

What Do You Expect from a Robot that Tells Your Future? The Crystal Ball

Wataru Takano, Hirotaka Imagawa, Dana Kulić and Yoshihiko Nakamura

Abstract—This paper proposes an approach to hierarchy formation of human behaviors, extraction of the behavioral transitions, and their application to prediction and automatic generation of behaviors. Human demonstrator motion patterns are stored as motion symbols, which abstract the motion data by using Hidden Markov Models. The stored motion patterns are organized into a hierarchical tree structure, which represents the similarity among the motion patterns and provides abstracted motion patterns. Concatenated sequences of motion patterns are stochastically represented as transitions between the abstracted motion patterns by using an Ngram Model, and the transitional relationships of the human behaviors are extracted. The behavioral hierarchy and transition model make it possible to predict human behaviors during observation and to generate sequences of motion patterns automatically while maintaining a natural motion stream, as if the system is a “crystal ball” to reflect future behaviors. The experiments validate the proposed framework by using a developed visualization system, which shows the demonstrator or the operator the established hierarchical tree and the transition network of the motion patterns, predicted behaviors and generated sequences of the motion patterns.

I. INTRODUCTION

Structuring knowledge about human motion through observation of human behavior is a notable research issue. The motion knowledge can be applied to intelligent humanoid robots which can recognize the human motion and generate human-like motion autonomously, realistic human-like motion generation for animated characters, and motion analysis during sports training or rehabilitation.

In robotics, there has been a long standing research work to develop algorithms of motion learning by imitation for robots [1]. Especially, various approaches to symbolize motion patterns through observation have been proposed.

Tani et al. describe an approach for modeling robot motion patterns by using Recurrent Neural Networks with Parametric Bias (RNNPB) [2]. Motion patterns are abstracted by a parameter bias layer added to an input layer in a recurrent neural network. Hidden Markov Models (HMMs) have been a prominent technique for motion pattern symbolization [3][4]. Motion patterns have been encoded by sets of parameters of HMMs, recognized as the HMMs and generated by the HMMs. However, clustering and sequential relationships between the motion patterns are not represented by these frameworks.

Kadone et al. uses an associative neural networks with non-monotonic sigmoid functions for hierarchical clustering of motion patterns [5]. Parameters in the sigmoid function correspond to a abstraction level. The observed motion are recognized as motion patterns at arbitrary abstraction levels by tuning the parameters. However, the time correlation between the motion patterns is not encapsulated by their model.

Kulić et al. have been developing algorithms for clustering motion patterns to form a hierarchical tree structure representing motions learned by robots, and for establishing a directed graph representing transitions between the motion patterns [6]. Observed motions are abstracted by using HMMs. The parameters of the model define a distance between the motions. The motions are clustered based on this distance to form a hierarchical tree structure. Each node in the tree represents a cluster of motion patterns, which can be used for motion recognition and generation. However, the tree structure depends on some parameters of distance thresholds for the cluster, which are manually specified. The parameter tuning requires expertise to construct a desired tree structure.

Kovar et al. proposed a motion graph [7]. The seamless motion transition candidates can be detected by distances between segmentation points of motion data. An edge on the motion graph corresponds to frames of motion data, and a node serves as a choice point for connecting motion frames. A realistic looking motion can be generated under dynamics of the motion graph. However, the motion patterns are not hierarchically represented by model parameters to reduce the dimensionality of the motion.

Sidenbladh et al. proposed a behavioral prediction method [8]. This method uses the dimensional reduction of motion data by the principal component analysis and the binary tree of the low dimensional data. This approach is suitable for generating motion prediction. The prediction based on a large database of motion data requires modeling the motion data as symbols in order to compress temporal spatial data.

A structure of motion primitives is represented by a context-free-grammar model [9]. A pair of adjacent motion primitives is grouped incrementally and a tree structure can be formed. A node in the tree corresponds to a short sequence of motion primitives. The motion primitives are not symbolized in various abstraction levels.

This paper describes a framework to extract knowledge of human motion by forming a hierarchical structure of motion patterns through binary tree clustering and by establishing a directed graph representing transitions between

W. Takano, H. Imagawa and Y. Nakamura are with Mechano-Informatics, University of Tokyo, {takano, imagawa, nakamura}@ynl.t.u-tokyo.ac.jp. D. Kulić is with Electrical and Computer Engineering, University of Waterloo {dkulic}@ece.uwaterloo.ca.

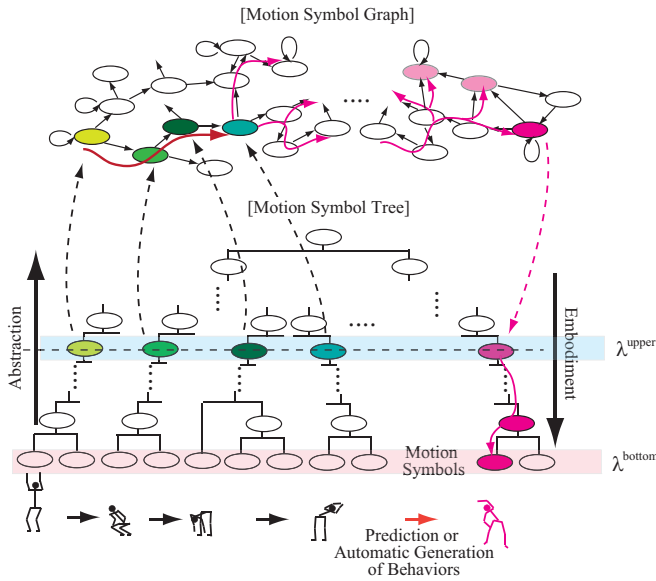


Fig. 1. The motion pattern data is encoded into a motion symbol. Similar motion symbols are grouped into another abstract motion symbol. The grouping process forms a motion symbol tree. A motion data stream is expressed as a sequence of the abstract motion symbols. The sequential relationship between the motion symbols is stochastically represented by a motion symbol graph. A current node in the motion graph can be estimated from an observed motion sequence. Multiple paths starting at the current node with large likelihood can be retrieved. The paths show future motion patterns.

abstracted motion patterns. The hierarchical structure can be constructed by the Ward clustering method. The motion sequences are memorized as transition probabilities between the abstracted motion patterns in the directed graph. The integration of the hierarchical structure and the directed graph make it possible to predict motion patterns following the observed motion sequences and to autonomously generate frequent motion sequences on the transition. An experiment using a motion capture system verified the validity of our proposed framework.

The proposed framework is a basic technology for robots to memorize motion patterns as symbols during observation, recognize human behaviors as memorized symbols, predict the human motions through observation and support the humans by understanding their behaviors.

II. MOTION SYMBOL TREE AND MOTION SYMBOL GRAPH FORMATION

Fig.1 shows an overview of the motion symbol tree and the motion symbol graph. The motion symbol tree hierarchically represents clustering relationships between motion patterns, and the motion symbol graph represents sequential relationships between the clusters of the motion patterns. The motion symbol tree and the motion symbol graph realize fast motion recognition processing, prediction of motion patterns through observation of whole body human motions, and automatic generation of frequently observed motion sequences.

A. Motion Symbol Tree

Continuous motion data is automatically segmented into motion pattern data [10][11][12]. The motion pattern data O_i ($i = 1, 2, 3, \dots, n$) is encoded into the corresponding HMM λ_i ($i = 1, 2, 3, \dots, n$), since the HMM can be used for both motion recognition and motion generation. n is the number of the segmented motion pattern data set. The dissimilarity between the HMMs, λ_i and λ_j , is defined by Kullback Leibler information :

$$D(\lambda_i, \lambda_j) = \frac{d(\lambda_i, \lambda_j) + d(\lambda_j, \lambda_i)}{2} \quad (1)$$

$$d(\lambda_i, \lambda_j) = \frac{\ln P(O_i|\lambda_i) - \ln P(O_i|\lambda_j)}{T_i}$$

where $P(O|\lambda)$ is the probability that the HMM λ generates the motion pattern data O , and T_i is the length of the motion pattern data O_i . The motion pattern data can be grouped into abstracted motion patterns based on the dissimilarity by using the Ward clustering method. When the dissimilarity between λ_1^{child} and λ_2^{child} is smallest, these two HMMs are grouped into λ^{parent} , which is optimized by using training motion pattern data encoded into λ_1^{child} and λ_2^{child} . The incremental grouping forms a hierarchical tree, where a node in the bottom layer represents HMMs corresponding to each motion pattern data and a node in the upper layer represents HMMs corresponding to multiple motion pattern data. Since the HMMs abstract motion pattern data, the HMMs and the formed tree are called “motion symbol” and “motion symbol tree” respectively.

The constructed motion symbol tree can be used to recognize observed motions as motion symbols in various abstractive resolutions according to following steps:

- step1 The observation O is given to motion symbol λ^{parent} .
- step2 The likelihood that the observation O is generated by motion symbol λ_1^{child} or λ_2^{child} is computed, where the motion symbols, λ_1^{child} and λ_2^{child} , correspond to child nodes of the motion symbol λ^{parent} .
- step3 The motion symbol (λ_1^{child} or λ_2^{child}) with the largest likelihood becomes the parent motion symbol, and go back to step1. If the motion symbol is located on the bottom layer and it has no child motion symbol, the iteration processing is terminated.

This computation allows for fast motion recognition.

B. Motion Symbol Graph

Continuous motion data is recognized as a sequence of the motion symbols. The sequential relationship between the motion symbols can be stochastically modeled by an ergodic HMM, where motion symbols are generated by each node through node transition. However it is difficult for the ergodic HMM to learn a large amount of the sequential relationship. The relationship need to be learned by a simple directed graph, where each node represents a motion symbol and each edge represents a transition between the motion symbols.

N-gram model is used for the motion symbol graph in this paper. N-gram is a stochastic model, which takes the assumption that each motion symbol depends only on its

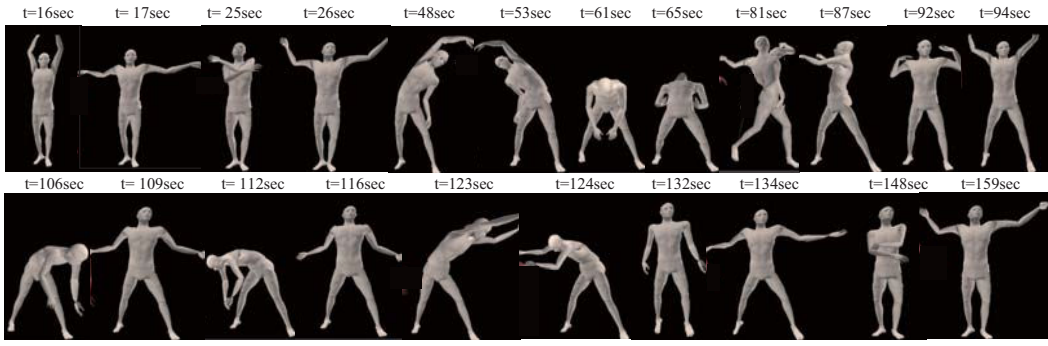


Fig. 3. The demonstrator performs radio physical exercises. Human whole body motion data of the demonstrator is collected in an optical motion capture system. The radio physical exercise consists of following motion patterns: moving both hands in a circle, arching the back, bending the body to the left, bending the body to the right, forward bending, backward bending, twisting the body around the vertical axis, raising both hands, twisting the body around the horizontal axis, jumping and deep breathing.

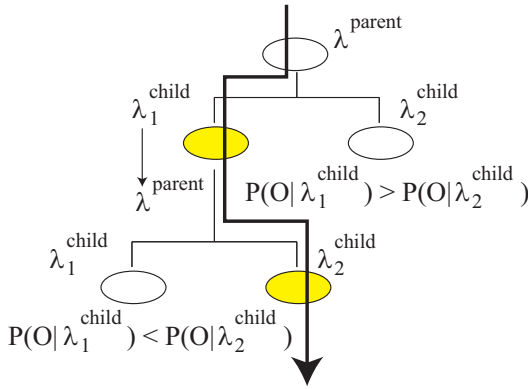


Fig. 2. The motion pattern data is incoming the motion symbol in the top layer. The motion symbol is a parent motion symbol and two motion symbols forming the parent motion symbol are child motion symbols. The system selects one child motion symbol with the larger likelihood that the motion symbol generates the motion pattern data. The selected motion symbol becomes a parent motion symbol, and one child motion symbol is chosen as recognition result in the same manner. This iteration leads to fast motion recognition.

$N - 1$ previous motion symbols. Each edge in the motion symbol graph is expressed by the transition probability $P(\lambda^{upper}[k] | \lambda^{upper}[k - N + 1], \dots, \lambda^{upper}[k - 1])$. Note that the segmented motion pattern data O_k is recognized as the motion pattern symbols $\lambda^{upper}[k]$, the abstraction level of which is manually given. The transition probability can be derived as :

$$P(\lambda_i | \lambda_1^{N-1}) = \frac{C(\lambda_1^{N-1}, \lambda_i)}{C(\lambda_1^{N-1})} \quad (2)$$

where λ_1^{N-1} is a sequence of $N - 1$ motion symbols, and the number of sequences λ in the training motion data is given by $C(\lambda)$.

The motion pattern data O_k is also recognized as the motion symbol $\lambda^{bottom}[k]$, which corresponds to a motion pattern and is located at the bottom in the motion symbol tree. The output probability that the motion symbol λ^{upper}

generates the motion symbol λ^{bottom} is estimated as :

$$P(\lambda^{bottom} | \lambda^{upper}) = \frac{C(\lambda^{bottom}, \lambda^{upper})}{C(\lambda^{upper})} \quad (3)$$

where $C(\lambda^{bottom}, \lambda^{upper})$ is the number of times that the motion pattern data is recognized as both λ^{bottom} and λ^{upper} , and $C(\lambda^{upper})$ is the number of times that the motion pattern data is recognized as λ^{upper} .

In this way, a sequential ordering of motion pattern data can be characterized by the transition relationship between the motion symbols in the upper layer and the output relationship between the motion symbols in the upper layer and in the bottom layer. The motion symbol graph is represented by a set of the transition probabilities and the output probabilities.

The motion symbol graph can be used to calculate possible sequences of motion patterns by searching the graph for valid paths starting at an initial motion symbol. The retrieved sequences of motion patterns leads to technologies for motion prediction and automatic motion generation. The search for the paths is done based on the following evaluation function :

$$E(\lambda^{bottom}_{k+1}^{k+M}) = \log P(\lambda^{bottom}_{k+1}^{k+M} | \lambda^{upper}_{k-N+1}^k) \quad (4)$$

$$= \sum_{t=k+1}^{k+M} \log P(\lambda^{bottom}[t] | \lambda^{upper}[t])$$

$$+ \sum_{t=k+1}^{k+M} \log P(\lambda^{upper}[t] | \lambda^{upper}_{t-N}^{t-1}) \quad (5)$$

Eqn.4 represents the likelihood that a sequence of motion symbols is generated given an initial sequence of motion symbols, $\lambda^{upper}_{k-N+1}^k$. The first term and the second term in Eqn.5 are the output probability and transition probability respectively. The path with the largest likelihood can be found by using the A* search algorithm. M is the length of a sequence of motion symbols for motion prediction and motion generation.

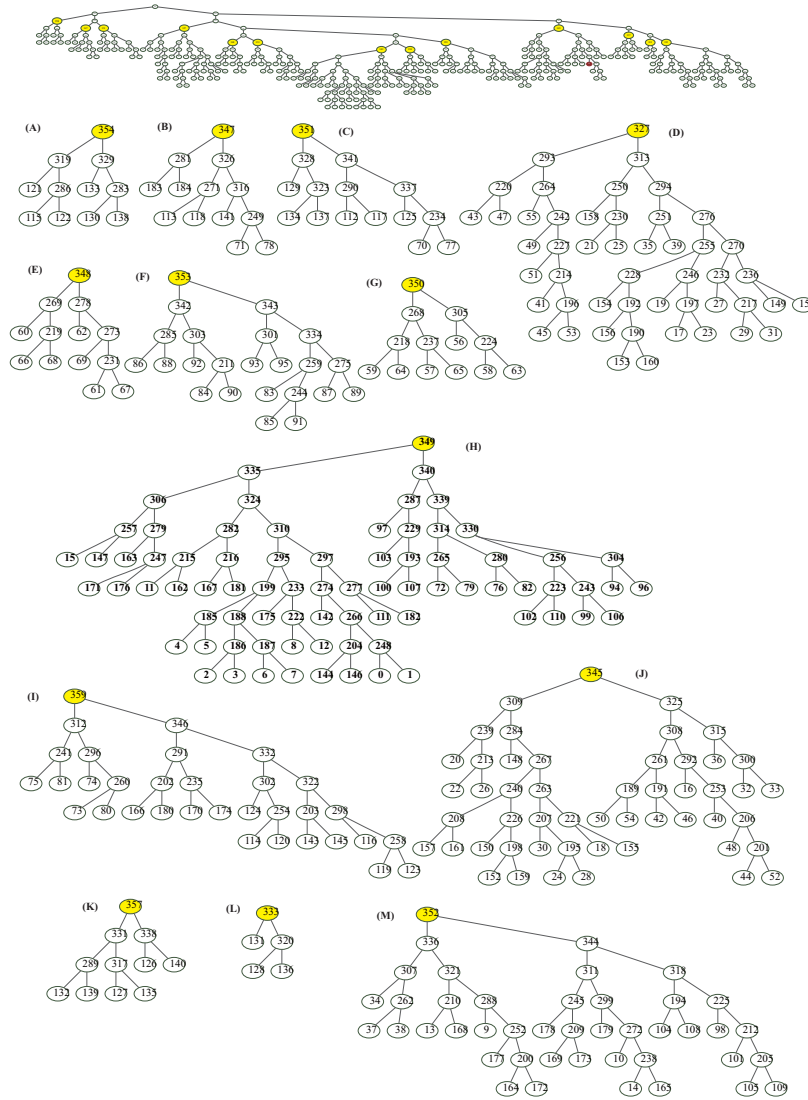


Fig. 4. The motion data collected during a radio physical exercise is segmented into 185 motion pattern data, which are encoded into the corresponding HMMs. The HMMs are labeled as motion symbol indices(from 0 to 184). They are located in the bottom layer. In this figure, 185 motion pattern data are grouped into 13 abstracted motion symbols.

III. EXPERIMENTS

The proposed framework of combining the motion symbol tree and the motion symbol graph was tested on human whole body motion data collected in an optical motion capture system. The motion capture system measures the positions of 34 reflective markers attached to a demonstrator, at a sample rate of 5ms. The marker positions are converted to positions of virtual markers attached to a demonstrator through inverse kinematics computation using a 40 DOF model. The virtual markers indicate representative points for each body part of the demonstrator. The motion pattern data is represented by a time series data of 112 dimensional vectors. The demonstrator performs exercises from an exercise video for about 3min. This exercise is call “radio physical exercise” in Japan.

The captured motion data is segmented into 185 motion pattern data. The motion symbol tree is established using

the motion pattern data, which are represented by HMMs, and is hierarchically grouped into a motion symbol tree. The resulting motion symbol tree is shown in Fig.4. The nodes in the bottom layer of the tree correspond to the captured motion pattern data, and the motion pattern data is abstracted by the nodes in the upper layers. In Fig.4, the 185 motion pattern data sets form 13 motion clusters. Fig.5 shows motion patterns in the 13 motion clusters. The motion patterns in the same cluster are similar. The Ward clustering algorithm based on Kullback Leibler information between the motion pattern data is found to be effective.

The motion data is represented by a sequence of the motion pattern symbols in the resulting motion symbol tree. The sequence has been learned as a motion symbol graph. Fig.6 shows the formed motion symbol graph, which consists of 60 motion symbols. Note that the motion symbol graph in Fig.6 is bigram model(N=2) for visualization. The 4-gram

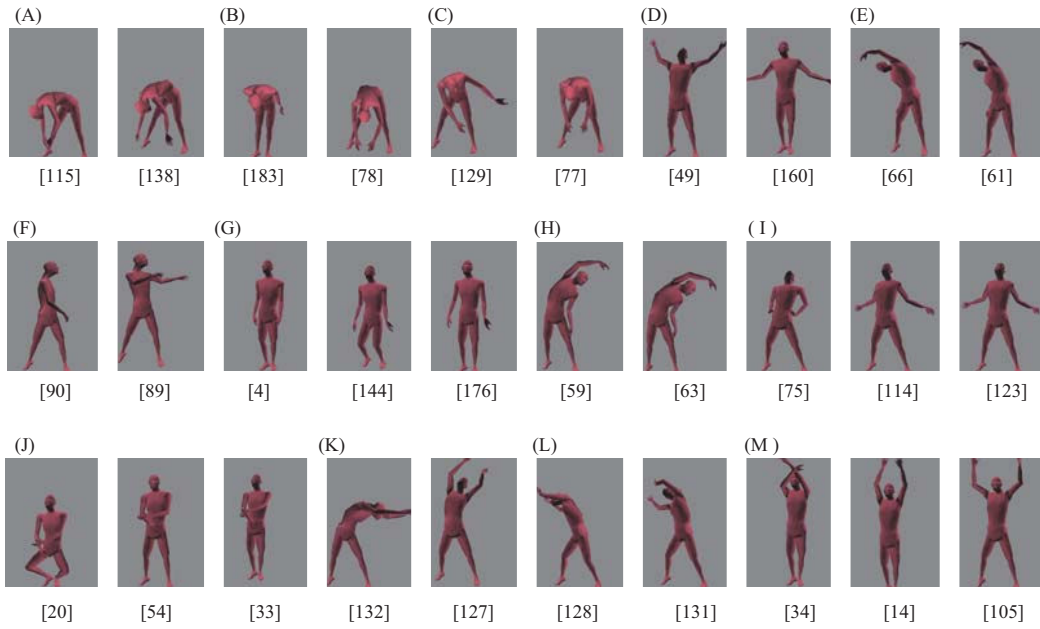


Fig. 5. Two Motion patterns classified into one of the 13 groups in Fig.4 are shown. Motion patterns of forward bending are categorized as Group (A). Motion patterns of twisting the body are included in Group (F). Motion patterns of bending the body to the right and left are categorized as Group(E) and (H) respectively. Similar motion pattern data are classified into the same Group.

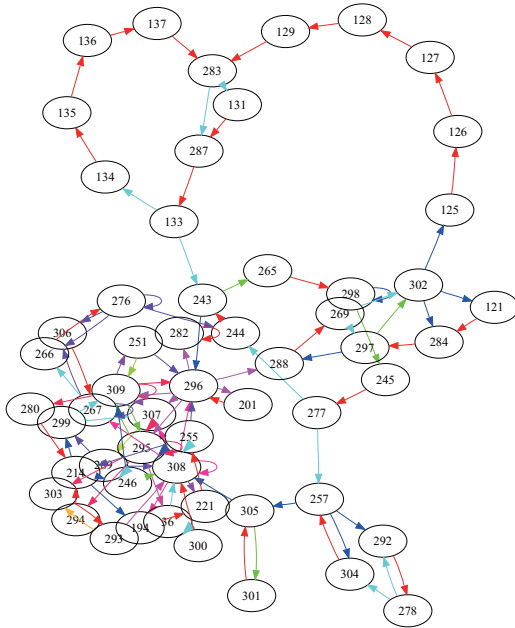


Fig. 6. The motion graph shows the bigram ($N = 2$) for visualization. The transition between the motion symbols can be represented by a edge in the graph.

is used for motion prediction.

Fig.7 shows sequences of motion patterns predicted from the observed motions. Note that 8 motion patterns following the current motion are predicted. Table.I shows the evaluation result of the motion prediction. The prediction performance

Z is derived as :

$$Z = \frac{N_{correct}}{N_{correct} + N_{wrong}} \quad (6)$$

$N_{correct}$ is the number of predicted motion patterns which correspond to actual motion patterns. If the predicted motion pattern does not coincide with the actual motion pattern, the motion pattern is counted as N_{wrong} . The coincidence means that an index of the predicted motion pattern symbol in the bottom layer of the motion symbol tree is equal to an index of the recognition result of an observed motion pattern. The number of predicted motion sequences is set to 10. The prediction performance is 82.2% for a motion pattern following the observed motion. The motion pattern which follows subsequently can also be predicted with the performance of 77.1%. In this way, it can be seen that the accurate prediction of the motion patterns can be achieved from Fig.7 and Table.I.

IV. CONCLUSIONS

The contributions of this paper are summarized as follows:

- 1) This paper presents a motion symbol tree ,which is formed by encoding motion pattern data into HMMs, grouping the HMMs by Ward clustering method. Each group is found to consist of similar motion pattern data through experiment with a motion capture system.
- 2) The sequential relationship between motion symbols is stochastically represented by a motion symbol graph, where a node corresponds to a motion symbol or a short sequence of motion symbols, and an edge corresponds to a transition between motion symbols or between the short sequences. This paper presents

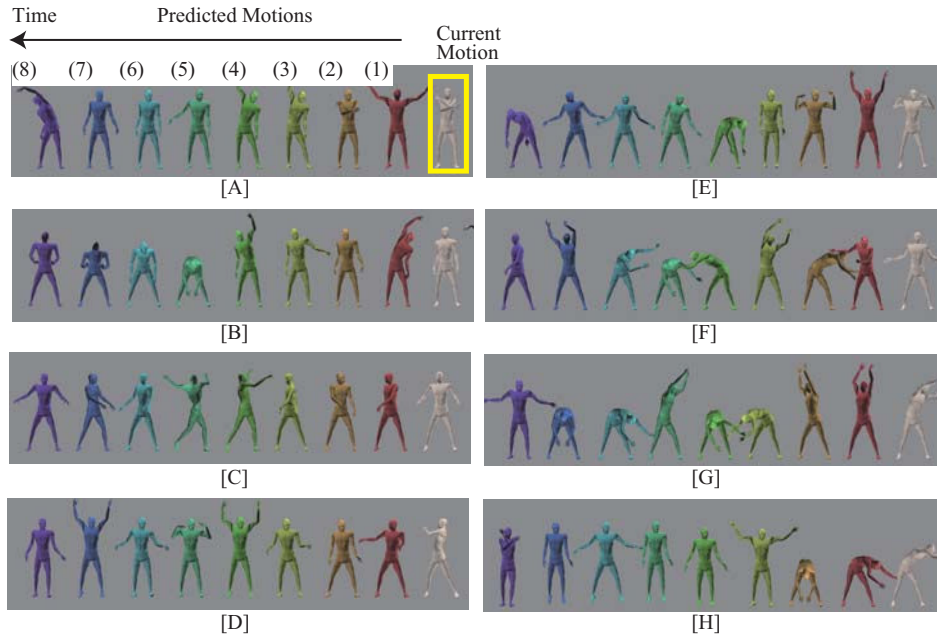


Fig. 7. Motion patterns “bending the body to the left and right” are predicted from the motion “arching the back” in [A]. Motion patterns “forward and backward bending” are predicted during observation of the motion “bending the body to the left and right” [B]. The motion patterns are accurately predicted from the observation in the same orders as the training radio physical exercise.

TABLE I

PREDICTION PERFORMANCE IS SHOWN IN THIS TABLE. THE MOTION PATTERNS FOLLOWING THE CURRENT MOTION CAN BE PREDICTED VERY ACCURATELY. THE PREDICTION PERFORMANCE IS 82.2%. THE EIGHTH NEXT MOTION PATTERN CAN BE ALSO PREDICTED WITH 65.1% ACCURACY.

| Prediction Term | Prediction Performance |
|--------------------|------------------------|
| 1st motion pattern | 82.2 |
| 2nd motion pattern | 77.1 |
| 3rd motion pattern | 74.2 |
| 4th motion pattern | 70.2 |
| 5th motion pattern | 69.7 |
| 6th motion pattern | 67.4 |
| 7th motion pattern | 64.5 |
| 8th motion pattern | 65.1 |

an approach to predict motion patterns using the motion symbol graph. The motion can be predicted by searching for the most probable paths starting at a node corresponding to the current observation.

- 3) We tested the proposed framework integrating the motion symbol tree and the motion symbol graph on the captured motion data of physical exercise. The motion following the current observation can be predicted with 82.2% accuracy. Even when the 8th next motion pattern is predicted, the accuracy is 65.1%. The experiment validates the prediction based on the motion symbol tree and the motion symbol graph.

V. ACKNOWLEDGMENTS

This research was supported by Category S of Grant-in-Aid for Scientific Research (20220001), Japan Society for

the Promotion of Science.

REFERENCES

- [1] C. Breazeal and B. Scassellati, “Robots that imitate humans,” *Trends in Cognitive Sciences*, vol. 6, no. 11, pp. 481–487, 2002.
- [2] J. Tani and M. Ito, “Self-organization of behavioral primitives as multiple attractor dynamics: A robot experiment,” *IEEE Transactions on Systems, Man and Cybernetics Part A: Systems and Humans*, vol. 33, no. 4, pp. 481–488, 2003.
- [3] T. Inamura, I. Toshima, H. Tanie, and Y. Nakamura, “Embodied symbol emergence based on mimesis theory,” *International Journal of Robotics Research*, vol. 23, no. 4, pp. 363–377, 2004.
- [4] A. Billard, S. Calinon, and F. Guenter, “Discriminative and adaptive imitation in uni-manual and bi-manual tasks,” *Robotics and Autonomous Systems*, vol. 54, pp. 370–384, 2006.
- [5] H. Kadone and Y. Nakamura, “Symbolic memory for humanoid robots using hierarchical bifurcations of attractors in nonmonotonic neural networks,” in *Proceedings of the 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2005, pp. 2900–2905.
- [6] D. Kulic, H. Imagawa, and Y. Nakamura, “Online acquisition and visualization of motion primitives for humanoid robots,” in *Proceedings of the 18th IEEE International Symposium on Robot and Human Interactive Communication*, 2009, pp. 1210–1215.
- [7] L. Kovar, M. Gleicher, and F. Pighin, *ACM Transactions on Graphics*, vol. 21, no. 3, pp. 473–482, 2002.
- [8] H. Sidenbladh, M. J. Blac, and L. Sigal, “Implicit probabilistic models of human motion for synthesis and tracking,” in *Proceedings of European Conference on Computer Vision*, 2002, pp. 784–800.
- [9] G. Guerra-Filbo and Y. Aloimonos, “A language for human action,” *IEEE Computer Magazine*, vol. 40, no. 5, pp. 60–69, 2007.
- [10] J. Kohlmorgen and S. Lemm, “A dynamic hmm for on-line segmentation of sequential data,” in *Advances in Neural Information Processing System*, vol. 14, 2001.
- [11] B. Janus and Y. Nakamura, “Unsupervised probabilistic segmentation of motion data for mimesis modeling,” in *Proceedings of IEEE International Conference on Advanced Robotics*, 2005, pp. 411–417.
- [12] D. Kulic and Y. Nakamura, “Scaffolding on-line segmentation of full body human motion patterns,” in *Proceedings of the IEEE/RSJ 2008 International Conference on Intelligent Robots and Systems*, 2008, pp. 2860–2866.