

# Comparative Study of Representations for Segmentation of Whole Body Human Motion Data

Dana Kulić and Yoshihiko Nakamura

**Abstract**—In previous work, the authors have been developing a stochastic model based approach for on-line segmentation of whole body human motion patterns during human motion observation and learning, using a simplified kinematic model of the human body. In this paper, we extend the proposed approach to larger, more realistic kinematic models, which can better represent a larger variety of human motions. These larger models may include spherical in addition to revolute joints. We examine the effects on segmentation performance due to motion representation choice, and compare the segmentation efficacy when Cartesian or joint angle data is used. The approach is tested on whole body human motion data modeled with a 42DoF kinematic model. The results indicate that Cartesian data seems to correspond most closely to the human evaluation of segment points. The experiments also demonstrate the efficacy of the segmentation approach for large kinematic models and a variety of human motions.

## I. INTRODUCTION

As robots enter the human environment, they will need to be able to recognize human actions during continuous on-line observation of human motion. For humanoid robots, recognizing and modeling human motion primitives can also be used for imitation learning [1], [2], [3]. In addition to applications to humanoid robots, the ability to recognize human actions during on-line observation can be useful in a wide variety of applications, including activity recognition for interaction, activity recognition for surveillance and security, as well as motion modeling for sports training, rehabilitation and health monitoring.

A key requirement of on-line recognition of motion primitives is automated motion segmentation, i.e., the automatic determination of the start and end points of motion primitives. In previous work [4] we have been developing an approach for autonomous segmentation of the observed continuous data stream into whole body motion primitives, based on a stochastic segmentation algorithm [5]. In the proposed approach, a hidden Markov model is used to represent the incoming data sequence, where each model state represents the probability density estimate over a window of the data. Based on the assumption that data belonging to the same motion primitive will have the same underlying distribution, the segmentation is implemented by finding the optimum state sequence over the developed model. In

previous work, a simplified 20DoF model of the human body, consisting of only revolute joint angles, was used to perform the segmentation. This simplified model was based on a specific robot kinematic model, simplifying the process of re-targeting the motion on a humanoid robot. However, when modeling human motion, especially in applications such as rehabilitation or activity recognition, such a simplified model may not be adequate to capture the important properties of human motion. Research focused on human motion animation, analysis and activity recognition typically uses more realistic models, with a higher number of degrees of freedom and including spherical joints. The larger kinematic model can better represent human motion, including motions such as torso and arm twisting motions. In this paper, we extend our previous method to handle larger kinematic models, and investigate how the choice of kinematic representation can affect segmentation accuracy for human motion data.

### A. Related Work

Segmentation of continuous human motion data has received significant attention in the robotics literature [6], [7], [8], [7], [8], where attention has focused on extracting motion segments for learning from imitation. There has also been extensive research in the graphics community [9], [10], [11], [12], [13], where research has focused on motion synthesis from motion capture data for animation characters. Existing data segmentation algorithms can be classified based on whether previous knowledge of the motion primitives to be segmented is required.

The first category of algorithms consider segmentation without prior information about the motion primitives, i.e., unsupervised segmentation. In these algorithms, some assumption must be made about the underlying structure of the data at a segmentation point. Earlier approaches proposed segmenting based on the velocity properties of the joint angle vector [6], [7], [8]. In Pomplun and Mataric [6], a segment is recognized when the root mean square (RMS) value of the joint velocities falls below a certain threshold, i.e., assuming that there will be a pause in the motion between motion primitives. In Fod et al. [7], it is assumed that there is a change in the direction of movement accompanying a change between motion primitives. Therefore, a segmentation point is recognized when a Zero Velocity Crossing (ZVC) is detected in the joint angle data, in a sufficient number of dimensions. However, with all the velocity based approaches, typically only a subset of the human body is considered, such as one arm [6], [7], or an arm and torso [8] as it becomes more difficult to tune the algorithm as the number of joints

Dana Kulić was with the Department of Mechano-Informatics, University of Tokyo, and is currently with the Department of Electrical and Computer Engineering, University of Waterloo, 200 University Avenue West, Waterloo, Ontario, N2L 3G1, Canada. Yoshihiko Nakamura is with the Department of Mechano-Informatics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, 113-8656 Tokyo, Japan. Corresponding Author: Dana Kulić [dkulic@ece.uwaterloo.ca](mailto:dkulic@ece.uwaterloo.ca)

increases.

Kulić et al. [14] propose an approach to using velocity information for segmentation without considering individual degrees of freedom. In the proposed approach, segmentation is based on visual information from a monocular camera only, without considering joint angle data or the kinematic model. In this approach, changes in direction of optical flow are used to indicate segment points.

Koenig and Matarić [15] develop a segmentation algorithm based on the variance of the feature data. The algorithm searches for a set of segment points which minimize a cost function of the data variance. In a related approach, Kohlmorgen and Lemm [5] describe a system for automatic on-line segmentation of time series data, based on the assumption that data from the same motion primitive will belong to the same underlying distribution. The incoming data is described as a series of probability density functions, which are formulated as the states of a Hidden Markov Model (HMM), and a minimum cost path is found among the states using an accelerated version of the Viterbi algorithm. Nakamura and colleagues [16], [17], [4] apply this approach to human motion capture data. However, in this case, a simplified 20DoF model of the human body, consisting only of single DoF revolute joints is used.

The second class of segmentation algorithms frames the problem as a supervised segmentation, where motion primitives are specified by the designer a-priori, and segmentation is based on the comparison between the known motions and the incoming data. Ilg et al. [18] use dynamic programming to find the best match between prototypical motion primitives and an observed sequence based on key features consisting of zero velocity points in key dimensions. In Takano and Nakamura [19], [20] the known motion primitives are encoded via short HMMs. Segmentation points are then decided based on the error between the motion data predicted by the HMM and the actual observed motion data. If the error increases above a certain threshold, a segment point is declared. Pardowitz et al. [21] propose a neural net based algorithm for motion segmentation. In the proposed approach, a Competitive Layer Model (CLM) type network is used to perform the segmentation, and the network is then trained on labeled data to find the optimum weight values of the network. The proposed approach is tested on a data set consisting of arm reaching movements.

### B. Proposed Approach

In previous research [4], [3], we have been developing algorithms for on-line segmenting and clustering of motion primitives, based on a simplified kinematic model of the human body, which can easily be re-targeted for imitating human movements with a humanoid robot. However, the simplified kinematic model may not be adequate for capturing the full range of human motions, and therefore may not be useful for segmentation where movements which are not captured by the model are performed. In this paper, we extend the previously developed approach to larger kinematic models, and analyze the performance on a variety of human

motions. We also consider the effect of representation choice on segmentation performance for larger kinematic models. Section II provides a brief review of the basic segmentation approach developed in our previous work [4], [3], Section III discusses the additional considerations and algorithm modifications required for extending the approach to larger kinematic models, Section IV provides the experimental validation of the proposed approach, while Section V concludes the paper and provides directions for future work.

## II. UNSUPERVISED PROBABILISTIC SEGMENTATION

This section provides an overview of our previously developed probabilistic segmentation algorithm [4], [3]. We briefly introduce the approach here to provide sufficient understanding for the novel components of the algorithm developed in Section III; for a detailed analysis of the previously developed algorithm performance, the reader is referred to [4], [3]. In the original approach, a modified version of the Kohlmorgen and Lemm segmentation algorithm [5] is applied to segment the continuous time series data into motion primitive segments [4]. After embedding the incoming data stream into a higher-dimensional space, the density distribution of the incoming embedded data is estimated over a sliding window of length  $W$ , via multivariate Gaussian kernels, centered on the data points in the window  $\{\vec{x}_{t-w}\}_{w=0}^{W-1}$ .

$$p_t(\mathbf{x}) = \frac{1}{W} \sum_{w=0}^{W-1} \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{\|\mathbf{x} - \vec{x}_{t-w}\|^2}{2\sigma^2}\right), \quad (1)$$

where  $\sigma$  is a smoothing parameter calculated proportional to the mean distance between each  $\vec{x}_t$  and its  $n$  nearest neighbors.

As more data are observed, the distance between successive data windows can be calculated based on the integrated square error between the two datasets as represented by the probability density functions. This distance can be calculated analytically in the case of mixtures of Gaussian density functions. Here, we use the heuristic integrated square error distance, as suggested by Kohlmorgen and Lemm [5], alternately, the Kullback-Leibler distance between the distributions could also be used. The segmentation analysis is carried out by defining a Hidden Markov Model over a set  $S$  of sliding windows. Each window corresponds to a state of the HMM. For each state, the observation probability distribution is defined as:

$$p(p_t(\mathbf{x})|s) = \frac{1}{\sqrt{2\pi\zeta}} \exp\left(-\frac{d(p_s(\mathbf{x}), p_t(\mathbf{x}))}{2\zeta^2}\right), \quad (2)$$

where  $p(p_t(\mathbf{x})|s)$  is the probability of observing the window represented by  $p_t(\mathbf{x})$  in state  $s$ , and  $d$  is the distance function based on the integrated square error. The initial state distribution is given by the uniform distribution, and the state transition matrix is designed such that transitions to the same state are  $k$  times more likely than transitions to any of the other states.

$$a_{ij} = \begin{cases} \frac{k}{k+N-1} & \text{if } i = j; \\ \frac{1}{k+N-1} & \text{if } i \neq j \end{cases} \quad (3)$$

where  $N$  is the number of states of the HMM. The Viterbi algorithm [22] can then be used to find the optimum state sequence given the current set of observations. The resulting state sequence directly leads to the segmentation result. Time windows which are classified to be in the same state are considered to belong to the same segment. Therefore each state change is considered a segmentation point. An on-line variant of the Viterbi algorithm is also developed [5], which incrementally builds the state path table as each new state is observed, by re-using the estimate of the likelihood and optimal state sequence from the previous time step. Note that this algorithm models the entire segment as a single state (single probability distribution), and is therefore not a generative model that can be used to generate simulations of the derived segments. Once motion segments are extracted, they can be incrementally clustered and modeled with more detailed stochastic models to produce generative models for use during motion synthesis [23].

To prevent the state list from growing to infinity as the number of observed data points increases, Kohlmorgen and Lemm [5] propose removing states following a segment away from that state. However, Janus [17] has found that this approach leads to over-segmenting, as the considered data range becomes too small (on the order of  $5W$ ) and therefore the algorithm becomes more prone to local minima. Instead, Janus propose that the algorithm runs in batch-mode over a larger, fixed number of windows, and that windows be discarded in a FIFO manner. The Janus approach is adopted herein.

### III. EXTENDING TO LARGE DOF KINEMATIC MODELS

The algorithm described in Section II has been applied to continuous human motion data, where the segmented data consisted of joint angle data obtained by a simplified 20DoF kinematic model [16], [4]. The segmentation points produced by the automated algorithm give a fairly good match to the segmentation points indicated by a human observer, achieving approximately 80% recognition accuracy. The 20DoF kinematic model is based on a humanoid robot model, and is used to facilitate re-targeting of the motions to a humanoid robot with the same kinematic structure and number of DoFs. However, for applications focusing on human motion analysis, typically a higher number of DoFs is required to adequately capture the human motion. In this case, typically a 34DoF to 50DoF model would be applied, to capture human-like movements such as twisting at the torso and twists in the arms and wrists. Such larger models introduce many additional elements in the observation vector, potentially affecting segmentation performance. On the other hand, a better kinematic model is better able to capture the characteristics of the human motion, and therefore using a better model could improve the segmentation performance,

if the higher order model can capture motion which is not visible with the reduced model.

In addition to the issue of dimensionality, large kinematic models typically do not use simple revolute joints to model the DoFs; instead, spherical joints are used. The use of spherical joints introduces a choice of representation of the orientation, which could affect segmentation performance.

The orientation of a spherical joint can be represented using the quaternion representation (also known as Euler parameters), Euler angles, or the angle axis representation [24]. The angle axis representation represents the orientation of a spherical joint as a 3D unit vector  $\vec{r}$  and an angle  $\theta$  expressing the rotation about that vector. A quaternion reformulates the angle axis representation to eliminate the non-uniqueness of the angle axis representation.

$$\eta = \cos\left(\frac{\theta}{2}\right) \quad (4)$$

$$\vec{e} = \sin\left(\frac{\theta}{2}\right)\vec{r} \quad (5)$$

Euler angles represent the spherical joint orientation as a sequence of 3 elementary (1DoF) rotations. Euler angles are a minimal representation (the 3DoFs in a spherical joint are represented by 3 linearly independent parameters), while the angle axis and quaternion representations are non minimal, representing the 3DoFs with 4 parameters (a scalar and a vector), which are related by a constraint equation. For the case of the quaternion, the constraint equation is given by:

$$\eta^2 + \epsilon_x^2 + \epsilon_y^2 + \epsilon_z^2 = 1. \quad (6)$$

The quaternion representation is typically favored in human motion animation applications, as it allows for smooth interpolation between poses without any representational singularities. However, due to the non-linear transformation between the joint angles and the quaternion representation, there is no guarantee that segment points in the quaternion representation will still correspond to segment points as observed by the human observer. In addition, the 4 quaternion values are linearly dependent, since they are related to each other through the constraint relationship, thus invalidating the HMM assumption that the elements in the observation vector be linearly independent. Since the quaternion is also a non-minimal representation, using quaternions to represent many joints in the kinematic model will introduce additional elements into the observation vector, thus increasing the size of the segmentation HMM and slowing computation.

The Euler angles have the advantage of minimal representation, and thus linear independence among the values. However, Euler angles suffer from the problem of representational singularities (at those values of the angles where the axes become parallel and thus no longer independent). Due to the large range of human motion, it is difficult to select a Euler angle representation which can ensure that a singular configuration is not encountered during arbitrary motion. A second, numerical issue with Euler angles is the angular representation, where a rotation of 0 degrees (or 0 radians)

is equivalent to 360 degrees (or  $2\pi$  radians) physically, but not numerically. Inverse kinematics solvers which impose a limit on the angle range can introduce a large step change in the joint angle data (for example from 361 degrees to 1 degrees), which represents a smooth motion during playback but appears as a large change in joint angles numerically, and affects the segmentation performance. On the other hand, not imposing a limit on the angle range may result in the drift of the angle values over time, which will reduce the effect on segmentation performance, but would negatively affect subsequent automatic motion recognition or clustering.

A final option for segmentation data input is to use Cartesian position data. If marker-based motion capture is being used, one alternative is to use the marker data directly as the input data, thus avoiding the need to compute inverse kinematics all together. However, this approach is problematic for on-line processing, as markers may not always be visible, so portions of the input vector may be missing or extremely noisy. There may also be many more markers than degrees of freedom, thus significantly increasing the size of the input vector and slowing down computation. A second approach is to use the inverse kinematics solver to compute the positions of the origin of each link frame, and use the link frame origin positions as the segmentation data input.

#### IV. EXPERIMENTS

To examine the effect of a larger kinematic model and the choice of orientational representation, the different representation approaches were tested and compared using a human motion capture data set. The data set was collected in a marker-based motion capture studio, using a set of 34 markers. The marker locations tracked are shown in Figure 1. The subject performed a variety of exercise motions, such as arm raises, squats and bends, for a sequence duration of approximately 3 minutes. The motions were performed in random order. Extracted frames from a portion of the data sequence are shown in Figure 2. The marker data was then converted to joint angle data using on-line inverse kinematics [25] based on a 43DoF model of the human body. The kinematic model is shown in Figure 3, and includes the 6 DoF base body joint, 7 DoFs for each arm and leg, 2 spherical joints representing the torso and a spherical joint representing the neck. For each arm, the shoulder and wrist are modeled as spherical joints, while the elbow joint is a single DoF revolute joint. For each leg, the hip and ankle joints are modeled as spherical joints, while the knee joint is a single DoF revolute joint. Degrees of freedom in the hands are not modeled. This model was selected because it has been found to provide a good tradeoff between model complexity and the ability to adequately represent most daily activities [26].

The inverse kinematics routine [25] outputs spherical angle data in terms of the quaternion representation. The data sequence was then also converted to Euler angle representation and relative Cartesian representation. For the quaternion representation, the quaternion data was verified to eliminate "flips" in representation which can occur due to the fact

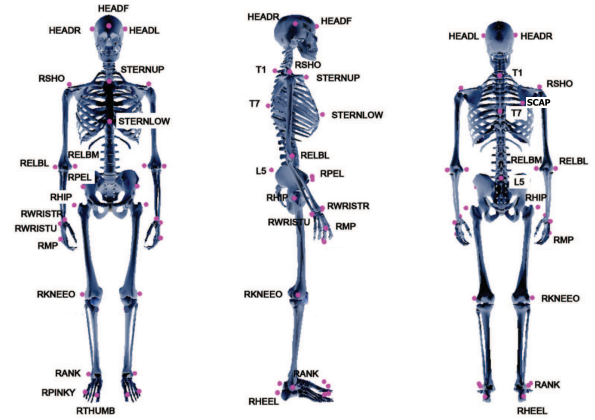


Fig. 1. Marker Setup used for the motion capture experiments

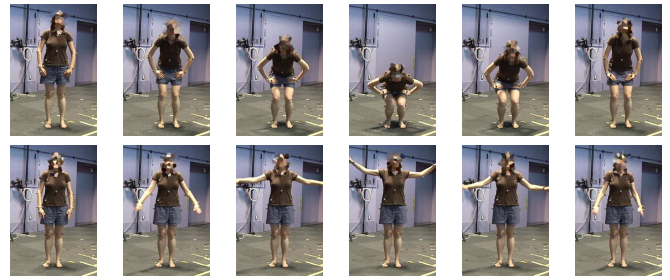


Fig. 2. Video frames taken during the motion capture experiment showing sample motions from the data set. Note that motion capture cameras are not visible from this viewpoint, as the video camera is zoomed in to the demonstrator.

that  $(\theta, \vec{r})$  and  $(-\theta, -\vec{r})$  represents the same rotation. The resulting data consisted of an observation vector with 51 elements. For the Euler angles, the quaternion values were converted to *ZYX* Euler angles, where the initial solution was selected such that the sequence did not include any singularities. The Euler angle observation vector consisted of 41 elements. For the Cartesian representation, the base joint together with the relative  $[x,y,z]$  locations origins of the shoulder, elbow, wrist, hip, knee and ankle joint frames were used, forming an observation vector of 39 elements. It is important to note that, other than the base joint, all the Cartesian data are relative (to the parent frame of each joint), such that the data will remain invariant in the presence of locomotion and translation in the workspace. Each data sequence was processed with the segmentation algorithm as described in Section II, using the same parameter settings as in [4], detailed in Table I.

TABLE I  
ALGORITHM PARAMETERS

Parameter	Value
$k$	1.18
$\varsigma$	0.7
$m$	5
$L$	5
$T$	20

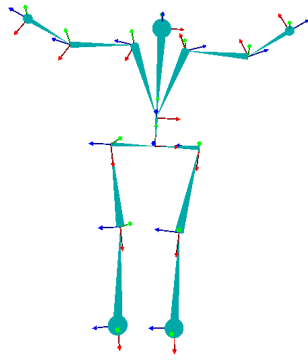


Fig. 3. Kinematic Model used to convert marker positions to joint angle data. Each joint origin frame is indicated by the blue/red/green frame indicator. The elbow and knee joints are single DoF revolute joints. All other joints are spherical.

TABLE II  
COMPARISON OF SEGMENTATION PERFORMANCE

Algorithm	Correct	False Pos	False Neg
Quaternion	74	42	17
Euler Angles	76	34	15
Cartesian	83	17	8

The data was also segmented manually by a human observer, who labeled the start, end and name of each motion performed in the sequence. The automatically segmented data was then compared to the manual labeling. A segment point was considered correct if it occurred within 4 windows of the manually obtained results. A segment point was counted as a false positive, if it occurred in a section where no manual segment point was specified within 4 windows of the given segment point. A false negative was counted if no segmentation point was specified within a 4 window frame of a manually found segmentation point. Table II shows the segmentation results for each representation type.

As can be seen from Table II, using a quaternion representation results in segmentation accuracy of approximately 81%, while the Euler angle representation achieves an accuracy of 83%. The Euler angle result is comparable to the 79% result achieved when using a revolute joint only model [4], which validates the consistency of the approach as the Euler angle representation is equivalent to a sequence of revolute joints. However the best results are achieved when Cartesian representation is used, which achieve a 91% segmentation accuracy while also reducing the number of false positives. It is important to note that the segmentation accuracy here is being judged against the "ground truth" of human labeled data. Therefore, it appears that Cartesian representation most closely corresponds to the human notion of motion primitives. This correlates with findings in neuroscience which indicate that human motions, and particularly arm motions, are planned and formulated in extrinsic-kinematic space [27].

Figure 4 shows an excerpt from the time series of the segmentation results, and Figure 5 shows a comparison of the segmentation outputs for each representation type. As can

be seen from these figures, the quaternion segmentation performs well when there is a change between the types of motion, for example the change from a squat motion to an arm motion in Figure 4, but performs poorly for distinguishing a change in the direction of motion, for example, the segment between an arm raise and an arm lower. This can also be seen in Figure 5, where the quaternion segmentation produces much lower accuracy segmentation results for motions which are a change of direction only, such as the "Right/Left/Both Arm Lower Start", which always follows the associated arm raise motion in the dataset, or the "Bend/Squat Raise Start", which always follows "Bend/Squat Lower". For these types of motions, the quaternion segmentation produces significantly more errors compared to the Cartesian result. The Quaternion segmentation also generates additional false positives, compared to the Cartesian segmentation results. A similar result is observed for the Euler angles. On the other hand, the Cartesian segmentation performs equally well at both inner and outer segment points.

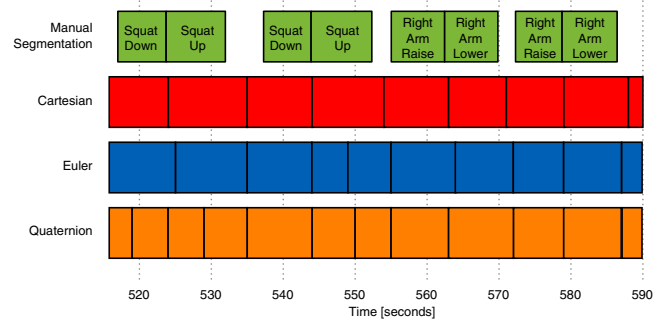


Fig. 4. Segmentation Sequence Excerpt

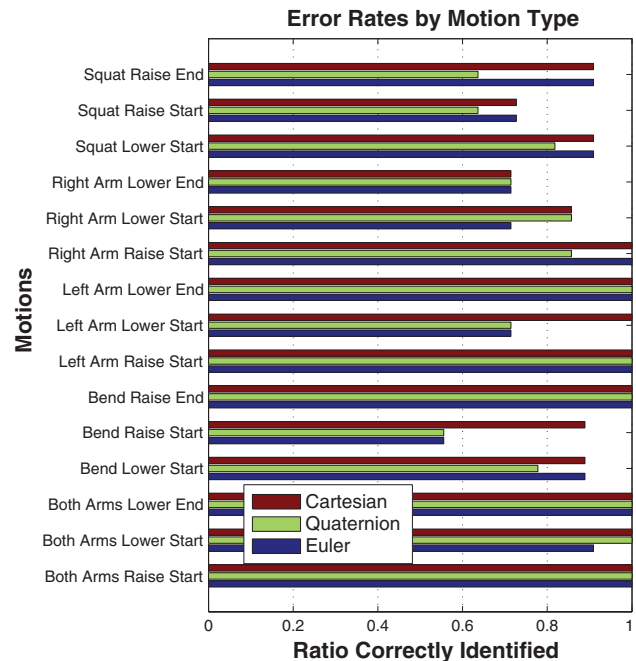


Fig. 5. Segmentation Performance Comparison by Motion Type

A second key finding from these experimental results is that the segmentation performance does not decrease when the number of degrees of freedom is increased from 20 DoF to 40 DoF to model the human motion, as comparable results were achieved for similar motion data modeled with a simplified model [4] and with a higher order model. This indicates that the proposed segmentation algorithm is suitable for higher order models. Unlike many previous approaches, which focus on off-line analysis or analysis of only parts of the body or only specific motions such as walking, the proposed method can perform on-line segmentation of full body motion, using a detailed kinematic model, and with no a-priori knowledge of the type of motion to be performed.

## V. CONCLUSIONS AND FUTURE WORK

This paper investigated the use of automated stochastic segmentation for full body human motion modeled by high DoF kinematic models. The choice of representation for spherical joints in the kinematic model was also considered. Experimental results on a human motion database confirm the feasibility of the automated segmentation approach for larger kinematic models, by showing that similar or improved segmentation performance can be achieved when more realistic models are used. The experiments also indicate that Cartesian representation corresponds most closely to the human observer generated segmentation.

In future work we plan to consider the issue of hierarchical segmentation, to handle those motions where multiple motion primitives are blended or overlaid, for example stepping while reaching. We are also investigating the use of stochastic segmentation for task based motions, where a combination of kinematic and environmental data is used to perform the automated segmentation.

## ACKNOWLEDGMENTS

The authors gratefully acknowledge the assistance of Hiro-taka Imagawa and Akihiko Murai with the data set collection. This work is supported by the Japanese Society for the Promotion of Science Category S Grant-in-Aid for Scientific Research 20220001.

## REFERENCES

- [1] C. Breazeal and B. Scassellati, "Robots that imitate humans," *Trends in Cognitive Sciences*, vol. 6, no. 11, pp. 481–487, 2002.
- [2] S. Schaal, A. Ijspeert, and A. Billard, "Computational approaches to motor learning by imitation," *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, vol. 358, pp. 537 – 547, 2003.
- [3] D. Kulić, D. Lee, Ch. Ott, and Y. Nakamura, "Incremental learning of full body motion primitives for humanoid robots," in *Proceedings of the IEEE International Conference on Humanoid Robots*, 2008, pp. 326–332.
- [4] D. Kulić and Y. Nakamura, "Scaffolding on-line segmentation of full body human motion patterns," in *Proceedings of the IEEE/RJS International Conference on Intelligent Robots and Systems*, 2008, pp. 2860–2866.
- [5] J. Kohlmorgen and S. Lemm, "A dynamic hmm for on-line segmentation of sequential data," in *NIPS 2001: Advances in Neural Information Processing Systems*, T. G. Dietterich, S. Becker, and Z. Ghahramani, Eds., vol. 14, 2002, pp. 793–800.
- [6] M. Pomplun and M. J. Matarić, "Evaluation metrics and results of human arm movement imitation," in *Proceedings of the IEEE International Conference on Humanoid Robotics*, 2000.
- [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] J. Lieberman and C. Breazeal, "Improvements on action parsing and action interpolatin for learning through demonstration," in *Proceedings of the IEEE/RJS International Conference on Humanoid Robots*, 2004, pp. 342–365.
- [9] T. Kwon and S. Y. Shin, "Motion modeling for on-line locomotion synthesis," in *Proceedings of the 2005 ACM SIGGRAPH/Eurographics symposium on Computer animation*, 2005, pp. 29–38.
- [10] G. Liu and L. McMillan, "Segment-based human motion compression," in *Proceedings of the 2006 ACM SIGGRAPH/Eurographics symposium on Computer animation*, 2006, pp. 127–135.
- [11] Y. Sakamoto, S. Kuriyama, and T. Kaneko, "Motion map: image-based retrieval and segmentation of motion data," in *Proceedings of the 2004 ACM SIGGRAPH/Eurographics symposium on Computer animation*, 2004, pp. 259–266.
- [12] T. Kwon, Y.-S. Cho, S. I. Park, and S. Y. Shin, "Two-character motion analysis and synthesis," *Visualization and Computer Graphics, IEEE Transactions on*, vol. 14, no. 3, pp. 707–720, 2008.
- [13] J. Barbič, A. Safonova, J.-Y. Pan, C. Faloutsos, J. K. Hodgins, and N. S. Pollard, "Segmenting motion capture data into distinct behaviors," in *Proceedings of Graphics Interface 2004*, 2004, pp. 185–194.
- [14] D. Kulić, D. Lee, and Y. Nakamura, "Whole body motion primitive segmentation from monocular video," in *Proceedings of the IEEE International Conference on Robotics and Automation*, 2009, to be published.
- [15] N. Koenig and M. J. Matarić, "Behavior-based segmentation of demonstrated tasks," in *Proceedings of the International Conference on Development and Learning*, 2006.
- [16] B. Janus and Y. Nakamura, "Unsupervised probabilistic segmentation of motion data for mimesis modeling," in *Proceedings of the IEEE International Conference on Advanced Robotics*, 2005, pp. 411–417.
- [17] B. Janus, "On-line motion segmentation algorithm for mimesis model," Master's thesis, University of Tokyo, 2006.
- [18] W. Ilg, G. H. Bakir, J. Mezger, and M. A. Giese, "On the representation, learning and transfer of spatio-temporal movement characteristics," *International Journal of Humanoid Robotics*, vol. 1, no. 4, pp. 613–636, 2004.
- [19] W. Takano and Y. Nakamura, "Humanoid robot's autonomous acquisition of proto-symbols through motion segmentation," in *Proceedings of the IEEE International Conference on Humanoid Robots*, 2006, pp. 425–431.
- [20] W. Takano, "Stochastic segmentation, proto-symbol coding and clustering of motion patterns and their application to significant communication between man and humanoid robot," Ph.D. dissertation, University of Tokyo, 2006.
- [21] M. Pardowitz, R. Haschke, J. Steil, and H. Ritter, "Gestalt-based action segmentation for robot task learning," in *Proceedings of the IEEE International Conference on Humanoid Robots*, 2008, pp. 347–352.
- [22] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.
- [23] D. Kulić, W. Takano, and Y. Nakamura, "Incremental learning, clustering and hierarchy formation of whole body motion patterns using adaptive hidden markov chains," *International Journal of Robotics Research*, vol. 27, no. 7, pp. 761–784, 2008.
- [24] L. Sciavicco and B. Siciliano, *Modeling and Control of Robot Manipulators*, 2nd ed., ser. Advanced Textbooks in Control and Signal Processing. Springer-Verlag, 2000.
- [25] K. Yamane and Y. Nakamura, "Natural motion animation through constraining and deconstraining at will," *IEEE Transactions on Visualization and Computer Graphics*, vol. 9, no. 3, pp. 352–360, 2003.
- [26] G. Venture, K. Ayusawa, and Y. Nakamura, "Motion capture based identification of the human body inertial parameters," in *IEEE/EMBS Proc. Int. Conf. on Engineering in Medicine and Biology*, 2008, pp. 4575–4578.
- [27] T. Flash and N. Hogan, "The coordination of arm movements: An experimentally confirmed mathematical model," *Journal of Neuroscience*, vol. 5, no. 7, pp. 1688–1703, 1985.