the lid. A fluctuation with an amplitude of approximately 0.5 N occurred. Subsequently, the value changed by approximately 3 N with a smaller slope. In this case, the first 0.5 N fluctuation should be considered for motion segmentation because the kinetic constraint changes with contact. However, such responses are often small in terms of the force dimension, as inferred from this example. This study assumes that this issue is a bottleneck in the

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.

performance of segmentation and therefore introduces a time derivative process to extract the feature. The time derivative of force and torque signals are calculated and used for primitive change detection. Time derivative, on the other hand, brings the problem of amplifying noise. However, due to the BOCPD's ability to statistically distinguish between high-frequency noises and characteristic signals, this problem is less significant. In other words, the proposed method has the characteristics of a complementary combination of the features of BOCPD and time derivative.

C. Primitive change detection

In this subsection, a method for unsupervised segmentation of human motion data by using time derivative of force and torque signals is presented. Force sensor signals often contain extensive noise, and the time derivative of values used in this study are even noisier. To robustly detect the segmentation point, a probability model was set up based on the Bayesian statistics [10] to account for such noise. BOCPD is a type of Bayesian change-point detection method that allows the online estimation of noise effects by estimating the probability distributions from time-series signals [24]. Because it is a suitable method for unsupervised segmentation while considering the effects of noise online, BOCPD was used to detect the change-points.

BOCPD estimates the elapsed time step from the previous change-point by using observed signals. In this study, r_t denotes the time step from the previous change point at time $t\,(=1,2,3,...)$ [step]. The time derivative of force and torque signals $\boldsymbol{x}_t$ at time t are defined as follows:

$$\boldsymbol{x}_t = \frac{d}{dt}\left[F_x(t), F_y(t), F_z(t), M_x(t), M_y(t), M_z(t)\right]^\top \quad (1)$$

where $F_x(t)$ denotes the value of F_x at time t, and the same rule is applied to the other axes.

The posterior distribution $P(r_t|\boldsymbol{x}_{1:t})$ is expressed as follows:

$$P(r_t|\boldsymbol{x}_{1:t}) = \frac{P(r_t, \boldsymbol{x}_{1:t})}{\sum_{r_t} P(r_t, \boldsymbol{x}_{1:t})} \quad (2)$$

where $\boldsymbol{x}_{1:t} = \boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_t$ denote all the observations from $t=1$. $P(r_t, \boldsymbol{x}_{1:t})$ in (2) can be expressed as follows:

$$P(r_t, \boldsymbol{x}_{1:t}) = \sum_{r_{t-1}} P(r_t|r_{t-1}) P(\boldsymbol{x}_t|r_{t-1}, \boldsymbol{x}^r_{t-1}) P(r_{t-1}, \boldsymbol{x}_{1:t-1}) \quad (3)$$

where $\boldsymbol{x}^r_{t-1} = \boldsymbol{x}_{(t-r_t):t-1}$ represents all the observations from the previous change point to $t-1$. If $r_t = 0$, then $\boldsymbol{x}^r_{t-1}$ is empty. $P(r_t|r_{t-1})$ is the prior changepoint, and $P\left(\boldsymbol{x}_t|r_{t-1}, \boldsymbol{x}^r_{t-1}\right)$ is the posterior predictive distribution of $\boldsymbol{x}_t$.

The prior changepoint $P(r_t|r_{t-1})$ is expressed by the following equation using the hazard function $H(t)$.

$$P(r_t|r_{t-1}) = \begin{cases} H(r_{t-1}+1) & (r_t = 0) \\ 1 - H(r_{t-1}+1) & (r_t = r_{t-1}+1) \\ 0 & (\text{others}). \end{cases} \quad (4)$$

$H(t)$ can be expressed by the following equation using geometric distribution $P_g(t)$ as follows:

$$H(t) = \frac{P_g(t)}{\sum_{s=t}^{\infty} P_g(s)}. \quad (5)$$

If the timescale of $P_g(t)$ is λ, then $H(t)$ is constant $\frac{1}{\lambda}$ regardless of t.

The posterior predictive distribution $P\left(\boldsymbol{x}_t|r_{t-1}, \boldsymbol{x}^r_{t-1}\right)$ is the predictive distribution for $\boldsymbol{x}_t$ and sets the probability distribution in advance. $P\left(\boldsymbol{x}_t|r_{t-1}, \boldsymbol{x}^r_{t-1}\right)$ is assumed to only depend on data from previous change points.

In this study, $P\left(\boldsymbol{x}_t|r_{t-1}, \boldsymbol{x}^r_{t-1}\right)$ is the predictive distribution of the time derivative of force and torque signals in a total of 6 dimensions. $P\left(\boldsymbol{x}_t|r_{t-1}, \boldsymbol{x}^r_{t-1}\right)$ is calculated by multiplying the probability density for each $\boldsymbol{x}_t$ element. Here, we define set $\boldsymbol{O}_t$, which denotes each element of $\boldsymbol{x}_t$.

$$\boldsymbol{O}_t = \left\{\frac{dF_x(t)}{dt}, \frac{dF_y(t)}{dt}, \frac{dF_z(t)}{dt}, \frac{dM_x(t)}{dt}, \frac{dM_y(t)}{dt}, \frac{dM_z(t)}{dt}\right\} \quad (6)$$

For each $\boldsymbol{O}_t$ component, a normal distribution was set as the probability distribution. Because the offsets of force and torque are removed by time derivative, the distribution was set as a normal distribution with a mean of zero and unknown precision $\lambda = \sigma^{-2}$. Note, a_0 and b_0 are parameters of the gamma distribution and are determined a priori.

$$P(\lambda) = \frac{1}{\Gamma(a_0)} b_0^{a_0} \lambda^{a_0 - 1} \exp(-b_0 \lambda). \quad (7)$$

The predictive distribution $P\left(x_t|r_{t-1}, x^r_{t-1}\right)$ for the observed value $x_t \in O_t$ is expressed as follows, which was obtained using Student's t-distribution $t_\nu(x_t|\mu, \sigma)$ where the mean is μ, standard deviation is σ, and degree of freedom is ν [25]:

$$P\left(x_t|r_{t-1}, x^r_{t-1}\right) = t_{2a_n}(x_t|0, \sigma_t) \quad (8)$$

$$\sigma_t^2 = \frac{b_t}{a_t} \quad (9)$$

$$a_t = \frac{r_t}{2} + a_0 \quad (10)$$

$$b_t = \frac{1}{2} \sum_{i=t-r_t}^{t-1} x_i^2 + b_0. \quad (11)$$

If $r_t = 0$ and $b_t = b_0$, the predictive distribution $P\left(\boldsymbol{x}_t|r_{t-1}, \boldsymbol{x}^r_{t-1}\right)$ is expressed as follows:

$$P\left(\boldsymbol{x}_t|r_{t-1}, \boldsymbol{x}^r_{t-1}\right) = \prod_{x_t \in O_t} P\left(x_t|r_{t-1}, x^r_{t-1}\right). \quad (12)$$

The time derivative of force and torque signals are input to BOCPD to calculate $P(r_t|\boldsymbol{x}_{1:t})$ and determine the $r_t^{\text{amax}}\,(=1,2,...,t-1)$ that obtains the highest probability at $P(r_t|\boldsymbol{x}_{1:t})$ as follows:

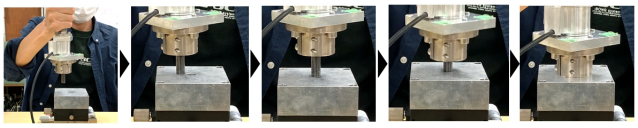
$$r_t^{\text{amax}} = \arg\max_{r_t} P(r_t|\boldsymbol{x}_{1:t}). \quad (13)$$

If $r_t^{\text{amax}} = r_{t-1}^{\text{amax}} + 1$, it is not a change point but a continuation, and if r_t^{amax} is smaller than r_{t-1}^{amax}, then t is a time of primitive change.

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.

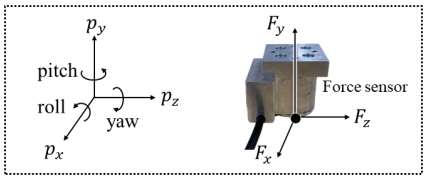
TABLE I
COMPARISON CONDITIONS AND DEFINITIONS

Method	Input	Definition
ZVC	Position, posture (Pos)	ZVC(Pos)
	Force, torque (F/T)	ZVC(F/T)
BOCPD	Pos	BOCPD(Pos)
	F/T	BOCPD(F/T)
	Time derivative of Pos	BOCPD(Vel)
	Pos, F/T	BOCPD(Pos+F/T)
	Pos, F/T, time derivative of Pos Time derivative of F/T	BOCPD(All)
	Time derivative of F/T	**Proposed**



(a) Peg-in-hole task

(b) Bottle-lid opening task



(c) Axis direction

Fig. 5. Flow of experiment tasks.

III. EXPERIMENTS

In this section, we describe the experiments conducted to evaluate the proposed motion-segmentation method. This study was conducted in accordance with Article 16 of the Code of Ethics, which was reviewed and approved by the ethics committee regarding research involving human subjects at the National University Corporation, Saitama University.

A. Segmentation of object manipulation tasks

We verified the proposed method on two different tasks, a peg-in-hole task and a bottle-lid opening task, to check for task dependence, as shown in Fig. 5. The participants performed 20 trials for each task without any instructions regarding the execution speed. The number of trials was determined based on a previous study [17]. We compared the proposed method with the baseline methods of ZVC and BOCPD with other input values. Table I lists a comparison of the conditions used in the experiment. Each definition in Table I represents a description of each method. Here, Pos, F/T and Vel stand for position, force/torque and velocity, respectively. In the baseline methods using BOCPD, the posterior predictive distribution $P\left(x_t|r_{t-1}, x^r_{t-1}\right)$ of the position, posture, force, and torque was a normal distribution with an unknown mean μ and known standard deviation σ_x. The prior distribution of the mean μ was set to the normal distribution $\mathcal{N}\left(\mu|\mu_0, \sigma_0\right)$ with a mean of μ_0 and standard deviation σ_0. The predictive distribution for one axis is as follows [25]:

$$P\left(x_t|r_{t-1}, x^r_{t-1}\right) = \mathcal{N}\left(x_t|\mu_t, \sigma_t\right) \tag{14}$$

$$\mu_t = \frac{r_t\sigma_o^2}{r_t\sigma_0^2+\sigma_x^2}\frac{\sum_{i=t-r_t}^{t-1} x_i}{r_t} + \frac{\sigma_x^2}{r_t\sigma_0^2+\sigma_x^2}\mu_0 \tag{15}$$

$$\sigma_t^2 = \left(\frac{r_t}{\sigma_x^2} + \frac{1}{\sigma_0^2}\right)^{-1} + \sigma_x^2. \tag{16}$$

If $r_t = 0$, then $\mu_t = \mu_0$.

The BOCPD hyperparameters for the proposed method are expressed in Eqs. (4), (10), and (11). They were empirically decided to be $a_0 = 1.0, b_0 = 20.0, H(r_{t-1}+1) = 200^{-1}$.

The hyperparameters of the baseline method are tuned based on the F-score. The value of μ_0 of the baseline BOCPD was set to the average value obtained for each position, posture, force, torque, position's time derivative, and posture's time derivative. The value with the highest F-score was chosen by grid search from $\sigma_x^2 \in [0.01, 0.05, 0.1, 0.5, 1.0, 5.0]$, $\sigma_0^2 \in [0.1, 0.5, 1.0, 5.0, 10.0]$ for BOCPD(Pos), BOCPD(F/T), and BOCPD(Vel). The same parameters indicated above were used for BOCPD(Pos+F/T) and BOCPD(All). Additionally, in these methods, $H(r_{t-1}+1) = 200^{-1}$. The ZVC determines the segmentation points at which the time derivative input value of each axis exceed a certain threshold rather than crossing zero. The threshold was determined by grid search to maximize the F-score.

The input signals were normalized using the standard deviation of all the trials for each position, posture, force, torque, force's time derivative, and torque's time derivative. Segmentation points were detected when r_t^{amax} was smaller than $\frac{1}{2}r_{t-1}^{\mathrm{amax}}$. For ZVC(Pos) and ZVC(F/T), we applied a first-order low-pass filter with a 1 Hz cutoff frequency to reduce the effect of noise after differentiation.

Figs. 6 and 7 present the segmentation results of the peg-in-hole and bottle-lid opening tasks using the proposed method, respectively. The black dotted lines in these figures represent the segmentation results obtained by using the proposed method. Fig. 8 presents the primitives obtained for each task. The red arrows indicate the direction in which the movement is restricted by physical constraints. The blue arrows indicate the direction of the tool movement.

Five primitives, (a) through (e), were detected in the peg-in-hole task, as shown in the top of Fig. 8. In primitive (a), the peg moves to the vicinity of the hole with no constraint owing to the contact. Primitive (b) is an exploratory movement in which a new constraint is generated vertically. Primitive (c) is the insertion movement, in which the peg reaches the top of the hole and is released from the constraint. Primitive (d) is a continuation of primitive (c) with the peg caught during insertion, resulting in horizontal constraints. In primitive (e), the peg is fully inserted and newly constrained in the vertical direction.

Five primitives, (a) through (e), were detected in the bottle-lid opening task, as shown in the middle of Fig. 8. In primitive (a), the opener moves to the lid without constraints. In primitive (b), the opener moves to engage its protrusion with the lid, and a new constraint is generated at the tip of the

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.

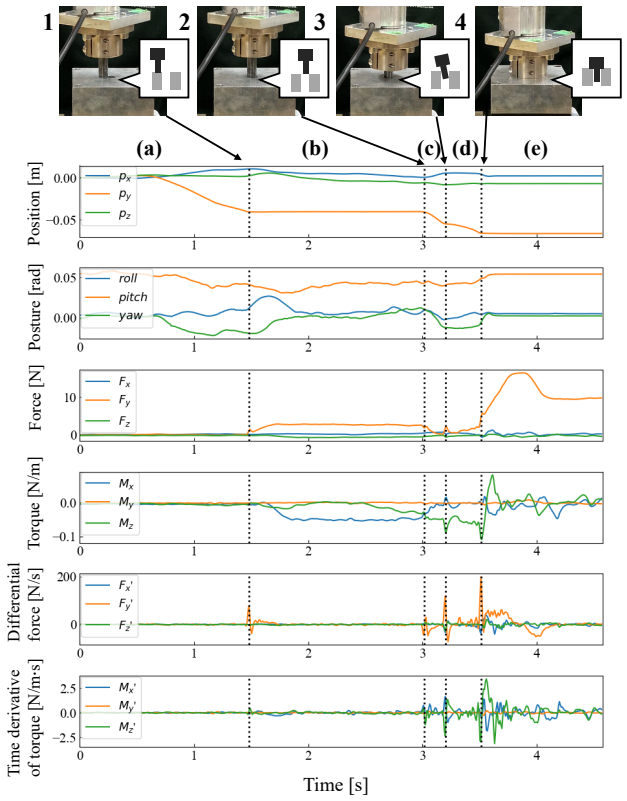


Fig. 6. Segmentation result (dotted lines) of the peg-in-hole task.

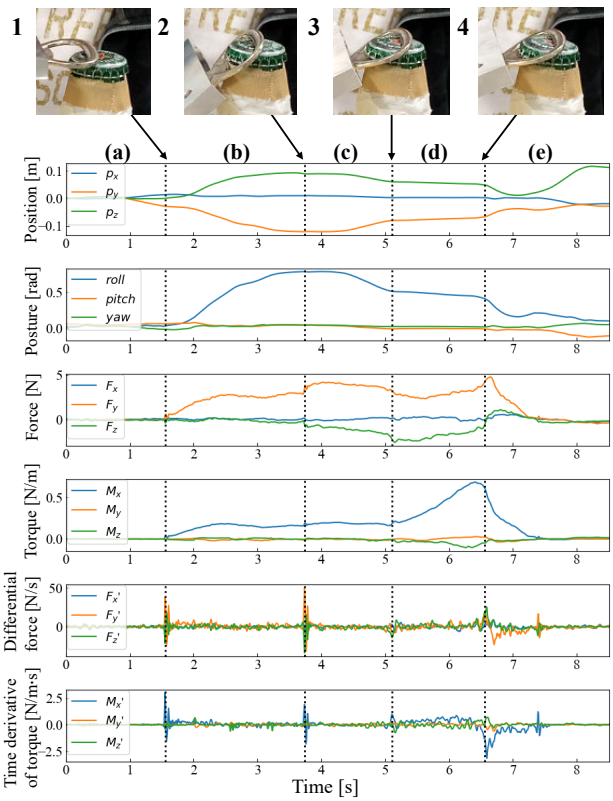


Fig. 7. Segmentation result (dotted lines) of the bottle-lid opening task.

opener. Primitive (c) is a continuation of primitive (b), with a new constraint between the protrusion of the bottle opener and bottle. In primitive (d), the bottle opener is lifted to open the lid with a new constraint on the protrusion, in which the opener moves in a constrained direction, but the movement is restricted by the strong friction between the lid and bottle. In primitive (e), the lid is opened and released from contact. The lid is no longer fixed to the bottle because both the lid and bottle opener are no longer restricted in their movements by the constraints. As indicated above, each primitive reflects the contact state transition owing to a change in the physical constraints.

The detection rate of the aforementioned segmentation points was verified for 20 trials to compare the proposed method with the baseline methods. Additionally, the average number of false positives (FPs) that did not correspond to the segmentation points and F-scores was calculated. The calculation of FPs was done per the number of segmentation points. To calculate the detection rates, FPs, and F-scores, all the trials of each task were labeled with segmentation points. Here, the ground-truth segmentation points were manually labeled. The labeling could be uniquely determined from the video and position/force responses, since the task had different mechanical constraints on all primitives as shown in Fig. 8. Each segmentation point was labeled at certain intervals based on previous relative studies [10], [17], [18], as shown in Fig. 9. The intervals were set to 10 steps, except for segmentation point 4 in the bottle-lid opening task. When the lid opens, the lid rubs against the bottle as it opens, thus requiring time for the transition of the contact state. Therefore, the interval of

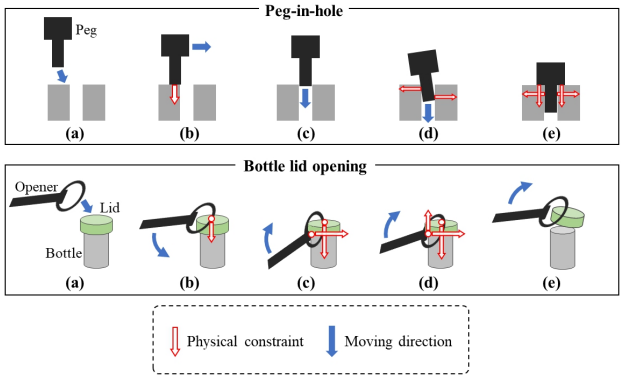


Fig. 8. Acquired motion primitives in Fig. 6 (peg-in-hole task) and Fig. 7 (bottle-lid opening task). The red arrows indicate physical constraints caused by contact. The blue arrows indicate the direction of tool movement.

the label at point 4 of the bottle-lid opening task was set to 25 steps, which was wider than 10 steps.

Tables II and III present the detection rates for the 20 trials of the peg-in-hole and bottle-lid opening tasks, respectively. The segmentation point numbers in Tables II and III correspond to those in Figs. 6 and 7, respectively. Figs. 10 and 11 present the FPs and F-scores of the peg-in-hole and bottle-lid opening tasks, respectively. In the peg-in-hole task, the proposed method demonstrated the highest detection rate for point 1, whereas the detection rate for point 4 was lower than those of other baseline methods. The proposed method achieved the lowest FPs and highest F-scores compared to the baseline methods. In the bottle-lid opening task, the proposed method had a 100% detection rate for points 1–3,

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.

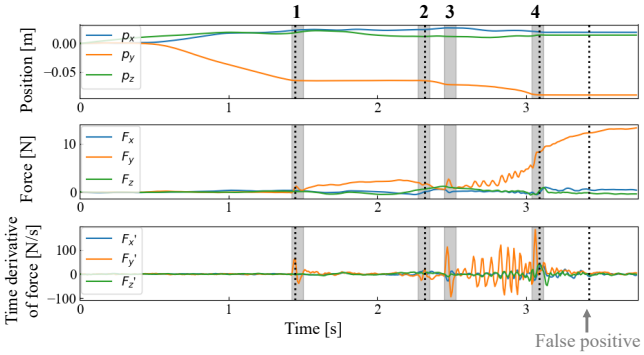


Fig. 9. Method for calculating detection rates. If the segmentation points are within gray area, we considered these points successfully detected the segmentation point. If multiple detections are made within the area, the second and subsequent points are also considered to be false positives.

TABLE II
DETECTION RATES OF PEG-IN-HOLE TASK

Method	Detection rates for each segmentation point [%]			
	1	2	3	4
ZVC(Pos)	15	80	40	90
ZVC(F/T)	0	10	65	**95**
BOCPD(Pos)	0	10	65	45
BOCPD(F/T)	10	55	70	80
BOCPD(Vel)	**95**	**95**	**90**	85
BOCPD(Pos+F/T)	5	25	60	80
BOCPD(All)	0	25	55	80
Proposed	**95**	90	80	65

whereas the detection rate for point 4 was inferior to that of other baseline methods. The F-score was the highest for the proposed method.

Although the time derivative of the F/T responses risks amplifying the noise, BOCPD has an ability to statistically distinguish between high-frequency noises and characteristic signals. In Fig. 9, the BOCPD did not misdetect the segmentation point even though there is a large noise in the time derivative of the force from 2 to 3s. The suppression of FPs in Figs. 10 and 11 also supports this claim.

To show that the ZVC parameters were adjusted to optimal values for each task, the results of varying the ZVC threshold on a separate dataset for parameter adjustment are presented in Fig. 12. The threshold with the highest F-score was used. In order to process inputs with different units in a unified manner, the input signals were time-differentiated after normalization.

TABLE III
DETECTION RATES OF BOTTLE-LID OPENING TASK

Method	Detection rates for each segmentation point [%]			
	1	2	3	4
ZVC(Pos)	0	0	0	90
ZVC(F/T)	0	5	0	**100**
BOCPD(Pos)	0	0	15	100
BOCPD(F/T)	**100**	**100**	90	95
BOCPD(Vel)	75	50	20	**100**
BOCPD(Pos+F/T)	70	**100**	75	85
BOCPD(All)	25	70	60	95
Proposed	**100**	**100**	**100**	90

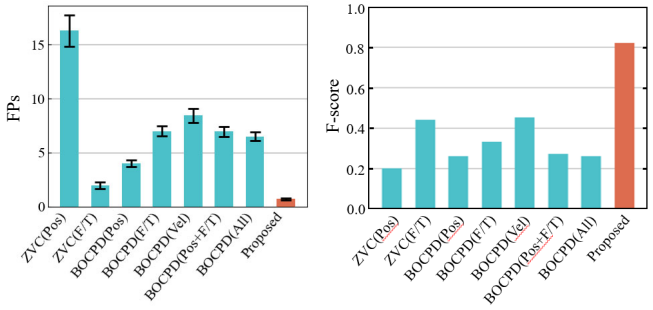


Fig. 10. FPs and F-score of peg-in-hole task

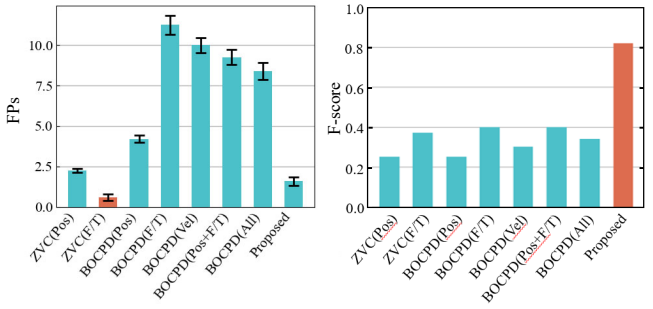


Fig. 11. FPs and F-score of bottle-lid opening task

B. Experiment regarding the inclusion of individual differences

Because the transition of the contact state in a contact-rich task mainly depends on the mechanical constraints between the contacting objects, the motion data are largely task-dependent. However, the collected motion data may include individual differences and fluctuations between trials. Furthermore, it is impractical to sequentially adjust the hyperparameters to accommodate different demonstrators. To test the ability of the proposed method to handle individual differences, we included two additional demonstrators: demonstrator (1) and demonstrator (2), and compared their respective movements. Their FPs and F-scores are compared in Figs. 13 and 14.

The target task was bottle-lid opening, and the same hyperparameters as those described in Section III-A were used. The label interval of segmentation point 4 was set for each demon-

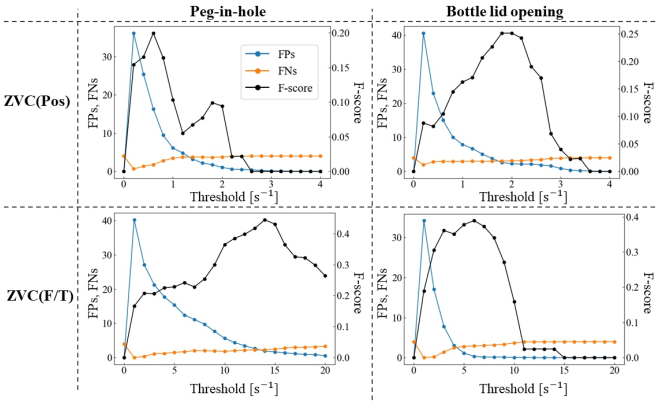


Fig. 12. Changes in FPs, FNs and F-score at different thresholds in ZVC.

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.

TABLE IV
DETECTION RATES OF DEMONSTRATOR (1)

Method	Detection rates for each segmentation point [%]			
	1	2	3	4
ZVC(Pos)	20	30	10	**100**
ZVC(F/T)	0	30	50	**100**
BOCPD(Pos)	10	0	40	50
BOCPD(F/T)	80	**90**	**80**	**100**
BOCPD(Vel)	20	10	0	50
BOCPD(Pos+F/T)	50	80	70	**100**
BOCPD(All)	10	80	**80**	**100**
Proposed	**90**	**90**	40	90

TABLE V
DETECTION RATES OF DEMONSTRATOR (2)

Method	Detection rates for each segmentation point [%]			
	1	2	3	4
ZVC(Pos)	0	12	30	**100**
ZVC(F/T)	0	0	10	60
BOCPD(Pos)	10	0	10	**100**
BOCPD(F/T)	20	**25**	60	**100**
BOCPD(Vel)	30	0	10	90
BOCPD(Pos+F/T)	10	**25**	70	**100**
BOCPD(All)	0	12	60	**100**
Proposed	**80**	**25**	**90**	90

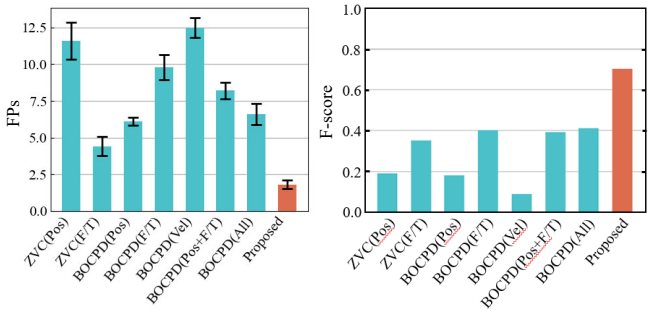


Fig. 13. FPs and F-score of demonstrator (1)

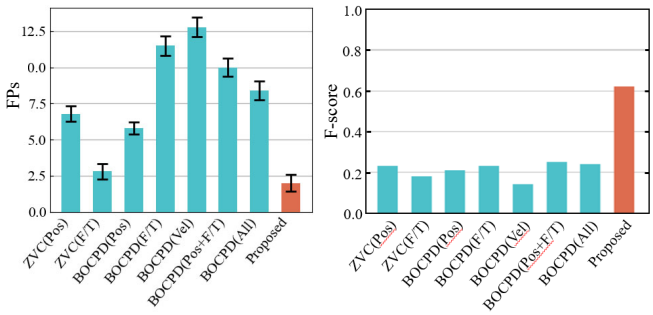


Fig. 14. FPs and F-score of demonstrator (2)

strator considering that each demonstrator had a different task execution time. The interval of segmentation point 4 was set to 10 and 20 steps for demonstrators (1) and (2), respectively. The thresholds of ZVC(Pos) and ZVC(F/T) were the same as those for the bottle-lid opening task described in Section III-A. Ten trials of the task were obtained from each demonstrator. While all the segmentation points 1 to 4 in the bottle-lid opening task were detected for demonstrator (1), point 2 occurred in 8 of 10 trials for demonstrator (2). In general, individual differences can be accommodated without adjusting parameters, but one exception is the intervals between the segmentation points. The intervals depend on the speed of task execution and therefore, the intervals needs to be adjusted depending on the speed. Tables IV and V list the detection rates for each demonstrator. There were a few difficult-to-detect segmentation points for each demonstrator. Despite the different demonstrators, the proposed method achieved the highest F-score compared to the baseline methods.

C. Robot task execution

Finally, we verified the applicability of the proposed method to LfD by having the robot perform a peg-in-hole task based on the acquired primitives. The command trajectory was generated based on the method in [18], [19], which set the final point of the primitives as subgoals. First, each of the multiple trials was segmented by the proposed method and clustered by DBSCAN [26] for positions p_x, p_y, p_z at the final point of the primitives. Next, the transition probabilities of each primitive were calculated, and the primitives were sorted in the order of the highest probability. Finally, the final points of the primitives were connected by an S-curve acceleration/deceleration to generate the target trajectory. Fig. 15 demonstrates the task execution with the primitives obtained by the proposed method. We performed the peg-in-hole task five times with a similar setup and were able to execute the task without failure all five times. Fig. 16 presents the response trajectories of the demonstrations and the command trajectories obtained by the proposed method, BOCPD(F/T), and ZVC(F/T). The orange line in the left graph indicates the subgoals generated based on primitives (a)–(e) that were obtained by using the proposed method. Here, $P_{y,abs}$ and $P_{z,abs}$ denote the absolute position in y and z axes, respectively. Because the subgoals were properly obtained, the peg-in-hole task was successfully completed. The green and blue lines indicate the subgoals generated based on ZVC(F/T) and BOCPD(F/T), respectively. Because primitives (b) and (c) were not segmented in these results and the subgoals were not properly generated accordingly, they were not tested in the experiment for safety reasons. The right graph in Fig. 16 shows the time response of the position of the proposed method. Here, superscript cmd denotes command values. Since primitives are extracted independent of the hole position, here, the position is expressed using relative values with the initial value as the origin. It was confirmed that the peg-in-hole task was carried out by giving the trajectory of each primitive in turn.

IV. CONCLUSION

In this study, we propose a motion segmentation method that utilizes the characteristic force and torque signals caused by contact state changes. We focused on small fluctuations as features for segmentation and extracted them through time derivative. These features were detected by using BOCPD for motion segmentation. In this experiment, we evaluated the segmentation method for peg-in-hole and bottle-lid opening tasks. In addition, we evaluated the differences in the movement and

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.

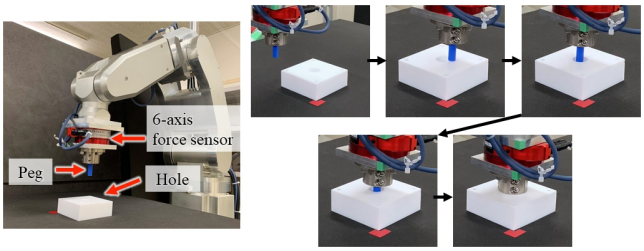


Fig. 15. Peg-in-hole task based on primitives obtained by the proposed method

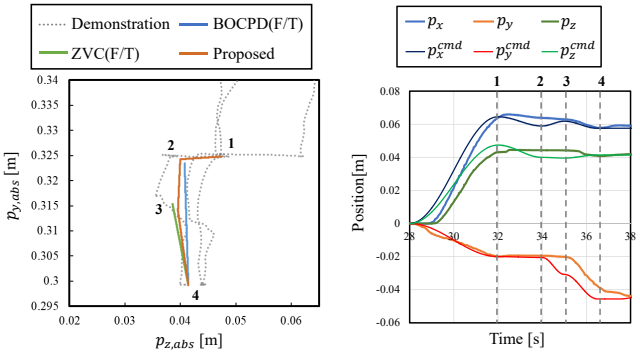


Fig. 16. Trajectories of demonstrations and the commands obtained by the segmentation methods.

force applications for different demonstrators in the bottle-lid opening task. We demonstrated that the time derivative of force and torque signals were effective features for motion segmentation.

The experimental results of BOCPD(all) demonstrate that excluding the force's and torque's time derivatives from the input data leads to poor accuracy. In BOCPD, a simple mixture model is established for the input signal. In such a setup, if any one of the axes is determined to be a segmentation point, the r_t inference is affected by the inappropriateness of the segmentation point. BOCPD(All) is assumed to be affected by the false positives observed in BOCPD(Pos) and BOCPD(F/T), resulting in a worsened accuracy.

The proposed method is based on the contact state transition and does not consider the characteristics of the position/posture information. Using this method combined with other trajectory-based segmentation methods would be effective in utilizing the results of both methods. Future studies will aim to extend this approach to robot-teaching using the motion primitives obtained by the proposed method and to investigate the relationship between the extracted feature axes and the contact state for task progress recognition.

REFERENCES

[1] B. D. Argall, S. Chernova, M. Veloso, and B. Browning: "A survey of robot learning from demonstration," *Robotics and Autonomous Systems*, vol. 57, no. 5, pp. 469–483, 2009.

[2] H. Ravichandar, A. S. Polydoros, S. Chernova, and A. Billard: "Recent advances in robot learning from demonstration," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 3, pp. 297–330, 2020.

[3] D. Furuta, K. Kutsuzawa, and T. Tsuji: "Motion planning with success judgement model based on learning from demonstration," *IEEE Access*, vol. 8, pp. 73142–73150, 2020.

[4] T. Zhang, Z. McCarthy, O. Jow, D. Lee, X. Chen, K. Goldberg, and P. Abbeel: "Deep imitation learning for complex manipulation tasks from virtual reality teleoperation," in *Proc. IEEE Int. Conf. on Robotics and Automation (ICRA)*, pp. 5628–5635, 2018.

[5] S. H. Lee, I. H. Suh, S. Calinon, and R. Johansson: "Autonomous framework for segmenting robot trajectories of manipulation task," *Autonomous Robots*, vol. 38, no. 2, pp. 107–141, 2015.

[6] S. Niekum, S. Osentoski, G. Konidaris, and A. G. Barto: "Learning and generalization of complex tasks from unstructured demonstrations," in *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pp. 5239–5246, 2012.

[7] A. Fod, M. J. Matarić, and O. C. Jenkins: "Automated derivation of primitives for movement classification," *Autonomous Robots*, vol. 12, no. 1, pp. 39–54, 2002.

[8] M. Wächter and T. Asfour: "Hierarchical segmentation of manipulation actions based on object relations and motion characteristics," in *Proc. Int. Conf. on Advanced Robotics (ICAR)*, pp.549–556, 2015.

[9] S. Niekum, S. Osentoski, C. Atkeson, and A. G. Barto: "Online Bayesian changepoint detection for articulated motion models," in *Proc. IEEE Int. Conf. on Robotics and Automation (ICRA)*, pp. 1468–1475, 2015.

[10] L. Senger, M. Schroer, J. H. Metzen, and E. A. Kirchner: "Velocity-based multiple change-point inference for unsupervised segmentation of human movement behavior," in *Proc. 2014 22nd Int. Conf. on Pattern Recognition (ICPR)*, pp. 4564–4569, 2014.

[11] D. Tanneberg, K. Ploeger, E. Rueckert and J. Peters: "SKID RAW: Skill discovery from raw trajectories," *IEEE Robotics and Automation Letters*, vol. 6, no. 3, pp. 4696–4703, 2021.

[12] J. R. Flanagan, M. C. Bowman, and R. S. Johansson: "Control strategies in object manipulation tasks," *Current Opinion in Neurobiology*, vol. 16, no. 6, pp. 650–659, 2006.

[13] M. Skubic and R. A. Volz, "Acquiring robust, force-based assembly skills from human demonstration," *IEEE Transactions on Robotics and Automation*, vol. 16, no. 6, pp. 772–781, 2000.

[14] J. Kober, M. Gienger, and J. J. Steil: "Learning movement primitives for force interaction tasks," in *Proc. IEEE Int. Conf. on Robotics and Automation (ICRA)*, pp. 3192–3199, 2015.

[15] A. Castellani, D. Botturi, M. Bicego, and P. Fiorini: "Hybrid HMM/SVM model for the analysis and segmentation of teleoperation tasks," in *Proc. IEEE Int. Conf. on Robotics and Automation (ICRA)*, vol. 3, pp. 2918–2923, 2004.

[16] M. Karlsson, A. Robertsson, and R. Johansson: "Segmentation of robot movements using position and contact forces," *arXiv preprint arXiv: 1909.08289*, 2019.

[17] Z. Su, O. Kroemer, G. E. Loeb, G. S. Sukhatme, and S. Schaal: "Learning to switch between sensorimotor primitives using multimodal haptic signals," in *Proc. Int. Conf. on Simulation of Adaptive Behavior*, pp. 170–182, 2016.

[18] Z. Su, O. Kroemer, G. E. Loeb, G. S. Sukhatme, and S. Schaal: "Learning manipulation graphs from demonstrations using multimodal sensory signals," in *Proc. IEEE Int. Conf. on Robotics and Automation (ICRA)*, pp. 2758–2765, 2018.

[19] S. Manschitz, M. Gienger, J. Kober, and J. Peters: "Probabilistic decomposition of sequential force interaction tasks into movement primitives," in *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, pp. 3920–3927, 2016.

[20] J. F. -S. Lin, M. Karg, and D. Kulić: "Movement primitive segmentation for human motion modeling: A framework for analysis," *IEEE Transactions on Human-Machine Systems*, vol. 46, no. 3, pp. 325–339, 2016.

[21] T. Tsuji, K. Sato, and S. Sakaino: "Contact feature recognition based on MFCC of force signals," *IEEE Robotics and Automation Letters*, vol. 6, no. 3, pp. 5153–5158, 2021.

[22] A. Stolt, M. Linderoth, A. Robertsson, and R. Johansson: "Detection of contact force transients in robotic assembly," in *Proc. IEEE Int. Conf. on Robotics and Automation (ICRA)*, pp. 962–968, 2015.

[23] R. P. Adams, and D. J. MacKay: "Bayesian online changepoint detection," *arXiv preprint arXiv:0710.3742*, 2007.

[24] N. Sakib, S. Tian, M. M. Haque, R. A. Khan, and S. I. Ahamed: "SepINav (Sepsis ICU Navigator): A data-driven software tool for sepsis monitoring and intervention using Bayesian Online Change Point Detection," *SoftwareX*, vol. 14, pp. 100689, 2021.

[25] K. P. Murphy: "Conjugate Bayesian analysis of the Gaussian distribution," 2007.

[26] M. Ester, H. P. Kriegel, J. Sander, and X. Xu: "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. 2nd Int. Conf. on Knowledge Discovery and Data Mining*, vol. 96, no. 34, pp. 226–231, 1996.

IEEE Robotics and Automation Letters (RA-L) paper, presented at ICRA 2024, Yokohama, Japan. Cite as RA-L paper.