



LUND UNIVERSITY

Learning Basis Skills by Autonomous Segmentation of Humanoid Motion Trajectories

Lee, Sang Hyung; Suh, Il Hong; Calinon, Sylvain; Johansson, Rolf

Published in:

12th IEEE-RAS International Conference on Humanoid Robots (Humanoids), 2012

DOI:

[10.1109/HUMANOIDS.2012.6651507](https://doi.org/10.1109/HUMANOIDS.2012.6651507)

2012

[Link to publication](#)

Citation for published version (APA):

Lee, S. H., Suh, I. H., Calinon, S., & Johansson, R. (2012). Learning Basis Skills by Autonomous Segmentation of Humanoid Motion Trajectories. In *12th IEEE-RAS International Conference on Humanoid Robots (Humanoids), 2012* (pp. 112-119). IEEE - Institute of Electrical and Electronics Engineers Inc.. <https://doi.org/10.1109/HUMANOIDS.2012.6651507>

Total number of authors:

4

General rights

Unless other specific re-use rights are stated the following general rights apply:

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: <https://creativecommons.org/licenses/>

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Learning Basis Skills by Autonomous Segmentation of Humanoid Motion Trajectories

Sang Hyoung Lee, Il Hong Suh[†], Sylvain Calinon, and Rolf Johansson

Abstract—Manipulation tasks are characterized by continuous motion trajectories containing a set of key phases. In this paper, we propose a probabilistic method to autonomously segment the motion trajectories for estimating the key phases embedded in such a task. The autonomous segmentation process relies on principal component analysis to adaptively project into one of the low-dimensional subspaces, in which a Gaussian mixture model is learned based on Bayesian information criterion and expectation-maximization algorithms. The basis skills are estimated by a set of Gaussians approximating quasi-linear key phases, and those times spent calculated from the segmentation points between two consecutive Gaussians representing the local changes of dynamics and directions of the trajectories. The basis skills are then used to build novel motion trajectories with possible motion alternatives and optional parts. After sequentially reorganizing the basis skills, a Gaussian mixture regression process is used to retrieve smooth motion trajectories. Two experiments are presented to demonstrate the capability of the autonomous segmentation approach.

I. INTRODUCTION

A humanoid robot task can be characterized by continuous motion trajectories with a set of key phases (i.e., sub-procedures). An intelligent robot, therefore, should be able to learn the key phases embedded in such tasks. Let us consider an example to intuitively describe such key phases. A humanoid robot learns the key phases from a manipulation task for cooking rice. The complete procedure is as follows: the robot first lifts a pot, which is attached to its right hand, toward a kitchen board. Next, the robot scoops some grains of rice from a rice bowl using a spoon attached to its left hand only once. The rice is delivered from the bowl to the pot. The robot pours the rice into the pot, and then it stirs the rice in the pot using the spoon. Finally, the pot is put on the stove. The robot can provide alternatives by reorganizing and repeating some sub-procedures such as three scooping rice, three delivering the rice, and two stirring the rice when detecting such sub-procedures from the complete procedure.

In this context, we propose an autonomous segmentation approach combining a continuous representation of the

task constraints with a segmentation process relying on the same statistical model. While the portions in-between two Gaussians in a Gaussian mixture model (GMM) are used to estimate basis skills as segmentation points, a Gaussian mixture regression (GMR) process ensures to retrieve the motion trajectories reproducible on the robot. Here, the key phases and those times spent (i.e., time durations) are referred to as the basis skills embedded in such a task.

Many researchers to date have studied such basis skills for achieving a robot task. Billing *et al.*, Cohen *et al.*, Bentivegna *et al.*, Nicolescu *et al.*, and Nejati *et al.* proposed methods for learning and combining pre-segmented and predefined basis skills for achieving given tasks [1]–[5]. Billing *et al.* presented a predictive sequence learning method for recognizing and controlling training data using known basis skills. Cohen *et al.* proposed a heuristic search-based manipulation planner using a set of predefined basis skills. Bentivegna *et al.* presented a method for increasing performance through repeated practices based on a set of predefined basis skills. Nicolescu *et al.* proposed a method for refining, learning, and generalizing basis skills using predefined action networks and instructive demonstration. Nejati *et al.* proposed a method for learning and generalizing known basis skills based on hierarchical task networks. All of these authors have worked on the methods for dealing with predefined basis skills but they did not consider unknown basis skills.

On the other hand, Drumwright *et al.*, Kulic *et al.*, Gribovskaya *et al.*, and Asfour *et al.* have proposed methods for learning unknown basis skills [6]–[9]. They focus on methods for segmenting continuous motion trajectories to learn such basis skills embedded in a task. In this context, Drumwright *et al.* proposed a method that segments joint trajectories. The segmentation points are determined based on the points in which the velocities of the joint trajectories maintained a fixed interval and the sum of the velocities was smaller than a threshold within the interval. Kulic *et al.* presented a method for learning basis skills which involved comparing density distributions with known or unknown models according to a fixed window. Gribovskaya *et al.* reported a method for acquiring basis skills using the sum of the velocities across the threshold in relation to the relative position in Cartesian coordinates. Finally, Asfour *et al.* proposed a method for learning basis skills by extracting common states from several hidden Markov models (HMMs). Here, an HMM is modeled using specific points of continuous trajectories. The specific points are determined by two criteria (i.e., changing direction of trajectories and

*This work was supported by the Global Frontier R&D Program on <Human-centered Interaction for Coexistence> funded by the National Research Foundation of Korea grant funded by the Korean Government(MEST) (NRF-M1AXA003-2011-0028353)

S. H. LEE is with the Department of Electronics and Computer Engineering, Hanyang University, Seoul, Korea. zelog@hanyang.ac.kr

[†]I. H. Suh is with the Department of Computer Science and Engineering, Hanyang University, Seoul, Korea. ihsuh@hanyang.ac.kr. All correspondence should be addressed to I. H. Suh.

S. Calinon is with the Department of Advanced Robotics, Istituto Italiano di Tecnologia (IIT), Genova, Italy. sylvain.calinon@iit.it

R. Johansson is with the Department of Automatic Control, LTH, Lund University, Lund, Sweden. Rolf.Johansson@control.lth.se

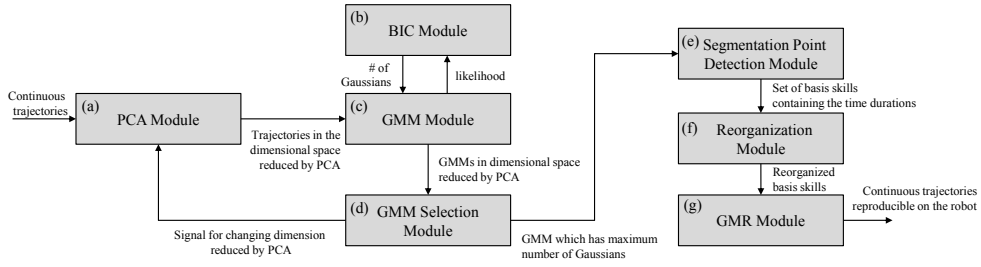


Fig. 1. Graphical flow for learning basis skills based on our autonomous segmentation process; (a) principal component analysis (PCA) module, (b) Bayesian information criterion (BIC) module, (c) Gaussian mixture model (GMM) module, (d) GMM selection module, (e) Segmentation point detection module, (f) Reorganization module, and (g) Gaussian mixture regression (GMR) module.

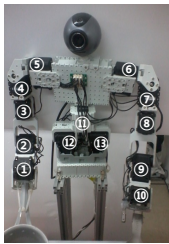


Fig. 2. Humanoid robot including thirteen motors developed at Hanyang University.

stopping trajectories in sufficient time). Although these authors have proposed the solutions for learning unknown basis skills, it is difficult to determine the fixed intervals, window size, fixed time, and predefined models. That is, it is not easy to predefine or tune these parameters according to the types of variables and tasks.

To learn such basis skills involved in a humanoid robot task, Fig. 1 shows an entire flow of our autonomous segmentation process. For segmentation, a GMM is modeled using motion trajectories extracted from a complete demonstration as shown in Fig. 1(c). Before the GMM is automatically estimated based on Bayesian information criterion (BIC) as shown in Fig. 1(b), the motion trajectories are first transformed to one of the low-dimensional spaces by principal component analysis (PCA) as shown in Fig. 1(a). While repeating the processes from Fig.1(a) to Fig.1(c), the GMM selection module chooses the GMM that contains the maximum number of Gaussians among all the GMMs as shown in Fig. 1(d). The segmentation points are next estimated in all points intersected along the time component of the selected GMM as shown in Fig. 1(e). The basis skills acquired by our process are sequentially reorganized using those time durations as shown in Fig. 1(f). Finally, continuous motion trajectories reproducible on the robot are retrieved by Gaussian mixture regression (GMR) as shown in Fig. 1(g). This core segmentation idea is exploited in order to provide alternative solutions by combining with existing methods that reorder the basis skills such as motion planning mechanisms [10]–[12].

The remainder of this paper is organized as follows: section II describes the details of the autonomous segmentation

process for learning basis skills. Section III presents evaluation results using publicly available motion data employed by TUM [13]. Section IV discusses our segmentation process. Finally, section V presents our conclusions and plans for future research.

II. LEARNING BASIS SKILLS EMBEDDED IN A TASK

As noted earlier, Gaussians of a GMM are used as important information for segmenting continuous motion trajectories. Representing the motion trajectories as such a GMM provides a way to encode the local directions of the trajectories through the shape of the Gaussians, as well as the local correlations among the variables taking part in the trajectories [14]. The mean and covariance information of the GMM to date have been exploited to reconstruct the trajectories [15], [16]. But, surprisingly, very little attention has been given to the intersections between two consecutive Gaussians in the trajectories. In this context, the autonomous segmentation process exploits these intersections in the learning process, for segmentation of the trajectories by estimating changes of the local shape and changes of the local correlations among the different variables. To intuitively describe our method, in this paper, we present our segmentation process using produces of each step acquired from a manipulation task for cooking rice introduced in Section I. The joint motion trajectories are recorded at rate of 10 Hz by a kinesthetic teaching method using a humanoid robot. The humanoid robot developed at Hanyang University includes thirteen motors as shown in Fig. 2.

A. Modeling GMM for Autonomous Segmentation

The joint motion trajectories, $\mathbf{X} \in \mathbb{R}^{(D+1) \times N}$, of a robot are extracted from a complete demonstration. Here, $(D+1)$ denotes the D -dimensional spatial variables (i.e., joint angles) and the one-dimensional temporal variable (i.e., time step), and N is the size of the trajectories. Fig. 3-(a) shows the joint motion trajectories, $\mathbf{X} \in \mathbb{R}^{(13+1) \times 578}$, extracted from a complete demonstration in the task for cooking rice.

Before modeling a GMM, the motion trajectories are first transformed to a low-dimensional space by PCA. The reason is as follows: the motion trajectories are segmented as per the Gaussians obtained from the GMM that indicate quasi-linear segments. For better characterizing the non-linear motion

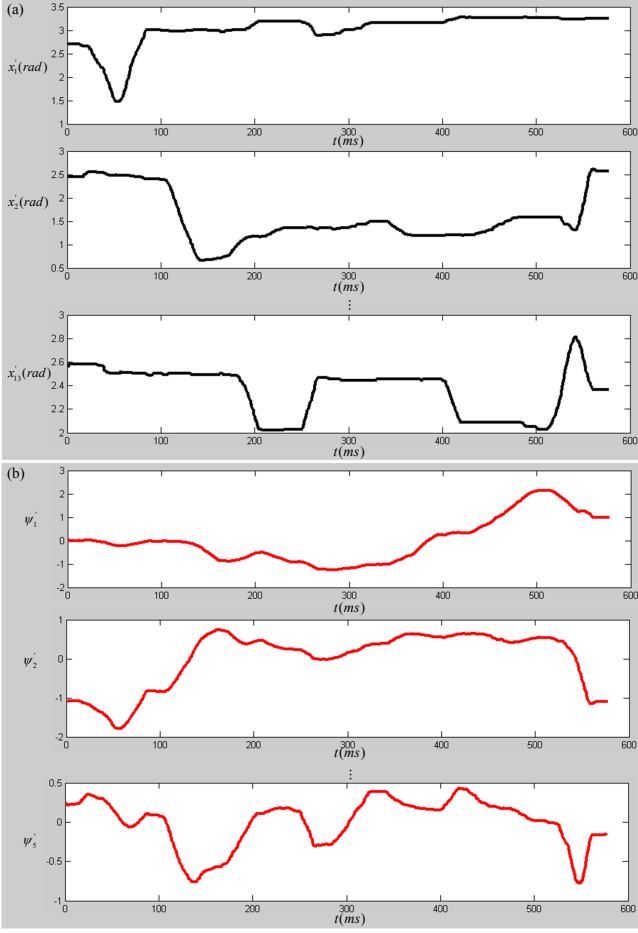


Fig. 3. Motion trajectories in the original space and in the dimensional space reduced by PCA in the task for cooking rice; (a) joint motion trajectories $\mathbf{X} \in \mathbb{R}^{(13+1) \times 578}$ in the original space: Here, (13+1) denotes thirteen joints of the robot and one time step, and 578 is the length of the trajectories. (b) Motion trajectories, $\Psi \in \mathbb{R}^{(5+1) \times 578}$, in dimensional space reduced by PCA: Here, (5+1) denotes a five-dimensional spatial variable in the dimensional space reduced by PCA and a one-dimensional temporal variable, and 578 is the length of the trajectories.

trajectories, the GMM should be modeled containing as many Gaussians as possible without overfitting. The BIC algorithm is a method for resolving the overfitting problem based on the criteria of minimum description length. The GMM is therefore well-fitted using the BIC and EM algorithms. In the BIC algorithm, however, the number of Gaussians depends on the dimensionality of the motion trajectories. The GMM tends to contain many Gaussians in the low-dimensional spaces reduced by PCA under the assumption that essential motion trajectory information is not eliminated. The motion trajectories are better represented by the GMM estimated in the low-dimensional space as more Gaussians are used than in the original space; when the GMM is estimated using expectation maximization (EM) algorithm, one well known limitation is that it gets increasingly difficult to converge to appropriate local optima when the dimension of the variables increases [17]. The original trajectories except

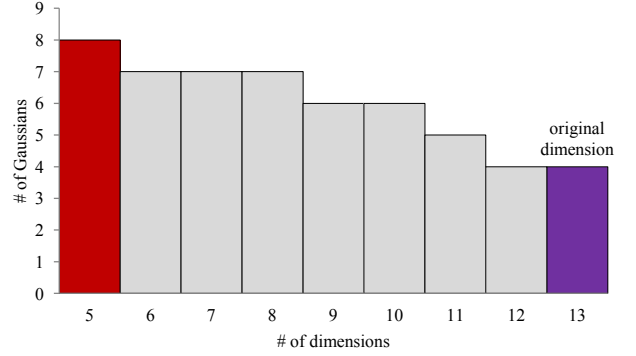


Fig. 4. The number of Gaussians according to dimension of PCA when using BIC in the task for cooking rice: Here, the red box is the maximum number in the dimensions reduced by PCA, and the violet box is the number of Gaussians in original dimensional space.

temporal variable are transformed as

$$\Psi' = \mathbf{A}^T (\mathbf{X}' - \bar{\mathbf{X}}'), \quad (1)$$

where $\mathbf{X}' \in \mathbb{R}^{D \times N}$, $\bar{\mathbf{X}}' \in \mathbb{R}^{D \times N}$, $\mathbf{A} \in \mathbb{R}^{D \times D'}$, and $\Psi' \in \mathbb{R}^{D' \times N}$ refer, respectively, to the original joint trajectories, the means of the original joint trajectories, the transformation matrix of PCA, and the trajectories in the reduced dimensional space. Here, D' denotes the D' -dimensional variables transformed by PCA. Here, the superscript T denotes the transpose operation.

To obtain the maximum number of basis skills from well-fitted GMM without overfitting, the dimension of transformation matrix \mathbf{A} is adaptively selected into the ranges from 0.9 to 1.0 of the sums of the eigenvalues while automatically determining the number of Gaussians using BIC. This is because the GMM should be modeled using as many Gaussians as possible in the ranges in which the essential information of the motion trajectories is not eliminated. Fig. 4 shows the number of Gaussians according to dimension of transformation matrix \mathbf{A} when using BIC. In the original trajectories for cooking rice, the thirteen-dimensional joint trajectories are transformed to the five-dimensional trajectories, $\Psi \in \mathbb{R}^{(5+1) \times 578}$, (determined by the sum of the eigenvalues as 0.94) as shown in Fig. 3-(b).

The GMM is estimated using the motion trajectories, $\Psi \in \mathbb{R}^{(D'+1) \times N}$, in the reduced dimensional space based on EM algorithm. The GMM is defined as

$$P(\Psi) = \sum_{i=1}^K w_i \cdot N(\Psi | \mu_i, \Sigma_i), \quad (2)$$

where w_i , μ_i , and Σ_i refer to the priors, the means, and the covariances of the i^{th} Gaussian, respectively. Here, the number K of Gaussians is determined by BIC. The score function of BIC is defined as

$$S_{BIC} = -2 \cdot \log L + n_p \cdot \log(N), \quad (3)$$

where L is the likelihood, n_p is the number of free parameters of the GMM, and N is the amount of trajectories. Here, n_p

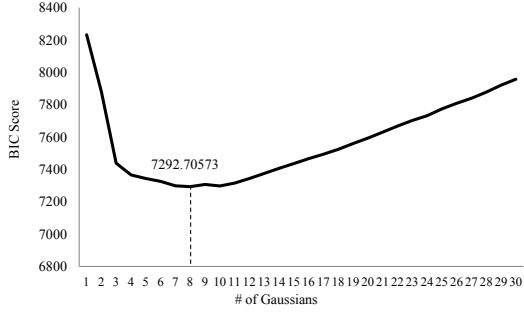


Fig. 5. Illustration of BIC scores on the number of Gaussians of the GMM when using the motion trajectories reduced by PCA in the task for cooking rice.

is defined as

$$n_p = \left(\frac{K \cdot (D' + 1)(D' + 2)}{2} \right) + (K - 1) + (K \cdot (D' + 1)), \quad (4)$$

where K is the number of Gaussians and $(D' + 1)$ is the number of variables. Fig. 5 shows the BIC scores when using the motion trajectories, $\Psi \in \mathbb{R}^{(5+1) \times 578}$, of the task for cooking rice. The GMM is estimated using $K = 8$ according to the result of BIC as shown in Fig. 6-(a).

B. Learning Basis Skills based on Segmentation Points

From the GMM, the segmentation points are detected in temporally overlapping points in-between two consecutive Gaussians as shown in Fig. 6-(a). The points are detected in all points intersected along the time component of the GMM. Before detecting the intersections, the means and covariances of the GMM are first divided into temporal and spatial components. In this context, the mean and covariance matrices of the i^{th} Gaussian are represented as

$$\mu_i = \{\mu_{i,t}, \mu_{i,\Psi'}\}, \quad (5)$$

and

$$\Sigma_i = \begin{pmatrix} \Sigma_{i,t} & \Sigma_{i,t\Psi'} \\ \Sigma_{i,\Psi't} & \Sigma_{i,\Psi'} \end{pmatrix} \quad (6)$$

where t and Ψ' refer to the one-dimensional temporal variable and the D' -dimensional spatial variable in $(D' + 1)$ -dimensional variable Ψ . All intersections in-between Gaussians are extracted by estimating weights of Gaussians along time component. Estimating the weights is defined as

$$h_i(t) = \frac{w_i N(t; \mu_{i,t}, \Sigma_{i,t})}{\sum_{k=1}^K w_k N(t; \mu_{k,t}, \Sigma_{k,t})}, \quad (7)$$

where i and K refer to the indices of Gaussians and the total number of Gaussians. Fig. 7 shows the weights $h_i(t)$ estimated along time component of the GMM and all intersections in the task for cooking rice. The segmentation points are 85, 146, 211, 284, 357, 428, and 509 steps in the entire 579 steps of continuous trajectories.

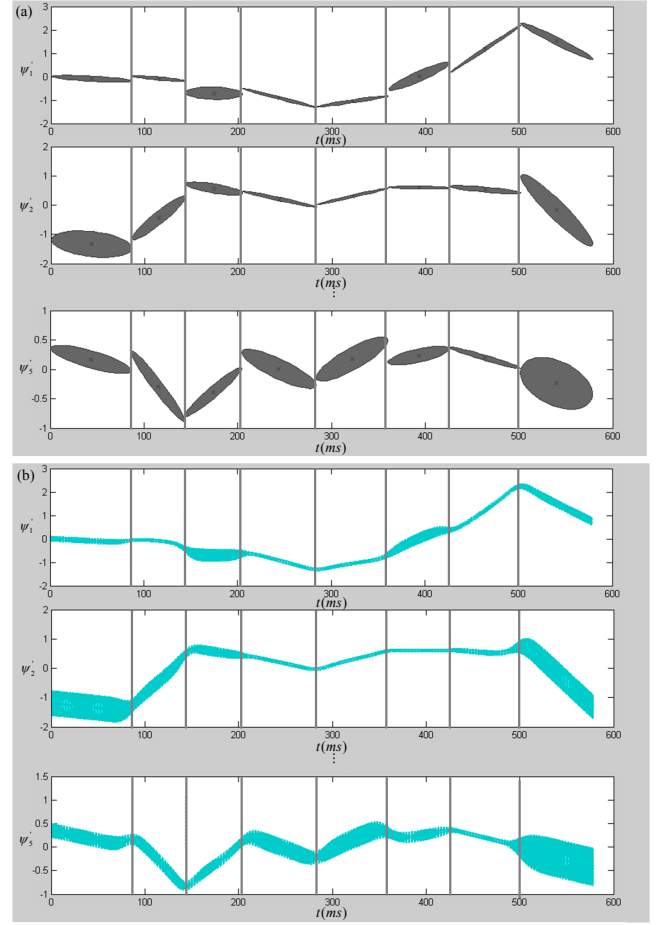


Fig. 6. Gaussian mixture model that consists of the eight Gaussians and temporally overlapping points in-between consecutive Gaussians in the task for cooking rice. Here, the black lines indicate the temporally overlapping points.

To achieve a manipulation task, continuous trajectories reproducible on the robot should be ensured in reorganization of all basis skills. For this, the parameters of (5) and (6) in the GMM should be able to be retrieved using queries of the temporal variable by GMR. The continuous trajectories retrieved by GMR are calculated as

$$\mu_{\Psi'}(t) = \sum_{i=1}^K h_i(t) (\mu_{i,\Psi'} + \Sigma_{i,\Psi't} \Sigma_{i,t}^{-1} (t - \mu_{i,t})), \quad (8)$$

and

$$\Sigma_{\Psi'}(t) = \sum_{i=1}^K h_i^2(t) (\Sigma_{i,\Psi'} - \Sigma_{i,\Psi't} \Sigma_{i,t}^{-1} \Sigma_{i,t\Psi'}), \quad (9)$$

where $h_i(t)$ is defined in (7). The motion trajectories are retrieved by GMR when organizing basis skills using those time durations as shown in Fig. 6-(b). Fig. 8 shows the set of eight basis skills embedded in a task for cooking rice. The labels of eight basis skills autonomously segmented by our segmentation process can be assigned as shown in Fig. 8. We can see that the basis skills are learned from joint motion trajectories through our autonomous segmentation process.

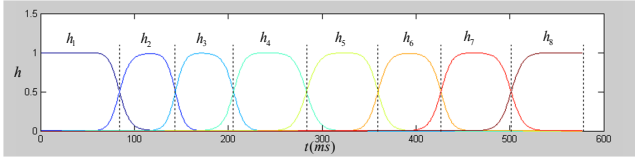


Fig. 7. Weights $h_i(t)$ estimated along time component of the GMM and the seven intersections by the weights in the task for cooking rice. Here, the dotted lines indicate the seven intersections by the weights.

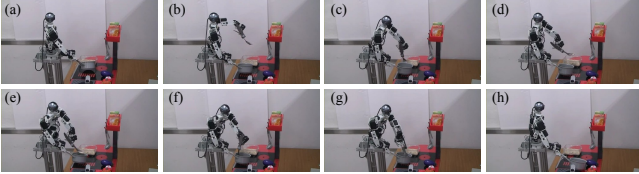


Fig. 8. Labels of eight basis skills embedded in the task for cooking rice: (a) [LiftingPot], (b) [LiftingSpoon], (c) [ApproachingRiceBowl], (d) [ScoopingRice], (e) [DeliveringRice], (f) [PouringRice], (g) [StirringRice], and (h) [PuttingOnStove].

The basis skills are exploited in alternate solutions of a manipulation task by reorganizing them according to different sequences. To sequentially rearrange the basis skills, in the alternative trajectories, the mean values of time component in the basis skills are replaced by those time durations (e.g., the task of three scooping, three delivering, and two stirring in a row) (See Fig. 14).

III. EXPERIMENTS AND EVALUATIONS

For quantitative analysis of our autonomous segmentation process, we evaluated our segmentation produces using motion data from the kitchen motion database presented in [13]. The motion data are the continuous positional trajectories extracted from demonstrations for setting the table in the kitchen of TUM. The motion trajectories are (84+1)-dimensional motion capture data extracted at 25 Hz. Here, the 85-dimensions indicate the 84 spatial variables (= 28 body parts \times (x, y, z) positions) and the one temporal variable. Our segmentation method was evaluated using four episodes [ID0-0], [ID0-2], [ID0-11], and [ID0-12] as shown in Fig. 9. Here, episode [ID0-0] was extracted from a demonstration in which a human separately transports each object like a robot. Episode [ID0-2] was also extracted from a complete demonstration in which the same human takes several objects at once like a human. For increasing the reliability, episodes [ID0-11] and [ID0-12] were extracted from demonstrations similar with [ID0-0] and [ID0-2] by another human. The episodes [ID0-0], [ID0-2], [ID0-11], and [ID0-12] have the lengths of 1241, 957, 1811, and 1136 frames.

In these motion data, especially, there were labels assigned by Gall *et al.* [18]. The labels were assigned by pre-learned models of manually segmented basis skills as shown in Fig. 10. For comparing our results, the labels were assigned after the motion data are autonomously segmented using our autonomous segmentation process. Our labels are physically

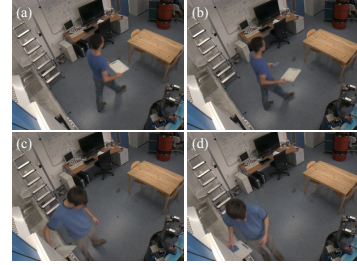


Fig. 9. Four episodes demonstrated for setting the table in the kitchen of TUM: (a) [ID0-0], (b) [ID0-2], (c) [ID0-11], and (d) [ID0-12].

Pre-learned labels of basis skills segmented by Gall <i>et al.</i>		Labels of basis skills autonomously segmented by our autonomous segmentation process	
Left Arm & Right Arm	Trunk	Left Arm & Right Arm	Trunk
CarryingWhileLocomoting (CWL)	StandingStill (Standing)	CarryingWhileLocomoting (CWL)	StandingStill (Standing)
Reaching	HumanWalkingProcess (Walking)	Reaching	WalkingForward (F_Walking)
TakingSomething (Taking)		TakingSomething (Taking)	WalkingBackward (B_Walking)
OpeningADoor (Opening)		StretchingToOpenDoor (O_Stretching)	WalkingSideways (S_Walking)
		FoldingToOpenDoor (O_Folding)	
LoweringAnObject (Lowering)		LoweringAnObject (Lowering)	TurningUsingLeftFoot (L_Turning)
ClosingADoor (Closing)		StretchingToCloseDoor (C_Stretching)	TurningUsingRightFoot (R_Turning)
		FoldingToCloseDoor (C_Folding)	
ReleasingGraspOfSomething (Releasing)		ReleasingGraspOfSomething (Releasing)	

Fig. 10. Labels of pre-learned models segmented by Gall *et al.* and labels assigned of basis skills autonomously segmented by our segmentation method.

identified as nine labels of two arms and six labels of trunk. Although these labels are greatly similar with the labels of Gall *et al.*, [OpeningADoor], [ClosingADoor], and [HumanWalkingProcess] in the labels of Gall *et al.* are segmented in detail as follows: [OpeningADoor] and [ClosingADoor] were segmented as each two labels for [StretchingArm] and [FoldingArm] to open or close a door. The label [HumanWalkingProcess] of Gall *et al.* is segmented as five labels of [WalkingForward], [WalkingBackward], [WalkingSideways], [TurningUsingLeftFoot], and [TurningUsingRightFoot].

In the motion data of episode [ID0-2], the 52 basis skills are autonomously estimated by our autonomous segmentation process. The basis skills are physically similar with 56 segments of Gall *et al.*. In details, 44 of 49 (=56-7) basis skills are the similar when timing differences of starting and ending points in the basis skills are allowed the averaged 3.24 frames (i.e., 0.13 s). Here, seven basis skills in which the time durations were 1~2 frames were removed from the 56 basis skills of Gall *et al.*, because it is difficult to find physical meaning in such basis skills. Even though 89.8 percentages of all basis skills are similar, even dissimilar basis skills contain physical meaning. The dissimilar basis skills are usually due to more segmented labels of [OpeningADoor], [ClosingADoor], and [HumanWalkingProcess].

Figs. 11-12 show the observed data from the procedures for segmenting the motion trajectories of the episode [ID0-2]. The (84+1)-dimensional trajectories are transformed as

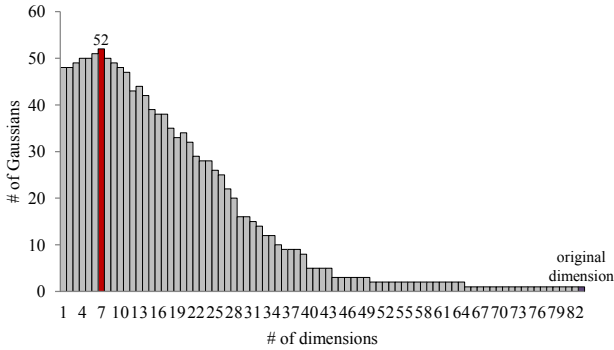


Fig. 11. The number of Gaussians according to dimension of PCA when using BIC in the episode [ID0-2]: Here, the red box is the maximum number of all dimensions of PCA, and the violet box is the number of Gaussians in original dimensional space.

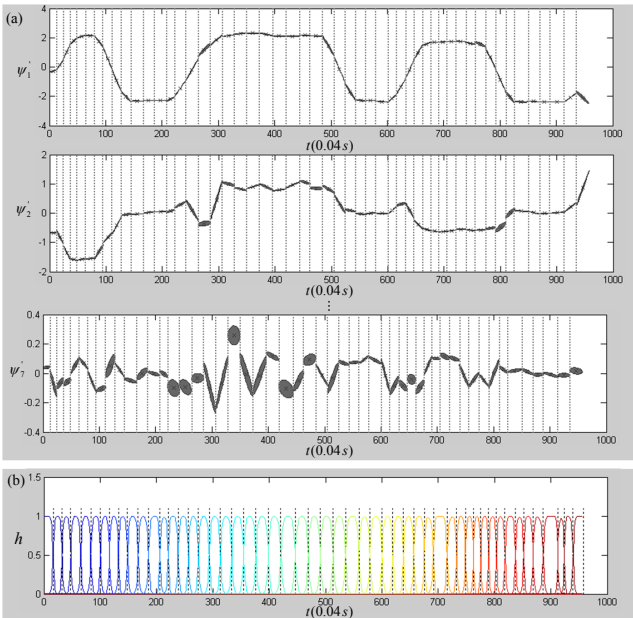


Fig. 12. GMM and weights estimated in the episode [ID0-2]: (a) GMM that consists of 52 Gaussians, and (b) weight $h_i(t)$ are estimated along the time component of the GMM. Here, the dotted lines indicate the 51 intersections by the weights.

(7+1)-dimensional trajectories by PCA. The dimension of PCA is determined to the seven-dimensional space (here, the sum of the eigenvalues is 0.98) for selecting the maximum number of Gaussians as shown in Fig. 11. The GMM is then estimated by EM algorithm in which the GMM consists of 52 Gaussians as shown in Fig. 12-(a). The segmentation points are detected in temporally overlapping points in-between 52 Gaussians by the weights estimated along the time component of the GMM. Fig. 12-(b) shows all intersections in-between consecutive Gaussians extracted by estimating the weights of Gaussians along the time component.

Fig. 13 shows the segmentation results obtained based on our segmentation process when using motion data of episodes [ID0-0], [ID0-2], [ID0-11], and [ID0-12]. In the case of episode [ID0-0], the motion trajectories are transformed to

	ID0-0	ID0-2	ID0-11	ID0-12
# of dimensions reduced by PCA	8	7	9	8
# of basis skills segmented by our method	72	52	86	51
# of basis skills segmented by Gall <i>et al.</i> (# of basis skills that the lengths are 1~2 frames)	100 (26)	56 (7)	97 (12)	54 (5)
# of similar basis skills	67	44	76	45
Similarity of basis skills	90.54%	89.79%	89.41%	91.84%

Fig. 13. Segmentation results obtained based on our segmentation process when using the motion data of episodes [ID0-0], [ID0-11], and [ID0-12].

(8+1)-dimensional space (here, the sum of the eigenvalues is 0.995) by PCA. The GMM is then estimated by BIC and EM algorithms using the motion trajectories in the dimensional space reduced by PCA. Here, the GMM is modeled by 72 Gaussians. When comparing similarity of segmentation produces by the same criteria of episode [ID0-2], 67 of total 74 (=100-26) basis skills by Gall *et al.* are found to be similar (i.e. 90.5%) to the basis skills given by our segmentation process. In 100 basis skills detected by Gall *et al.*, 26 basis skills were removed for reasonable comparison, and dissimilar basis skills are usually due to more segmented labels of [OpeningADoor], [ClosingADoor], and [HumanWalkingProcess]. In the cases of episodes [ID0-11] and [ID0-12], the motion trajectories are transformed to each (9+1)-dimensional space (here, the sum of the eigenvalues is 0.985) and (8+1)-dimensional space (here, the sum of the eigenvalues is 0.98) by PCA. The GMMs are modeled by each 86 Gaussians and 51 Gaussians based on BIC and EM algorithms. 76 of total 85 basis skills (in 97 basis skills by Gall *et al.*, 12 basis skills were removed) are similar (about 89.4%) to the basis skills given by our process in the case of episode [ID0-11], and 45 out of 49 (in 54 basis skills by Gall *et al.*, 5 basis skills were removed) are similar (about 91.8%) to the basis skills given by our process in the episode [ID0-12]. In these all episodes, the number of basis skills produced by our autonomous segmentation method is low than the basis skills by Gall *et al.*'s method. It is because there are many segments with 1~2 frames (i.e. 0.04~0.08 sec) in the basis skills of Gall *et al.* for which it is not easy to find physical meaning.

IV. DISCUSSION

We have segmented joint motion trajectories of a cooking task and positional trajectories of human's body parts extracted from four demonstrations based on our autonomous segmentation process. The segmentation processes are executed without adjusting parameters according to the types of variables and tasks. That is, there are no constraints to adapt to such variations. Although we have automatically determined the number of Gaussians by BIC, it can adjust the segmentation resolution using the number of basis skills as priors if a human teacher knows the number of basis skills embedded in a task. That is, the segmentation resolution can

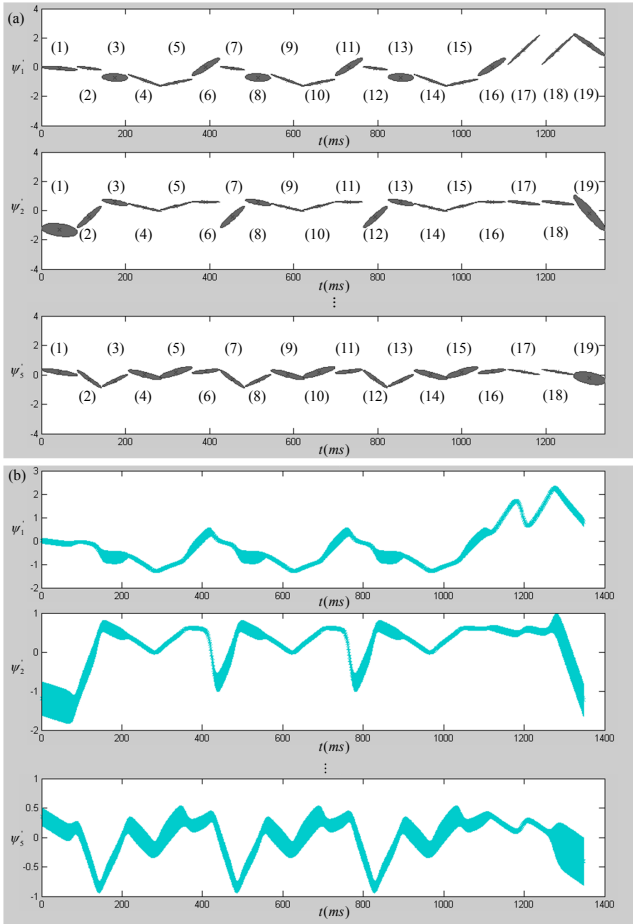


Fig. 14. Sequence of the basis skills and continuous motion trajectories acquired by reorganizing eight basis skills of rice cooking task for alternative solution (i.e. the task of three scooping, three delivering, and two stirring in a row): (a) Sequence of the basis skills acquired by reorganizing eight basis skills of the rice cooking task: Here, the basis skills correspond respectively to (1) [LiftingPot] (time duration: 86); (2), (7), (12) [LiftingSpoon] (time duration: 56); (3), (8), (13) [ApproachingRiceBowl] (time duration: 66); (4), (9), (14) [ScoopingRice] (time duration: 75); (5), (10), (15) [DeliveringRice] (time duration: 73); (6), (11), (16) [PouringRice] (time duration: 70); (17), (18) [StirringRice] (time duration: 81); and (19) [PuttingOnStove] (time duration: 71). (b) Continuous motion trajectories retrieved by the GMR process when using the sequence in (a).

be simply adjusted by adjusting the number of Gaussians in the GMM. Whereas we restricted ourselves to kinematic data in the experiments, the approach proposed extends to all kinds of sensor data such as force, torque, or signals of end-effectors.

It can take lots of time to determine the dimension of PCA for extracting the maximum number of basis skills (especially, in the cases of TUM data), since it should be verified in many dimensions that can be reduced by PCA. Even though the segmentation resolution is usually higher in the lower dimension of PCA, it can be sometimes lower in the lower dimension. It is because necessary information of the motion trajectories can be removed in the low-dimensional space reduced by PCA. In most cases, therefore, the dimension of PCA usually needs to be checked between

the dimensions in which the sum of the eigenvalues ranges from 0.9 to 1.0 based on our various experiments.

In this paper, the motion trajectories are extracted from a complete demonstration. However, our approach can be applied in the set of motion trajectories extracted from multiple demonstrations. In such a case, the GMM is first estimated using the set of motion trajectories after temporally aligning the motion trajectories using dynamic time warping (DTW), and the basis skills are estimated from the GMM. Even though the DTW algorithm should additionally be used to deal with such multiple demonstrations, it does not need to tune or predefine any constraints according to the types of tasks and/or variables.

Finally, task skills for humanoid robots are very different than task skills in industrial settings. Since humanoids are developed to share home and office environments, the range of possible tasks cannot be determined in advance, and the robot needs to reuse previously learned skills in a flexible and efficient manner without having to relearn everything from scratch. In a robot cooking scenario, the sequential organization of basis skills depends on many unpredictable factors related to the home environment setup, the kitchen tools, and cooking ingredients being used by the robot. The reorganization of basis skills also depend on human factors such as the user introducing perturbation by sharing the same space of the robot, and regarding the user preferences and variants of cooking recipes. For example, if the user is hungry, the quantity can be increased by repeating some basis skills. A food ingredient can also be removed the recipe if the user does not like it. Therefore, the reorganization reflecting such human factors is also another issue to be solved for reasonably reusing the basis skills.

V. CONCLUSIONS AND FUTURE WORKS

In this paper, we have proposed a method to autonomously segment the motion trajectories from a complete demonstration for learning basis skills embedded in a task. For segmentation, the motion trajectories are modeled by a GMM, since segmentation points are to be extracted by segmenting the Gaussians of the GMM. To obtain the basis skills as many as possible from a task, the motion trajectories are first transformed to a low-dimensional space reduced by PCA when the number of Gaussians is automatically determined by BIC. After estimating the GMM using the motion trajectories transformed by PCA, the segmentation points are detected in points intersected along the time component of the GMM. The intersections indicate temporally overlapping points in-between two consecutive Gaussians in the GMM. Finally, continuous motion trajectories are retrieved by GMR when sequentially reorganizing basis skills using those time durations.

To quantitatively analyze our segmentation process, we have evaluated our method using four episodes of activities from the kitchen motion database. When timing differences of starting and ending points in the basis skills are allowed the averaged 3~4 frames, these segmentation results are

similar with the basis skills of Gall *et al.* in each 90.54%, 89.79%, 89.41%, and 91.84% of entire segments. Moreover, the dissimilar basis skills can be easily explained by the difference of segmentation granularity that can be considered in motions such as opening, closing, and walking. There are certain advantages to our segmentation method. First, the core process does not require tuned or predefined parameters according to the types of variables and tasks. Second, the process can learn basis skills in which non-linear motion trajectories can be better characterized than using original motion trajectories, since the GMM is modeled using more Gaussians by investigating the reduced dimensional spaces. Finally, the proposed approach can retrieve the novel motion trajectories reproducible on robots based on the GMR process by temporally rearranging the basis skills when the sequences of basis skills are provided by the existing planning methods or human experts.

In future work, we intend to apply our scheme to various types of continuous trajectories. Furthermore, we shall extend the basis skills so as to use reorganization strategy using pre- and post-conditions based on surrounding environment. In this regard, segmentation points can play a key role towards determining these pre- and post-conditions.

REFERENCES

- [1] E. Billing, T. Hellstrom, and L. Janlert, "Behavior recognition for learning from demonstration," in *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*, 2010, pp. 866–872.
- [2] B. Cohen, S. Chitta, and M. Likhachev, "Search-based planning for manipulation with motion primitives," in *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*, 2010, pp. 2902–2908.
- [3] D. Bentivegna and C. Atkeson, "Learning from observation using primitives," in *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*, 2001, pp. 1988–1993.
- [4] M. Nicolescu and M. Mataric, "Natural methods for robot task learning: Instructive demonstrations, generalization and practice," in *Proceedings of the Second International Joint Conference on Autonomous Agents and Multiagent Systems*, 2003, pp. 241–248.
- [5] N. Nejati, P. Langley, and T. Konik, "Learning hierarchical task networks by observation," in *Proceedings of the 23rd International Conference on Machine Learning*, 2006, pp. 665–672.
- [6] E. Drumwright, O. Jenkins, and M. Mataric, "Exemplar-based primitives for humanoid movement classification and control," in *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*, 2004, pp. 140–145.
- [7] D. Kulic, W. Takano, and Y. Nakamura, "Online segmentation and clustering from continuous observation of whole body motions," *IEEE Transactions on Robotics*, vol. 25, no. 5, pp. 1158–1166, 2009.
- [8] E. Gribovskaya and A. Billard, "Combining dynamical systems control and programming by demonstration for teaching discrete bimanual coordination tasks to a humanoid robot," in *Proceedings of ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2008.
- [9] T. Asfour, F. Gyrfas, P. Azad, and R. Dillmann, "Imitation learning of dual-arm manipulation tasks in humanoid robots," in *Proceedings of 6th IEEE-RAS International Conference on Humanoid Robots*, 2006, pp. 40–47.
- [10] K. Hauser, V. Ng-Thow-Hing, and H. Gonzalez-Baños, "Multi-modal motion planning for a humanoid robot manipulation task," in *Robotics Research*, pp. 307–317, 2011.
- [11] L. Montesano, M. Lopes, A. Bernardino, and J. Santos-Victor, "Learning object affordances: From sensory-motor coordination to imitation," in *IEEE Transactions on Robotics*, vol. 24, no. 1, pp. 15–26, 2008.
- [12] S. Lee and I. Suh, "Motivation-based dependable behavior selection using probabilistic affordance," in *Advanced Robotics*, vol. 26, no. 8-9, pp. 897–921, 2012.
- [13] M. Tenorth, J. Bandouch, and M. Beetz, "The TUM kitchen data set of everyday manipulation activities for motion tracking and action recognition," in *Proceedings of IEEE 12th International Conference on Computer Vision Workshops (ICCV Workshops)*, 2009, pp. 1089–1096.
- [14] O. Sigaud, C. Salaün, and V. Padois, "On-line regression algorithms for learning mechanical models of robots: a survey," in *Robotics and Autonomous Systems*, 2011, pp. 1115–1129.
- [15] S. Calinon, F. Guenter, and A. Billard, "On learning, representing, and generalizing a task in a humanoid robot," in *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 37, no. 2, pp. 286–298, 2007.
- [16] B. Akgun, M. Cakmak, J. Yoo, and A. Thomaz, "Trajectories and keyframes for kinesthetic teaching: A human-robot interaction perspective," in *Proceedings of the ACM/IEEE international conference on Human-robot interaction (HRI)*, 2012, pp. 391–398.
- [17] S. Eddy *et al.*, "Multiple alignment using hidden Markov models," in *Proceedings of the Third International Conference on Intelligent Systems for Molecular Biology*, vol. 3, 1995, pp. 114–120.
- [18] J. Gall, A. Yao, and L. Van Gool, "2D action recognition serves 3D human pose estimation," in *Computer Vision—ECCV 2010*, pp. 425–438, 2010.