Missing Motion Data Recovery using Factorial Hidden Markov Models

Dongheui Lee, Dana Kulić and Yoshihiko Nakamura

Department of Mechano-Informatics
The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, JAPAN
{dhlee, dana, nakamura}@ynl.t.u-tokyo.ac.jp

Abstract— This paper proposes a method to recover missing data during observation by factorial hidden Markov models (FHMMs). The fundamental idea of the proposed method originates from the mimesis model, inspired by the mirror neuron system. By combining the *motion recognition from partial observation* algorithm and the *proto-symbol based duplication of observed motion* algorithm, whole body motion imitation from partial observation can be achieved. The algorithm for *missing data recovery* uses the same basic strategy as the *whole body motion imitation from partial observation*, but requires more accurate spatial representability. FHMMs allow for more efficient representation of a continuous data sequence by distributed state representation compared to hidden Markov models (HMMs). The proposed algorithm is tested with human motion data and the experimental results show improved representability compared to the conventional HMMs.

Index Terms— factorial hidden Markov model, mimesis, motion recovery

I. INTRODUCTION

In order for robots to successfully interact with humans and operate in daily life, they need to not only perform a variety of tasks but also understand human beings. The imitation learning mechanism provides a means of automatic programming of complex systems such as dexterous anthropomorphic robots, without extensive trials or complex programming. In particular, for humanoid robots, learning motion patterns by observing and imitating humans is not only highly efficient but also desirable for social interaction.

The neuroscience evidence of motor primitives and mirror neurons [1] [2], found in humans and other primates, has provided a scientific background and suggested a direction of technical studies for robot imitation learning [3] [4] [5] [6] [7] [8] [9] [10]. Bentivegna and Atkeson [3] [4] used the idea of primitives for motor learning to play air hockey and marble maze. Billard and Matarić [6] [7] used connectionist-based approaches to represent movements. Fod et al. [8] automatically derived primitives through an off-line process of segmentation and application of the principal component analysis. Inamura et al. [9] proposed the HMM (hidden Markov model) based mimesis model, where observed human motion primitives are encoded using HMMs and the HMMs are used for humanoid robot motion generation. Breazeal and

Scasellati [11] and Schaal et al. [12] provided reviews on motion learning by imitation.

Deacon [13] explained the developmental process of symbols in three stages: icon, index, and symbol. The idea of symbols or motion primitives has been used in many imitation learning algorithms [3] [4] [5] [6] [8] [9] [14]. Abstraction is an important strategy for handling the huge search space in the real world and the limited computation capability of the human brain. The real world is full of complex spatiotemporal experiences. By using the semiotics system, which segments meaningful chunks from the physical information, abstracts the complex spatiotemporal information into symbols, and infers the relation between symbols, human beings can handle the complex real world quickly in a reduced search space. The semiotics system also provides efficient memory management.

HMMs have been frequently used for modeling human motions because HMMs efficiently abstract time series data [9] [14]. When using HMMs for both motion recognition and motion generation, there is a tradeoff between recognition and generation performance, in particular when selecting the number of states of the model [15]. A small number of states give good generalization and recognition performance, while a large number of states give better generation performance, at the risk of over-fitting and poor generalization. Although the model selection problem for HMM models has been studied based on Bayesian [16] and Akaike [17] information criteria, recognition criteria only are used to select the appropriate number of states. In a prediction programming method [18], both the structure and the parameters of the HMM are trained on-line. However, the method has been applied in a simple task (i.e. manipulator trajectory prediction), not verified in complex tasks (i.e. human motion abstraction). Kulić et al. use factorial hidden Markov models (FHMMs) for human motion representation [15]. Another alternative has been proposed based on regression [19].

One of the most critical points in real applications is how to deal with partial observations. Among the research dealing with partial observations, Ghahramani and Jordan [20] proposed using the expectation-maximization (EM) algorithm to fill in missing feature values of examples when learning from

incomplete data by assuming a mixture model.

Full body motion imitation from occluded observations includes two issues. One is *how to understand motion patterns from partial observations*. Human beings can recognize and imitate the demonstrator's motion, even when a part of demonstrator's body is occluded, as long as the key features are visible. The binocular vision system of human beings has low depth sensitivity, specially for objects at distance. Humans can understand 3D images with the vision system of low depth sensitivity.

The second issue is *how to generate whole body motion similar to the demonstrator's motion using knowledge*. Imitation is not just copying the demonstrator's motion. It must be an inferred result by the semiotics system because it allows that one is aware of meanings of one's motions. Human infants develop their motion repertoire from the innate set of motions and the skill of mimesis. It is an iterative strategy to acquire a new motion pattern by modifying the generated one based on the previously acquired ones. It would be more efficient than acquiring by copying an unknown motion from scratch.

The authors have been developing the mimesis model which enables the humanoid to imitate complete motion patterns from incomplete human motion observations: full body motion imitation from occluded observations [21], a motion recognition algorithm from observing labeled markers on humans by an onboard monocular vision system [22], and a 3D motion imitation algorithm from 2D image sequences using vector field of feature points[23]. The underlying common strategies are *motion recognition from partial observation* and *proto-symbol based duplication of observed motion*.

Through the proposed symbol based duplication scheme, the repertoire of the humanoid's motion patterns is enriched, compared with replaying existing proto-symbols. A humanoid robot also can imitate a new motion pattern more quickly by applying prior knowledge, compared with learning the motion from scratch and generating it.

By combining *motion recognition from partial observation* algorithm and *proto-symbol based duplication of observed motion* algorithm, whole body motion imitation from partial observation can be achieved. However, the recovery accuracy of missing data is not sufficient. The insufficiency is inherited from not having enough discrete states of the model. The problem is similar to the tradeoff between recognition and generation performance. As the number of states of HMMs increases, motion representation becomes more accurate. On the other hand, the HMMs with a large number of states are hard to train without computational overflow or underflow and prone to over-fitting. Due to the over-fitting, recognition of new motion patterns becomes weak and the optimal state-sequence search performance becomes worse. In this paper, we develop a methodology for missing motion data recovery using Factorial Hidden Markov Models (FHMMs).
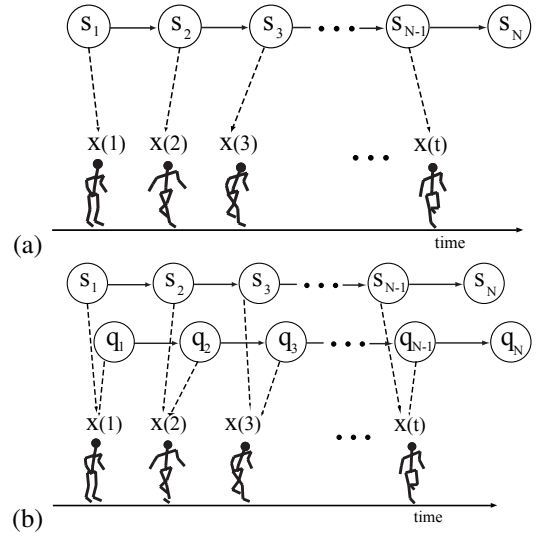

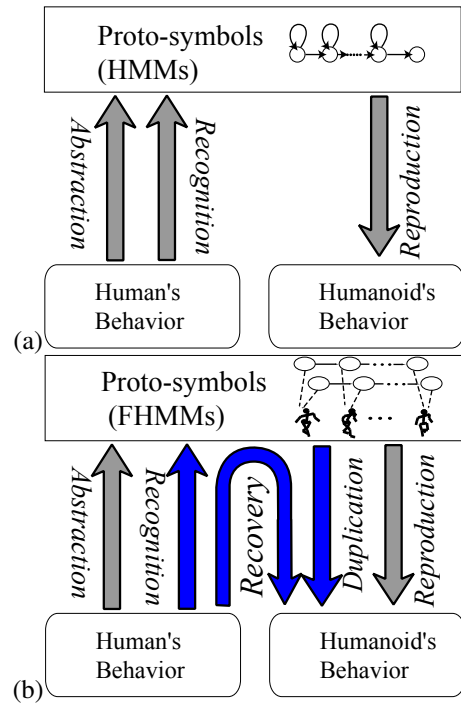Fig. 1.   (a) Hidden Markov Model, (b) Factorial Hidden Markov Model

Fig. 2.   (a) Mimesis Model [9]: The mimesis model is a mathematical model, inspired by the mirror neuron system. It is a bidirectional model which performs learning, recognition and generation functions through proto-symbols. The proto-symbols are defined through the HMM parameters. (b) The proposed strategy: Abstraction is done by the FHMM exact training [24]. Recognition is implemented by an adaptation of the forward-backward algorithm [24]. Reproduction is implemented by the greedy generation algorithm [15] or a stochastic generation algorithm . Algorithms of *motion recognition from partial observation* and *proto-symbol based duplication of observed motion* for FHMMs are developed. By combining the two algorithms, missing observation data are recovered accurately.

II. Motion Representation by Factorial Hidden Markov Models

A hidden Markov model (HMM) is a representation of a Markov process which cannot be directly observed. Efficient algorithms have been developed for model training (the Baum-Welch algorithm), pattern recognition (the forward algorithm and the backward algorithm) and hidden state sequence estimation (the Viterbi algorithm) [25].

HMMs have been frequently used for modeling human motions [26] [27] [28] [9]. In the mimesis model [9] shown in Fig. 2 (a), motion learning, recognition and generation are performed by using continuous HMMs (Fig. 1 (a)). In the motion learning procedure, time series of an observed human motion are abstracted as HMM parameters $\lambda = (A, B, \pi)$ via the Expectation Maximization (EM) technique. The trained HMM $\lambda = (A, B, \pi)$ is called a proto-symbol. A is the state transition probability matrix. π is the initial state probability vector. B is the observation emission probability distribution.

In motion recognition, the likelihood $P(O|\lambda)$, the probability that a time-series sequence of an observed motion pattern O is generated by a proto-symbol λ, is calculated against each proto-symbol. The most likely proto-symbol which generates the observed motion is sought.

In the motion generation function, a motion pattern is decoded from a selected proto-symbol via two processes: generation of the state sequence and generation of the output motion elements. The state sequence is decoded from the state transition probability matrix A and the initial state probabilities vector π. The output motion elements at each state are decoded from the observation symbol probability distribution B.

A factorial hidden Markov model (FHMM) is a hidden Markov model with a distributed state representations, as shown in Fig. 1 (b). An FHMM consists of multiple hidden Markov chains. Each dynamic chain is represented by its own state transition probability matrix, initial state probabilities vector, and emission probability distribution. Chains are independent of each other; they do not interact when generating data. The output of the model at each time step is, however, dependent on the values of the state variables of all the chains because the outputs from all the dynamic chains are summed.

Several FHMM training algorithms have been developed; the exact method [24], mean field approximation [24], backfitting algorithm [29], and sequential training method [30]. The exact method is an adaptation of the Baum-Welch algorithm but computationally expensive. The others are faster than the exact method but their training performance is not as good as that of the exact method. Once an FHMM is trained, pattern recognition can be implemented by an adaptation of the forward-backward algorithm [24]. A representative sequence from an FHMM can be generated by an adaptation of the motion generation of the mimesis model [15].

In contrast of the conventional HMMs, FHMMs are ef-ficient from a representational viewpoint. While in HMMs there is inherited tradeoff between recognition and generation performance in particular when selecting the number of states, FHMMs provide more efficient approach for combining good generalization for recognition purposes with sufficient detail for better generation. Kulić et al. [15] proposed a methodology for human motion representation based on FHMMs to avoid over-fitting problems with large number of states and poor generation and discrimination performance with small number of states, which are encountered with HMMs.

III. Missing Motion Data Recovery using FHMMs

One of the most critical points in real applications is how to deal with missing observation data. The authors have proposed algorithms which enable the humanoid to imitate complete motion patterns from incomplete human motion observations; full body motion imitation from occluded observations [21], a 3D motion imitation algorithm from observing labeled 2D markers on humans [22], and a 3D whole-body motion imitation from 2D occluded image sequences using vector field [23].

In [21], algorithms of *motion recognition from a partial observation* and *proto-symbol based duplication of an observed motion* are developed. The *motion recognition from a partial observation* algorithm may recall the mirror neuron's attribution of *activation from hidden observations* [31] [2]. According to congruence type of observation and generation, mirror neurons have been subdivided into "strictly congruent" and "broadly congruent" neurons [2]. Two thirds of mirror neurons are broadly congruent neurons which do not require that observation and action are exactly the same. Mirror neurons in which the effective observed and executed motions correspond in terms of goal (e.g., grasping) and means for reaching the goal (e.g., precision grip) have been classed as strictly congruent neurons. The algorithm of *proto-symbol based duplication of an observed motion*, which generates a motion pattern close to the observation, provides the similar attribution of the strictly congruent mirror neurons.

The *proto-symbol based motion duplication* algorithm performs motion imitation similar to the observation. By estimating the optimal state sequence using the Viterbi algorithm, the imitated motion pattern is temporally synchronized with the observation. Different motion patterns corresponding to the same proto-symbol can be imitated with different temporal sequences. This allows situated motion generation by temporal synchronization.

By combining the two algorithms, we achieved imitation of whole-body motion patterns from occluded observations [21] [23]. *Missing data recovery* can be achieved by the same basic strategy as the *whole body motion imitation from partial observation*. This is an interesting approach to recover missing data based on the mimesis model. However, the detail of the generated motion data by our previous research is

not good enough to say that the missing data is recovered accurately. The insufficiency to represent continuous human motion is inherited from discrete states of the hidden Markov chains. This problem is similar to the tradeoff between recognition and generation performance. As the number of states of HMMs increases, motion representation becomes more accurate. However, due to the over-fitting, recognition of new motion patterns becomes weak and the optimal state-sequence search performance becomes worse. On the other hand, an HMM with low number of states is good at recognizing new data, but cannot reproduce the observed data through generation. In this paper, we develop a methodology for accurate recovery of missing motion data using Factorial Hidden Markov Models (FHMMs), as shown in Fig. 2 (b).

A. Recognition of motion patterns with missing data

Motion recognition is a problem to find the most probable proto-symbol for the input observation sequences.

$$\lambda^* = \arg\max_\lambda P(x|\lambda) \tag{1}$$

Log-likelihood, $\log P(x|\lambda)$, for FHMMs is calculated by an adaptation of the forward-backward algorithm [24]. When there are missing motion elements in the input observation sequence x_t, the emission probability distribution $b_i(x_t)$, which is represented with a Gaussian distribution, is modified.

$$b_i(x_t) = \frac{exp\{-\frac{1}{2}(x_t - \mu_i)^T \Sigma_i^{-1}(x_t - \mu_i)\}}{\sqrt{(2\pi)^M \det \Sigma_i}} \tag{2}$$

For the missing motion elements $\{x_k\}_t$, either eq. (3) or eq. (4) is substituted into eq. (2),

$$\{x_k\}_t - \mu_{ij} = * \tag{3}$$
$$\Sigma_{ij} = \infty \tag{4}$$

so that the invisible motion elements do not affect the output probability density function. In eq. (3), $*$ indicates a constant value and in our experiments it is set to zero.

B. Proto-symbol based Motion Duplication

Motion patterns are decoded using the expectation operator in the stochastic model. The general motion generation is a two-stage stochastic process; state transition generation and motion output generation from the state sequence.

Missing data in the observed motion sequences are recovered by applying previous knowledge and current observations. The state sequence is obtained by applying the Viterbi algorithm [25], which computes the single best state sequence for the given observation sequence. Thus, this optimal state transition generation enables us to generate a whole body motion pattern close to the observed target motion pattern.

The Viterbi algorithm for FHMMs is developed. It is a three-step procedure. (1) An FHMM, consisting of M chains of N states, is expanded as a single hidden Markov chain with N^M states. The expanding procedure is adopted from

the exact training method [24]. (2) The single optimal state sequence S^* given an observation for the converted single hidden Markov chain with N^M states is calculated by the conventional Viterbi algorithm. The best state at time t, s_t^*, is a value between 1 and N^M. (3) The calculated optimal state sequence S^* for a single chain is converted into $S = \{S_1, \cdots, S_M\}$ for M chains. The best state at time t for the i-th chain, s_{it}, is a value between 1 and N.

When calculating the optimal state-sequence, the output probability density function is required. In order for the invisible motion elements not to affect the output probability density function, for the invisible motion elements, either $x - \mu = *$ or $\Sigma = \infty$ is applied, where $*$ indicates a constant value. After the optimal state sequence for each chain is obtained, the output observation sequence from each chain is calculated according to its output emission probability distribution in state i, i.e., $b_i(x)$. The simplest way to generate the output is by taking the mean vectors of each gaussian for the state-sequence. Then, the output observation sequence y is calculated by summing the contribution from each chain at each time step. After the trajectory is obtained, post processing like a smoothing technique is required before use as motor command input.

The proto-symbol based motion duplication for an FHMM is summarized as the following pseudo-code.

Step1 Expand an FHMM λ (M chains of N states) to an HMM λ' (N^M states)

Step2 Calculate the optimal state-sequence $S^* = [s_1^*, \cdots, s_T^*]$ given an observation for λ' by the conventional Viterbi algorithm

Step3 Convert S^* into $S = [S^1, \cdots, S^M]^T$ for M chains

Step4 Generate a motion by summing the contribution from each chain at each time step: $y_t = \sum_{m=1}^M B^m(s_t^m)$, for $0 < t < T$.

IV. EXPERIMENTS

The performance of the FHMMs is compared to the conventional HMMs for missing motion data recovery. Both models are tested on a data containing a series of 9 different human movement observation sequences obtained through a motion capture system. The human movement is converted into joint angle data for a 20 degree of freedom humanoid model. The data set contains multiple observations of walking (28 observations), cheering (15 observations), dancing (14 observations), kicking (19 observations), punching (14 observations), sumo leg raise motion (15 observations), squatting (14 observations), throwing (13 observations), and bowing (15 observations). Figure 3 shows selected frames of an animation for a walking motion pattern from the data set.

A set of 9 HMMs and FHMMs are trained on the data, one for each motion type. Each FHMM consists of two periodic left-to-right hidden Markov chains of 10 states each. The corresponding HMMs are also periodic left-to-right type with
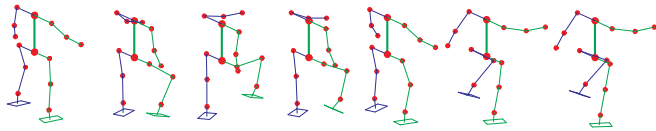
Fig. 3. Sample Walking Motion - Animated from joint angle data provided by the motion capture system. (1, 11, 21, 31, 41, 51 and 61 frames)
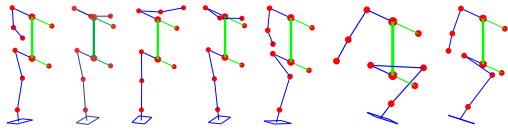

Fig. 4. Partially Observed Walking Motion - Animated from joint angle data. Right shoulder, arm, and leg are invisible. (1, 11, 21, 31, 41, 51 and 61 frames)
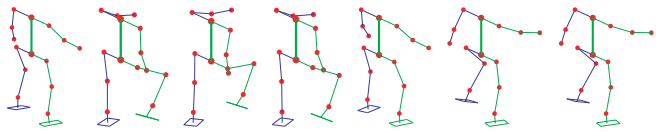

Fig. 5. Recovered Walking Motion using FHMM - Animated from recovered joint angle data using the trained walk motion model by the proposed method. (1, 11, 21, 31, 41, 51 and 61 frames)

different number of states. Among several training methods of FHMMs [24] [29] [30], the exact training method [24] was used in the experiments. To avoid overflow or underflow problem during implementing HMMs, it is often necessary to use a scaling parameter [32] or a logarithmic probability form. On top of these techniques, in a number of algorithms [33] [14] [15], the covariance matrix was constrained to be diagonal and the minimum covariance was constrained. A reason for the overflow or underflow problem is the HMM's over-fitting tendency, in particular with the large number of states. In the experiments, the full covariance matrix is used, without any constraint on the minimum covariance, during training in order to check the over-fitting problem. Kulić at el. [15] compared HMMs and FHMMs for recognition and generation, and suggested the use of FHMMs in a dense area of proto-symbol space to discriminate the proto-symbols clearly. This paper skips the recognition and generation performance comparison, but concentrates on missing data recovery performance.

In the first set of experiments, each FHMM consists of 2 chains of 10 states and the corresponding HMM contains 10 states. Each model is trained on 7 exemplars of a motion type. Input data is walking motion pattern whose right side of body was occluded, as shown in Fig. 4. The input motion is a different example which was not used for training. Figure 5 shows the snapshots from an animation of a recovered whole-body walking motion using the walking FHMM. Figure 6 shows an example of motion recovery results for the left and right knee joints during a walking motion. The shown

trajectory is mean value trajectory with respect to the optimal state-sequence given the observation, prior to applying any post processing (i.e., low-pass filtering). As can be seen in the figure, due to the higher number of states available to represent the motion, FHMMs achieve better spacial accuracy compared to single chain HMMs.
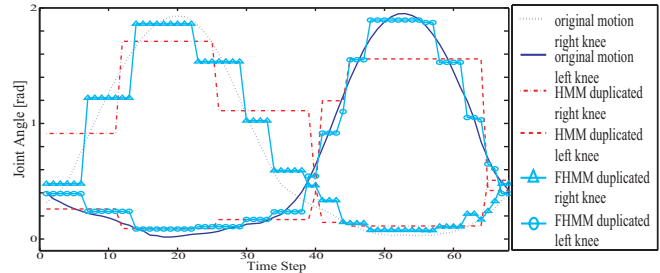

Fig. 6. Comparison of missing motion data recovery results of HMM (10 states) and FHMM (2 chains of 10 states) for the right and left knee joints during a walking motion. In observed data, joint angles for right half of the body was missing. Invisible right knee joint angle is recovered closely to the true value.

In the second set of experiments, each FHMM consists of 2 chains of 10 states and the corresponding HMM contains 10×2 states. They are compared with the same occluded walking motion data, as shown in Fig. 4. Figure 7 shows an example of motion recovery results for the left and right knee joints during a walking motion, prior to applying any post processing. Although both HMMs and FHMMs contain the same number of states (20 states), FHMMs achieve better spacial accuracy because FHMMs can represent 10^2 statuses by factorizing distributed states.
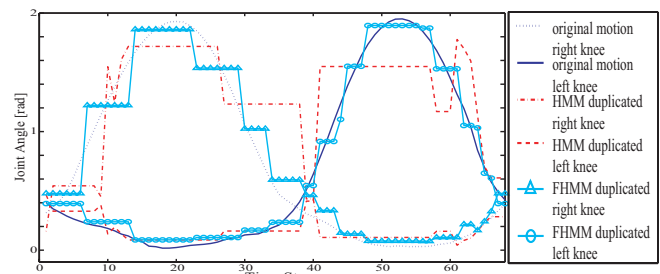

Fig. 7. Comparison of missing motion data recovery results of HMM (10×2 states) and FHMM (2 chains of 10 states) for the right and left knee joints during a walking motion, before any post processing is applied.

In the third set of experiments, each FHMM consists of 2 chains of 10 states and the corresponding HMM contains 10^2 states. The same occluded walking motion data, which is displayed in Fig. 4, is given as the recovery target. Figure 8 shows an example of the motion recovery results. The equivalent number of states of HMMs to FHMMs (2 chains of 10 states) is 10^2. Although both models have the equivalent number of states to represent the motion, it can be seen

that FHMMs outperform the HMMs from Fig. 8. A single hidden Markov chain with a large number of states is prone to over-fitting, therefore the HMM shows a performance drop when calculating the optimal state-sequence against new motion data. On the other hand, FHMMs demonstrate better generalization and good performance at optimal state-sequence calculation.
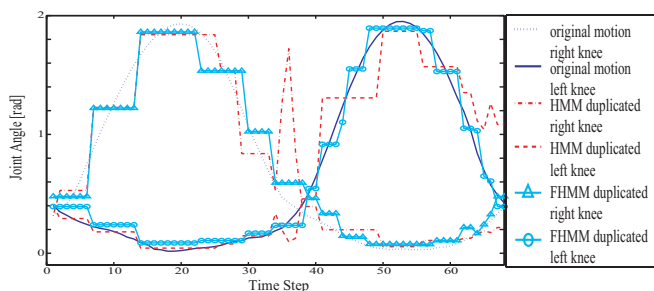

Fig. 8. Comparison of missing motion data recovery results of HMM (10^2 states) and FHMM (2 chains of 10 states) for the right and left knee joints during a walking motion, prior to any post-processing.

One of reasons for the computational overflow or underflow problem during training is the over-fitting. In the fourth set of experiments, the over-fitting tendency with respect to different model structures is surveyed. The following nine motion patterns are trained; walk, cheer, dance, kick, punch, sumo leg raise (SUMO), squat, throw and bow. Each motion pattern is trained into eight different model structures; periodic HMM with 10 states, periodic HMM with 20 states, periodic HMM with 100 states, periodic FHMM with 2 chains of 10 states, left-to-right(LR) HMM with 10 states, left-to-right HMM with 20 states, left-to-right HMM with 100 states, and left-to-right FHMM with 2 chains of 10 states. Training of each motion pattern into a model structure is run 1000 times. Among the 1000 trials of training, the rate of success without computational overflow or underflow problem is calculated and shown in table I. The full covariance matrix is used and there is no constraint on the minimum covariance during training. From the table, FHMMs which consist of 2 chains with 10 states are less prone to overfitting than HMMs with 10 states. The table shows that the parameters of HMMs with 10^2 states are hard to train. Even when the HMMs with many states are trained successfully, the trained HMMs do not represent abstract forms of motion patterns and the mimesis model using such HMMs becomes distant from the semiotic system.

In the fifth set of experiments, the recovered accuracy with respect to different model structures is surveyed. Multiple partial observations of nine motion patterns are recovered; walk (28 observations), cheer (15 observations), dance (14 observations), kick (19 observations), punch (14 observations), sumo leg raise (15 observations), squat (14 observations), throw (13 observations), and bow (15 observations). Among the 20

TABLE II
ERRORS OF THE RECOVERED MOTION [UNIT: RAD]

model	HMM 10 states	HMM 20 states	HMM 100 states	FHMM 2 chains, 10 states
walk	0.067345	0.057648	0.054110	0.043207
cheer	0.072403	0.064836	0.058101	0.064642
dance	0.076132	0.068385	0.051436	0.067164
kick	0.077991	0.073480	0.073277	0.067525
punch	0.075862	0.083210	0.053234	0.059824
SUMO	0.087385	0.068190	0.057218	0.065885
squat	0.063387	0.051660	0.061371	0.049082
throw	0.073762	0.063399	0.068947	0.054224
bow	0.066992	0.065015	0.062311	0.052278

joints, 8∼10 joints are invisible for each motion pattern. The invisible parts are the left half of the body for cheer and squat, the upper half for kick and sumo leg raise, and the lower body for punch and throw. In the case of walk, dance and bow, the right half is invisible. The difference between the true joint angles and the recovered joint angles is calculated. Table II shows the averaged errors of each joint at each frame. The table shows that FHMMs provide better recovery accuracy than the HMMs in general. Although sometimes HMMs with 100 states outperform FHMMs, HMMs with a large number of states are distant from the semiotic system and inappropriate for the mimesis model.

V. CONCLUSIONS

As an application of the mimesis model, a novel algorithm for missing data recovery using factorial hidden Markov models is proposed. The missing data recovery takes the same strategy as the whole body motion imitation from partial observations, but requires better motion representability. With the distributed state representation, FHMMs achieve better spatial representation with a small number of states and robust recognition performance against new data. A symbol based motion duplication algorithm is developed for FHMMs. In the experiments, the proposed algorithm is used to estimate the invisible joint angles of the human. From the results, it is shown that FHMMs are more flexible to new motion data, and improve the recovery accuracy compared to the HMMs.

REFERENCES

[1] G. Rizzolatti, L. Fadiga, V. Gallese, and L. Fogassi, "Premotor cortex and the recognition of motor actions," *Cognitive Brain Research*, vol. 3, pp. 131–141, 1996.
[2] G. Rizzolatti and L. Craighero, "The mirror-neuron system," *Annual Reviews of Neuroscience*, vol. 27, pp. 169–192, 2004.

TABLE I

TRAINING SUCCESS RATIO WITHOUT OVER/UNDERFLOW [UNIT: %]

model	periodic HMM 10 states	periodic HMM 20 states	periodic HMM 100 states	periodic FHMM 2 chains of 10 states	LR HMM 10 states	LR HMM 20 states	LR HMM 100 states	left-to-right FHMM 2 chains of 10 states
walk	33	15.6	3.6	44.6	29.7	20.7	0	61
cheer	100	100	32	100	100	99.4	0	100
dance	100	100	32.6	100	100	100	0.3	100
kick	100	98.8	29.6	100	99.9	99.5	0	100
punch	100	100	46.67	100	100	100	0	100
SUMO	100	100	33.67	100	100	99.8	0.2	100
squat	100	100	39	100	100	99.7	0	100
throw	100	100	41.8	100	100	99.8	0	100
bow	100	100	26	100	99.8	98.4	0	99.9

[3] D. C. Bentivegna and C. G. Atkeson, "Using primitives in learning from observation," in *First IEEE-RAS International Conference on Humanoid Robots (Humanoids 2000)*, 2000.

[4] D. C. Bentivegna, C. G. Atkeson, and G. Cheng, "Learning similar tasks from observation and practice," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'06)*, Oct. 10-13 2006, pp. 4994–5000.

[5] M. J. Matarić, "Getting humanoids to move and imitate," *IEEE Intelligent Systems*, vol. 15, no. 4, pp. 18–24, 2000.

[6] A. Billard and M. J. Matarić, "Learning human arm movements by imitation: Evaluation of biologically inspired connectionist architecture," *Robotics and Autonomous Systems*, vol. 37, pp. 145–160, 2001.

[7] ——, "Betty: Robot, play with me! robot: O.k. how do we play? betty: You watch me and do like i do. look!" in *Workshop on Interactive Robotics and Entertainment (WIRE-2000)*, Pittsburgh, 2000.

[8] A. Fod, M. J. Matarić, and O. C. Jenkins, "Automated derivation of primitives for movement classification," in *IEEE-RAS International Conference on Humanoid Robots (Humanoids 2000)*, Oct 2000.

[9] T. Inamura, Y. Nakamura, and I. Toshima, "Embodied symbol emergence based on mimesis theory," *International Journal of Robotics Research*, vol. 23, no. 4, pp. 363–377, 2004.

[10] S. Schaal, "Is imitation learning the route to humanoid robots?" *Trends in Cognitive Sciences*, vol. 3, no. 6, pp. 233–242, 1999.

[11] C. Breazeal and B.Scassellati, "Robots that imitate humans," *Trends in Cognitive Science*, vol. 6, no. 11, pp. 481–487, 2002.

[12] S. Schaal, A. Ijspeert, and A. Billard, "Computational approaches to motor learning by imitation," *Philosophical Transaction of the Royal Society of London: Series B, Biological Sciences*, vol. 358, pp. 537–547, 2003.

[13] T. Deacon, *The Symbolic Species: The Co-Evolution of Language and the Brain*. W.W.Norton and Company Inc, 1997.

[14] W. Takano, K. Yamane, T. Sugihara, K. Yamamoto, and Y. Nakamura, "Primitive communication based on motion recognition and generation with hierarchical mimesis model," in *IEEE International Conference on Robotics and Automation*, Orland, Florida, USA, May 2006, pp. 3602–3609.

[15] D. Kulić, W. Takano, and Y. Nakamura, "Representability of human motions by factorial hidden markov models," in *IEEE/RSJ International Conference on Intelligent Robots and Systems(IROS'07)*, 2007.

[16] 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.

[17] D. Kulić, W. Takano, and Y. Nakamura, "Incremental on-line hierarchical clustering of whole body motion patterns," in *IEEE International Symposium on Robot and Human Interactive Communication*, 2007.

[18] K. R. Dixon, J. M. Dolan, and P. K. Khosla, "Predictive robot programming: Theoretical and experimental analysis," *International Journal of Robotics Research*, vol. 23, pp. 955–973, 2004.

[19] S. Calinon, F. Guenter, and A. Billard, "On learning, representing and generalizing a task in a humanoid robot," *IEEE Transactions on Systems, Man and Cybernetics, Part B. Special issue on robot learning by observation, demonstration and imitation*, vol. 37, no. 2, pp. 286–298, 2007.

[20] Z. Ghahramani and M. I. Jordan, "Supervised learning from incomplete data via an EM approach," in *Advances in Neural Information Processing Systems*, J. D. Cowan, G. Tesauro, and J. Alspector, Eds., vol. 6. Morgan Kaufmann Publishers, Inc., 1994, pp. 120–127.

[21] D. Lee and Y. Nakamura, "Mimesis from partial observations," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'05)*, Edmonton, Canada, August 2005, pp. 1911–1916.

[22] ——, "Mimesis scheme using a monocular vision system on a humanoid," in *IEEE International Conference on Robotics and Automation (ICRA'07)*, Rome, Italy, April 2007, pp. 2162–2168.

[23] ——, "Motion capturing from monocular vision by statistical inference based on motion database: Vector field approach," in *IEEE/RSJ International Conference on Intelligent Robots and Systems(IROS'07)*, San Diego, USA, 2007, pp. 617–623.

[24] Z. Ghahramani and M. I. Jordan, "Factorial hidden markov models," in *Machine Learning*, vol. 29, 1997, pp. 245–275.

[25] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proc. IEEE*, pp. 257–286, 1989.

[26] K. Tokuda, H. Zen, and T. Kitamura, "Trajectory modeling based on hmms with the explicit relationship between static and dynamic features," in *Proceedings of European Conference on Speech Communication and Technology*, Geneva, Switzerland, Sept. 2003, pp. 1–4.

[27] W. Takano, H. Tanie, and Y. Nakamura, "Key feature extraction for probabilistic categorization of human motion patterns," in *the 12th IEEE International Conference on Advanced Robotics (ICAR'05)*, Seattle, USA, July 2005, pp. 424–430.

[28] A. P. Shon, J. J. Storz, and R. P. Rao, "Towards a real-time bayesian imitation system for a humanoid robot," in *IEEE International Conference on Robotics and Automation (ICRA'07)*, 2007, pp. 2847–2852.

[29] R. A. Jacobs, W. Jiang, and M. A. Tanner, "Factorial hidden markov models and the generalized backfitting algorithm," *Neural Computation*, vol. 14, pp. 2415–2437, 2002.

[30] D. Kulić, W. Takano, and Y. Nakamura, "Incremental learning of full body motions via adaptive factorial hidden markov models," in *7th International Conference on Epigenetic Robotics*, 2007.

[31] M. Umilta, E. Kohler, V. Gallese, L. Fogassi, and L. Fadiga, ""i know what you are doing": a neurophysiological study," *Neuron*, vol. 32, pp. 91–101, 2001.

[32] H. Juang, B. and R. Rabiner, L., "A probabilistic distance measure for hidden markov modeling," *AT&T Tech. J., vol. 64, no. 2*, vol. 64(2), pp. 391–408, 1985.

[33] D. Lee and Y. Nakamura, "Stochastic model of imitating a new observed motion based on the acquired motion primitives," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'06)*, Oct. 10-13 2006, pp. 4994–5000.