Segmentation, Memorization, Recognition and Abstraction of Humanoid Motions Based on Correlations and Associative Memory

Hideki Kadone and Yoshihiko Nakamura
Department of Mechano-Informatics, Graduate School of Information Science and Technology, The University of Tokyo, 7-3-1 Hongo, Bunkyo, Tokyo, Japan. {kadone,nakamura}@ynl.t.u-tokyo.ac.jp

Abstract—In order to self-organize symbols from observed motion patterns, it is necessary to temporally segment the continuous motion pattern flows into meaningful chunks. For reusability of the acquired information, repeatedly observed patterns are important, which means that segmentation, memorization, recognition and abstraction depend on each other. From this point of view, we propose methods for motion patterns of humanoid robots observed as a continuous flow using pattern correlations and associative memory. Initially, patterns are segmented by pattern correlations and then stored into the associative memory. Afterwards, only new kinds of motions are fed through this process. Associative memory is capable of segmentation, recognition and abstraction, and has ease in incremental update of the storage for new patterns.

I. INTRODUCTION

Symbols are crucial for intelligent systems. Extracting important informations from specific physical informations and memorizing them as abstract symbols enable one to apply the acquired information to other specific situations. Deacon[1] explains the developmental process of symbols as three stages which consists of icon, index and symbol. Icons refer to specific physical objects based on similarity of the patterns. Indexes are the relationships between icons based on physical causality. Symbols are referential relationships without physical causality. According to Deacon, developments from lower stages to higher stages emerge from learning associations in the lower stages. Memory systems for humanoid robots are expected to have these functions, to self-organize symbols from accumulated associations of specific patterns. As for the patterns, having our base on the concept of Mimesis[2], [4], which is about development of symbols from imitating others’ motion patterns, we use motion patterns of humanoid robots.

In the process of symbolization from physical experiences or pattern flows, it is necessary to find out “meaningful” chunks in the time domain from raw pattern flows, which is called segmentation. On the other hand, to evaluate and find out what is “meaningful” or not, recognition using memories and symbols organized from the accumulated experiences of the agent itself is necessary, which means segmentation, recognition, memorization and symbolization interact with each other. One of the possible criteria for “meaningful” or not is the encountering frequency of the pattern. If a pattern appears repeatedly, it is advantageous to remember it since the information can be reused at some time in the future in other situations. If it appears once and never appears again, it is meaningless to memorize the pattern. In addition, to make the representation efficient, the chunks should be as long in time as possible, while keeping the above sense.

There are some other methods already proposed on the segmentation of motion patterns, some based on patterns and others based on controls. These can be classified into the followings. (i) Competitive learning[3], [9], [10]. Initially randomized competitive learners are made adapt to the incoming pattern flows. Segments are the time points at which the corresponding learner is replaced by another one. The result depending on the matching of the initial distribution of the learners and that of the patterns to be learned, the problem with these methods is that it needs prior knowledge about the distribution of the incoming patterns. (ii) Generalization of teacher data[8]. These methods focus on teacher data and structures that generalizes better on some given criteria. We rather focus on the self-organizing and emergent properties of memories and symbols reflecting the inherent structures of motion patterns. (iii) Generalization of heuristic rules[5]. Some characteristic features in the flows are defined to be the boundaries of the segments. For example, drastic changes are adopted as the boundaries. This definition is inadequate since some kinds of motions should include drastic changes inside their segments.

In this paper, based on the idea that repeated patterns are important, we propose a system which incorporates segmentation, recognition, memorization and symbolization of motion pattern flows utilizing pattern correlations and an associative memory model. For symbolization, we use the dynamical mechanism of the associative neural network model[6], which forms abstract representations from accumulated specific memories of patterns, and hierarchically maintains specific representations and emergent abstract representations. Because the proposed method has its base on direct comparison of observed motion patterns, it does not depend on initial states of the model. The result totally depends on the inherent structure of the observed patterns. In addition, it has advantages in the stability of results by directly calculating the pattern correlations, and the constant computational burden against increasing number of memorized patterns by using dynamics of the associative memory model. Refer to [6] and our another presentation in this conference [7] for the detailed description of model and dynamics of the associative memory model.
II. SEGMENTATION BASED ON PATTERN CORRELATIONS
AND ABSTRACTION BY ASSOCIATIVE MEMORY

A. Segmentation based on pattern correlations

At the first stage, having no patterns in its memory and therefore being unable to do segmentation by memory, the initial process is the segmentation purely based on the pattern correlations. As stated earlier, assuming that repeated patterns has important information, if two highly correlated patterns appear in the pattern flow, the two patterns are memorized.

The correlation between the parts in the initial pattern flow is computed in the following way. Fig.1 depicts the concept. Let \( \theta[k] \in \mathbb{R}^{20} \) be the angular vector of the observed humanoid robot at time step \( k \). Like in Fig.1, two windows are set starting at the time \( b_1 \) and \( b_2 \), with the same length \( L \), having no overlap each other. The norm of the difference of the patterns in these windows are defined by the following equation.

\[
d(b_1, b_2, L) = \frac{1}{L} \sum_{l=0}^{L} ||\theta[b_1 + l] - \theta[b_2 + l]||
\]

The smaller the values is, the larger the correlation is. Minimal value for fixed \( L \) and variable \( (b_1, b_2) \) is defined.

\[
\hat{d}(L) = \min_{b_1, b_2} d(b_1, b_2, L)
\]

If highly correlated patterns appear repeatedly in the pattern flow, \( \hat{d}(L) \) becomes small if \( L \) is small enough, and rapidly becomes large at the instant when \( L \) goes across the length of the correlated patterns. Here, the range of \( L \) takes minimum at about 100msec, maximum at the possible largest length for the given pattern flow. Actually computing with real data shows that the curves of \( \hat{d}(L) \) has a characteristic tendency depicted in Fig.2. Hence we choose maximum \( L \) that satisfies \( \frac{d\hat{d}(L)}{dL} \leq 0 \) and \( L \leq L_0 \), where \( L_0 \) is the intersection point of \( \frac{d\hat{d}(L)}{dL} \) and the average value of the maximum and the minimum of \( \frac{d\hat{d}(L)}{dL} \). Once we have \( L \), we can obtain \( (b_1, b_2) \) by eq.(2), and therefore segments of motions. For the rest of the motion flow from which segments are removed, we again evaluate the curves of \( \hat{d}(L) \) and \( \frac{d\hat{d}(L)}{dL} \), and then segments. This procedure is repeated until the lengths of all the resting patterns do not show the characteristic curves of Fig.1, or the lengths of them becomes shorter than a some small value, for which we adopted 100msec.

The motion pattern flow used here is an about 66sec continuation of stepping, squat, stretch, bend, stretch, squat, bend, stepping, kick, stretch, squat, stretch, stretch, stepping, bend, bend, kick. It is made by converting motion capture data of a human motion into 20 DOF humanoid configuration, the procedure of which is the same for motion data used in the following sections. Even if the motions are labeled the same, the real patterns are different.

Fig.5 shows comparison between manual labeling and segmentation by the proposed algorithm. stp, sq, str and bn in the upper row stands for stepping, squat, stretch, and bend respectively. Same alphabets in the lower row shows segments that were found out to have large correlations and segmented out at the same iteration step. Except one error in J, mixing stepping and kick, the result is appropriate.
motions. of same kinds of motions, showing the inherent structure of are added manually. Cluster structure can be seen with clusters visualized by principal component analysis (PCA). The labels Feature vectors of the segmented motions are shown in Fig. 3.

Let \( \theta_i[k] \in \mathbb{R}^{20} \) be the angular vector of segmented motion \( i \) at time step \( k \). The feature vector \( m_i \) of each motion \( i \) is defined using auto-correlation matrix \( M_i(l) \) with temporal difference, which was reported to be able to clarify the global structure of motion patterns by showing cluster structures [6].

\[
M_i(l) = \frac{1}{T_i} \sum_{k=1}^{T_i} \theta_i[k] \theta_i^T[k-l]
\]  

where \( T_i \) is the length of the motion \( i \). About 66msec is adopted for \( l \). By rearranging the elements of matrix \( M_i \in \mathbb{R}^{20 \times 20} \), we obtain the feature vector \( m_i \in \mathbb{R}^{400} \). Feature vectors of the segmented motions are shown in Fig. 3 visualized by principal component analysis (PCA). The labels are added manually. Cluster structure can be seen with clusters of same kinds of motions, showing the inherent structure of motions.

B. Feature vectors of segmented motions by auto-correlation

C. Formation of symbolic representations by bifurcations of attractors in associative memory

By using dynamics of the associative memory model with nonmonotonic activation functions, we obtain symbolic encodings. In associative memory models, the target pattern to be recognized is given as the initial state of the dynamics, and the result is obtained as the attractor. When hierarchically
correlated patterns with hierarchical cluster structures are stored, this model shows bifurcations of attractors depending on the parameter \( h \) of the activation nonmonotoncity. Each stored pattern is the attractor for small \( h \). For large \( h \), each pattern not being the attractor, the attractors are at the centers of clusters[6]. \( h \) determines the size of the cluster to be integrated into one attractor and basin. For larger \( h \), larger clusters are integrated and vice versa. Therefore \( h \) corresponds to the height in the hierarchy. Attractors and basins at the centers of clusters are the symbolic representation of motions, integrating accumulated memories of motion patterns[6]. Fig.6 simply depicts the idea.

Connection weights \( W \) of the associative memory is determined by

\[
W = \frac{1}{N} \sum_i \xi_i \xi_i^T \tag{4}
\]

where \( N \) the number of neurons, which is 4000, and \( \xi_i \in \{-1, 1\}^{1000} \) is obtained by quantizing each element of \( m_i \) into 10 bit digits. Dependency of the attractors and the basins on \( h \) are obtained by simulation and shown in Fig.4. The initial states are the stored patterns. The space is equivalent to that of Fig.3. At \( h = 0.5 \), each stored pattern is an attractor. With increasing \( h \), symbolic attractors are formed, widening their basins, integrating the patterns in the clusters. At \( h = 2.5 \), clusters of stepping and kick are integrated, which gives more abstract representation.

III. SEGMENTATION AND RECOGNITION BASED ON ASSOCIATIVE MEMORY

A. Feature vectors of motion pattern flow

In order to use the associative memory model that is stored patterns as in section II-C, we have to define feature vectors of motion pattern flow to give the model as the initial state of the dynamics. To deal with various \( T \), which differs by motion segments, we use the average vector of various feature vectors obtained by setting windows of various lengths, at the newest part of the input flow with their centers aligned, the concept of which is depicted in Fig.7.

\[
\hat{M}(k, T, l) = \frac{1}{T} \sum_{s=k-T/2}^{s=k+T/2} \theta[s] \theta^T[s - l] \tag{5}
\]

Fig. 7. To deal with segments of different lengths, multiple windows with various lengths are set, at the newest part of the flow, with their centers aligned.

\[
M(k, l) = \frac{1}{C} \sum_{i=1}^{C} \hat{M}(k, D_i, l) \tag{6}
\]

where \( C \) is the number of the windows, and \( D_i \) is the length of the \( i \)th window. We use \( C = 10 \) and \( D_i \) is set from 1.5sec to 5sec with a constant interval.

B. Segmentation and recognition based on associative memory

By giving the feature vectors of motion pattern flows defined in the previous subsection as the initial states of the associative memory dynamics, the attractors to which they converge are obtained by simulation. The motion pattern flow used here is an about 30sec continuation of stepping, stretch, squat, bend, kick, squat, bend, stepping, stepping, stretch, kick, kick.

Fig.8 shows the attractors visualized in the same space as Fig.3. What is worth to be noted here is that the attractors are discrete in time, while the given feature vectors are temporally continuous and smooth. To identify to which attractors they converge along the time course, Fig.9 shows the 1st element of the vectors of attractors in the feature vector space. As can be inferred from Fig.8, discrete jumps from attractors to attractors can be seen. Comparison to the manual segmentation is shown by the dotted line and the alphabets above. Except at the beginning and at the end, where feature vector cannot be defined, the result is appropriate.
IV. SEGMENTATION AND ABSTRACTION OF UNKNOWN MOTIONS

A. Segmentation of unknown motion patterns

In this section, newly observed motion pattern flows are evaluated whether they are unknown to the system or not, and if the pattern is unknown, it is segmented based on the pattern correlations, and then stored in to the memory incrementally, which is described in the next subsection. Initialization by the method described in section II is assumed, which means that segmentation based on pattern correlations, storage of the patterns in to the connection weights of the associative model, and formation of symbols are completed for an initial motion flow.

Because of the basic properties of associative memory models, if the initial state is not among the stored patterns, the distance between the initial state and the attractor becomes large. Taking advantage of this mechanism, the observed pattern can be determined if it is a kind of known patterns or not, by measuring the traveled distance of the dynamics. Fig.10 shows the traveled distances between the initial states and the attractors along the motion pattern flow. By setting an appropriate threshold, unknown patterns are found out when the distance keeps being larger than the threshold for a certain period of time (100msec.).

The obtained unknown patterns are segmented by the method of the previous section based on pattern correlations, and stored into the associative memory. Since this method does not have no need to compare the input pattern to all of the stored patterns, the computational cost does not increase against increasing stored patterns.

The motion pattern flow used here is an about 90sec continuation of stepping, sumo-stomp, squat, punch, stretch, dance, squat, bend, sumo-stomp, bend, punch, stretch, squat, dance, bend, sumo-stomp, punch, stepping, punch, dance, kick, dance, squat, sumo-stomp. Compared to the one used in II-A, sumo-stomp, punch, dance, kick are added. Feature vectors of motions, including the ones newly segmented here, are shown in Fig.11 by the same procedure as in section II-B using eq.(3). Cluster structure of motions can be seen including the newly added motions.

B. Formation of symbolic representations in associative memory

As the same in section II-C, formation of symbolic representations in associative memory is described. When adding a new pattern to the storage, from eq.(4), connection weights of the associative memory is updated naturally by the next equation,

\[ W[n+1] = W[n] + \frac{1}{N} \sum_{ni} \xi_{ni} \xi_{ni}^T \]

where \( n \) is the number of renewal iterations and \( ni(i = 1, \ldots) \) is the index of motions added at the \( nth \) renewal.

For motions newly segmented out from the motion flow in section IV-A, feature vectors are obtained by eq.(3) and digitized and stored into the associative memory by eq.(7). Dependency of the initial states and attractors on the non-monotonicity parameter \( h \) is shown in Fig.12 by simulation. At \( h = 0.5 \), the attractors are the stored patterns. As \( h \) increases, stored patterns not being the attractors, attractors are formed at the centers of clusters. The larger the \( h \) is, attractors are formed integrating the higher level clusters. Compared to Fig.4, because of the increased number of storage patterns and clusters, this case shows more complex attractor structures integrating at more number of levels of abstraction.

V. CONCLUSION

We proposed a system that performs segmentation, memorization, recognition and abstraction of motion patterns of humanoid robots observed as a continuous flow, based on pattern correlations and associative memory. Firstly, an initial motion pattern flow is segmented based on pattern correlations, and then stored into the associative model, where symbolic representations are formed. Next, segmentation of motion pattern flow consisting known kinds of motions is possible.
utilizing the discrete property of attractors of the associative memory. Finally, using the dynamics of the associative memory model, newly observed motion pattern flows are evaluated whether they are unknown to the system, and if some of them is unknown, they are segmented and stored into the memory in an incremental way. In this way, new symbolic representations are formed according to the new motions, bringing more complex hierarchical representations. The result does not depend on the initial parameters of the model. Hierarchical abstract representations are formed in the memory, by bifurcations of attractors and their basins, from the specific level to various abstract levels, reflecting the inherent structure of motion patterns.

REFERENCES