Construction of Symbolic Representation from Human Motion Information

Araki, Yutaka
Department of Intelligent Systems, Kyushu University

Arita, Daisaku
Department of Intelligent Systems, Kyushu University

Taniguchi, Rin-ichiro
Department of Intelligent Systems, Kyushu University

Uchida, Seiichi
Department of Intelligent Systems, Kyushu University

他

出版情報：Lecture notes in computer science. 4252, pp.212-219, 2006-01. Springer Berlin, Heidelberg
バージョン：
権利関係：(c)2006 Springer

Construction of Symbolic Representation from Human Motion Information

Yutaka Araki, Daisaku Arita, Rin-ichiro Taniguchi, Seiichi Uchida,
Ryo Kurazume and Tsutomu Hasegawa

Department of Intelligent Systems, Kyushu University
6-1, Kasuga-koen, Kasuga, Fukuoka, 816-8580, JAPAN

Abstract In general, avatar-based communication has a merit that it can represent non-verbal information. The simplest way of representing the non-verbal information is to capture the human action/motion by a motion capture system and to visualize the received motion data through the avatar. However, transferring raw motion data often makes the avatar motion unnatural or unrealistic because the body structure of the avatar is usually a bit different from that of the human beings. We think this can be solved by transferring the meaning of motion, instead of the raw motion data, and by properly visualizing the meaning depending on characteristics of the avatar's function and body structure. Here, the key issue is how to symbolize the motion meanings. Particularly, the problem is what kind of motions we should symbolize. In this paper, we introduce an algorithm to decide the symbols to be recognized referring to accumulated communication data, i.e., motion data.

1 Introduction

Non-verbal information is very important in human communication, and video-based communication seems to be the simplest way. However, it has several problems, such as use of large network bandwidth, lack of spatioperceptual inconsistency, restriction of the number of participants, etc. As a possible solution to these problems, we are developing an avatar-based communication system[1]. It has an important merit that a virtual scene can be constructed as a communication environment, which can make the communication richer and efficient. Recently, this idea is extended to robot-based communication[2], where a robot is used instead of avatar and where a virtual communication environment is established in a physical 3D space.

In general, the avatar-based communication consists of acquisition of the contents of human communication, its transmission, and presentation of the transmitted contents. To present the non-verbal information, it seems that the simplest way is to capture the human action/motion by a motion capture system and to visualize the received motion data through the avatar. However, transferring the raw motion data causes several problems:

– The difference of body structure between a human and an avatar makes the reconstructed avatar motion unrealistic or physically impossible.

– The disturbance in communication network makes the avatar motion unnatural, because raw motion data is time synchronous data.

We think this can be solved by transferring the meaning of motion, instead of the raw data of motion, and by properly visualizing the meaning depending on characteristics of the avatar's function and its body structure. In this framework, the key issue is how to represent the meaning of motion referring to observed motion data, or how to symbolize the motion meaning. Particularly, the problem is what kind of motions we should symbolized. Of course, one method is to decide the symbols according to observation by ourselves, but it requires much time when we examine a large amount of accumulated communication data. Here, instead, we introduce an algorithm to decide the symbols to be recognized referring to accumulated communication data, i.e., motion data. Our basic idea is that frequent occurring motion patterns, i.e., motion motifs (or motifs for short), usually convey meaningful information, and that we automatically extract such motifs from the accumulated motion data.

2 Motif Extraction

To extract motifs[3], we propose a three-step procedure: the first step is compressing multi-dimensional motion information into lower-dimensional one by Principal Feature Analysis (PFA)[4]; the second is labeling time slices of each dimensional motion information according to its value and generating label sequences; the third step is recursive extraction of frequently occurring label patterns from multi-dimensional label sequences as motifs based on Minimum Description Length (MDL) principle[5]. Although motion motif extraction based on the MDL was used in [6], here, we deal with extraction of multiple motifs and explicit integration of multiple features.

2.1 Reduction of redundant dimension by PFA

Feature space reduction of high dimensional feature data such as human motion information is a common preprocessing step used for pattern recognition, clustering, compression, etc. Principal component analysis (PCA) and independent component analysis (ICA) have been extensively used for the space reduction. These methods find a mapping function from the original feature space to a lower space. However, they mix the original feature components and the original feature components are not handled directly. It is not easy to directly extract and describe motion patterns of subsets of body parts, such as a motion pattern of arms, or a motion pattern of legs.

Therefore, Principal Feature Analysis (PFA), which automatically determine a subset of feature components representing the original feature space, is used instead. Human motion information is described as a set of measured positions of body parts[1]. We represent each feature as a single vector $\mathbf{f}_i = [f_{i,1}\ f_{i,2}\ \cdots\ f_{i,n}]^T \in \Re^n$ and all motion information as a matrix $\mathbf{M} = [\mathbf{f}_1\ \mathbf{f}_2\ \cdots\ \mathbf{f}_s] \in \Re^{n \times s}$, where s is the number of features and n is the length of motion information. Here, we suppose each feature is normalized

[1] Each position is composed of three features, i.e., 3D spatial coordinates, (x, y, z)

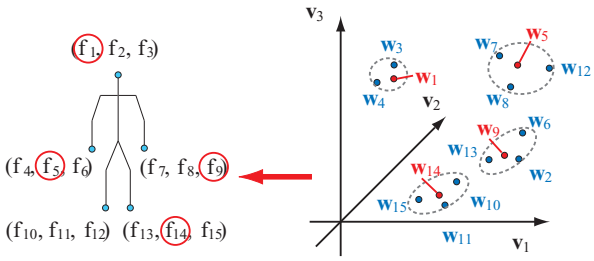

Fig. 1. Selection of principal features by PFA, where $q = 3$, $p = 4$. Each $\mathbf{w}_i$ written in red is closest to the mean of the cluster and corresponding $\mathbf{f}_i$ circled in red is a principal feature.

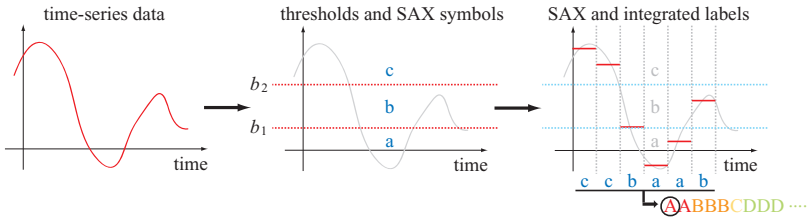

Fig. 2. Transforming a time-series data into a label sequence with SAX algorithm.

so that the average of the feature values is zero. Then, principle features are selected by three steps as follows.

First, the eigenvectors $\mathbf{v}_i \in \Re^s$ of $\mathbf{M}$ and their matrix $\mathbf{V} = [\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_s]$ are calculated. Second, principal components $\mathbf{M}_{pc} \in \Re^{n\times q}$ of $\mathbf{M}$ is calculated by the following equations:

$$\mathbf{M}_{pc} = \mathbf{M}\mathbf{W}^T \tag{1}$$

where $\mathbf{W} = [\mathbf{w}_1\ \mathbf{w}_2\ \cdots\ \mathbf{w}_s] \in \Re^{q\times s}$ is the first q columns of $\mathbf{V}^T$. q is decided according to the variability of all the features. Then, vectors $\{\mathbf{w}_i\}$ are clustered into $p(\geq q)$ clusters by K-Means algorithm using Euclidean distance. Here each vector $\mathbf{w}_i$ represents the projection of $\mathbf{f}_i$ to the principal component space.

Finally, in each cluster, the representing vector $\mathbf{w}_j$, which is the closest to the mean of the cluster, is selected. Thus, each $\mathbf{f}_j$ corresponding to $\mathbf{w}_j$ is selected a dominant feature. Figure 1 illustrates the concept of PFA.

2.2 Transformation of time-series data into label sequences

To reduce the computational complexity, time-series data $\{\mathbf{f}_j\}$[2] selected by PFA is transformed into a label sequence by Symbolic Aggregate approXimation (SAX)[7]. A label sequence is acquired by reducing the length of $\mathbf{f}_j$ and re-quantization. First, a time series data $\mathbf{f}_j$ of length n, i.e., n-dimensional vector $\mathbf{f}_j$ is regularly re-sampled into

[2] Subscript j indicates the label of the selected feature.

Fig. 3. Recursive motif extraction: (a)an original label sequence, (b)extracting the first motif, (c)extracting the second motif, (d)final segments.

a w-dimensional vector $\bar{\mathbf{f}}_j = [\bar{f}_{j,1}, \ldots, \bar{f}_{j,w}]^T$. The ith element of $\bar{\mathbf{f}}_j$ is the average of its corresponding interval of $\mathbf{f}_j$.

$\bar{f}_{j,k}$ is re-quantified to a label-based form, SAX symbol, by thresholding. Each $\bar{f}_{j,k}$ is symbolized to $\hat{f}_{j,k}$ as follows:

$$\hat{f}_{j,k} = \text{sym}_l, \text{ iff } b_{l-1} \leq \bar{f}_{jk} \leq b_l \quad (l = 1, \ldots, N) \tag{2}$$

where b_l is a threshold and sym_l is a SAX symbol. The thresholds are decided so that generation probability of each SAX symbol is equal to others, assuming that distribution of $\bar{f}_j$ is Gaussian. Then, each series of c SAX symbols, "$\hat{f}_{j,k} \cdots \hat{f}_{j,k+c-1}$" is assigned to a single unique label $l_{j,k}$, and a label sequence $\mathbf{L}_j = [l_{j,1} \cdots l_{j,w-c+1}]$ is constructed.

Finally, in order to reduce the influence of the variation in motion speed, $\mathbf{L}_j$ is compressed into $\hat{\mathbf{L}}_j$ by the run-length coding.

2.3 Recursive motif extraction based on MDL principle

For example, when a label sequence $\hat{\mathbf{L}}$ shown in Figure 3(a) is given, intuitively "BCD" can be a frequently occurring symbol pattern. Here, we need clear definition of what a frequently occurring pattern, or a motif, is, and we define the motif based on the MDL principle.

MDL is to find the best model which most efficiently compresses a label sequence by means that label patterns are replaced by unique meta-labels. We can consider that label patterns replaced by unique meta-labels in the best model are the motifs, because those label patterns are frequently occurred. In other words, the frequency of label patterns is evaluated based on the MDL.

To use the MDL principle, the description length of a label sequence should be defined, and it is defined based on a description model h and a label sequence $\hat{\mathbf{L}}$ as follows[6]:

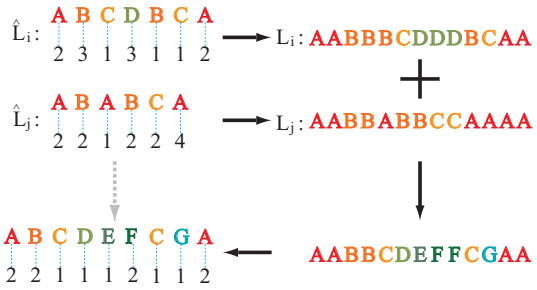

Fig. 4. Integration of $\hat{\mathbf{L}}_i$ and $\hat{\mathbf{L}}_j$.

$$DL = DL(\hat{\mathbf{L}}|h) + DL(h) \tag{3}$$

$$DL(\hat{\mathbf{L}}|h) = \sum_{i}^{m} \sum_{j} -w_{ij} \log_2 \frac{w_{ij}}{t_i} \tag{4}$$

$$DL(h) = \sum_{i}^{m} \log_2 t_i + m \log_2 (\sum_{i}^{m} t_i) \tag{5}$$

Where the model h is a segmentation of the label sequence $\hat{\mathbf{L}}$, m is the number of segments, w_{ij} is the frequency of the j-th label in the i-th segment, t_i is the length of the i-th segment.

Suppose t_L is the length of the label sequence $\hat{\mathbf{L}}$, the number of possible segmentations is $O(2^{t_L})$, which is too much computation to find the best segmentation in a naive manner. Therefore, we propose a sub-optimal method to solve this problem approximately by a recursive scheme as follows.

First, the most frequent label pattern, or a motif candidate, is searched by traversing a label sequence with a fixed size search window. By changing the size of search window, we can find the best pattern, or a motif, from all the selected candidates based on equation (3). The this process is show in Figure 3(a), and the total computation of finding a motif is $O(t_L^3)$.

Second, the selected frequent pattern, or the selected motif, is replaced by a unique meta-label, and the next motif is searched in the same way as the first step. Iterate the second step until no frequent pattern whose length is more than or equal to 2 is found. All the computation cost of finding possible motifs is $O(t_M t_L^3)$, where t_M is the number of the iterations. This process is illustrated in Figure 3(b), (c). When the process finishes, the all the motifs are detected as shown in Figure 3(d).

2.4 Integration of features

In the previous sections, we mentioned a method to extract motifs from one feature of human motion information. To analyze full-body motion information, it is necessary to integrate all the features. In this paper, simply, the integrated label sequences are

generated from the all combinations of each $\hat{\mathbf{L}}_i$ and the motifs are extracted from each integrated label sequence based on MDL. Each $\hat{\mathbf{L}}_i$ is integrated using corresponding $\mathbf{L}_i$. For example of integrating $\hat{\mathbf{L}}_i$ and $\hat{\mathbf{L}}_j$, each pair (l_{it}, l_{jt}) is re-labeled to the new label to be unique of the other pairs' shown in Figure 4. Where l_{it} correspond to tth element of $\mathbf{L}_i$.

Thus, when we have N features, we have $2^N - 1$ label sequences, and we extract motifs from each of the label sequence. After extracting motifs, we check similarity of the motifs and select proper motifs. However, there still remain improper motifs and we have to check them manually. Improvement of motif extraction accuracy is the future work.

3 Experimental results

We demonstrate the motif extraction method using motion data from Carnegie Mellon University's Graphics Lab motion-capture database [3]. The motion data used in this experiment is composed of 25 measured markers and 2486 frames. Each marker is composed of data of (x, y, z)-axis, i.e. the number of features is 75. The input motion data includes three basketball signals (A) waving hands up and down, (B) putting hands on the shoulders, and (C) thrusting hands forward shown in Figure 5(a), (b) and (c), each of which is repeated three times and connected by standing posture shown in Figure 5(d). In this experiment, the origin and the orientation of the coordinate system of motion information is fixed on a marker called "root" shown in Figure 6.

Five features shown in Figure 6 are selected by PFA: the left hand's x-axis, the left clavicle's y-axis, the right humerus' z-axis, the right elbow's y-axis and the right elbow's z-axis. In this experiment, the number of recursions of motif extraction is decided to eight empirically since it is enough for extracting all essential motions from the input motion information. For example, three motifs extracted from the integrated feature of (4) and (5) are shown in Figure 7. These motifs are extracted as the third, fifth and seventh motifs and correspond to motion (B), (A) and (C) respectively. The other motifs correspond to the standing posture. The third and seventh motifs are corresponding to the whole motion (B) and (C) shown in Figure 7(b),(c),(e),(f). However, the fifth motif is corresponding to a part of the motion (A) shown in Figure 7(a),(d). It is because motion (A), shaking hands up and down, is an iterative motion in comparison with motion (B) and (C) which are non-iterative. It is difficult to extract an iterative motion as a motif since an iterative motion includes sub-motifs, which are extracted as frequent patterns.

4 Discussion

We have proposed a motion motif extraction method from human motion information. The method automatically extracts all frequent motions as motifs from the whole body motion information. Our idea is that meaningful motion patterns, which are symbolized for avatar-based communcation, are frequently appeared, and that they can be extracted

[3] http://mocap.cs.cmu.edu/

(a) (b) (c) (d)

Fig. 5. The human motion scene: (a)waving hands up and down, (b) putting hands on the shoulders, (c)thrusting hands forward, (d)standing.

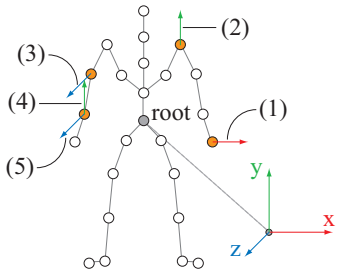

Fig. 6. Selected features: (1)left hand's x-axis, (2)left clavicle's y-axis, (3)right humerus z-axis, (4)right elbow's y-axis, (5)right elbow's z-axis.

by the proposed method. Of course, it is just preliminary study and we have to investigate the effectiveness of the idea, i.e., the following issues:

- The meaning of motion pattern sometimes differs depending on the context where the motion pattern appears. We should incorporate a context depending interpretation mechanism.
- Criteria for meaningful motion pattern other than the frequency of motion pattern should be investigated. Relationship between human posture and other persons (or objects) in the environment should be considered.

The preliminary experimental result of the motif extraction shows that the our method is effective to extract all motifs. However, there remain two problems. One is that our algorithm can not extract an iterative motion but just one cycle of the iterative motion. Another is that the computation cost is not small especially when we handle large motion databases including a variety of motion patterns, which are essentially high dimensional data.

Acknowledgment This work has been partly supported by "Intelligent Media Technology for Supporting Natural Communication between People" project (13GS0003, Grant-in-Aid for Creative Scientific Research, the Japan Society for the Promotion of Science), "Real-time Human Proxy for Avatar-based Distant Communication" (16700108, Grant-in-Aid for Young Scientists, the Japan Society for the Promotion of Science), and "Embodied Proactive Human Interface," (the Ministry of Public Management, Home

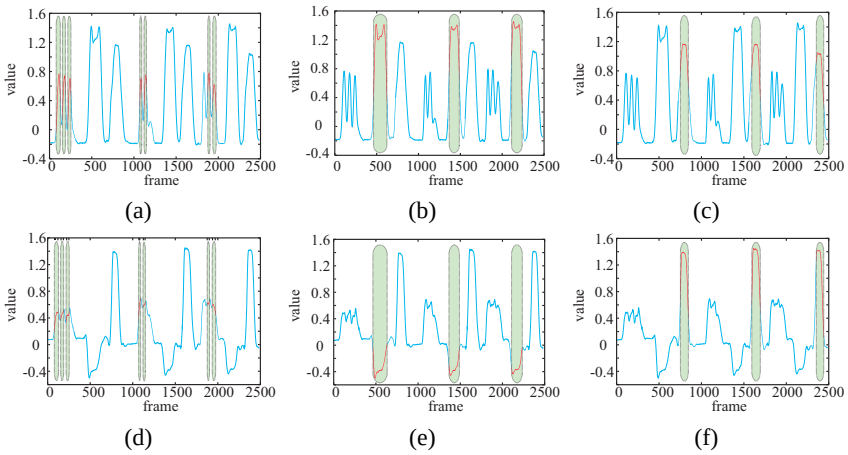

Fig. 7. The extracted 3rd(center), 5th(left) and 7th(right) motifs from integrated features: (a),(b),(c) illustrate y-position of the right elbow, (d),(e),(f) illustrate z-position of the right elbow.

Affairs, Posts and Telecommunications in Japan under Strategic Information and Communications R&D Promotion Programme (SCOPE)).

Our thanks to Carnegie Mellon University's Graphics Lab for allowing us to use their Motion Capture Database, which was supported with funding from NSF EIA-0196217.

References

1. D. Arita, H. Yoshimatsu, D. Hayama, M. Kunita, R. Taniguchi, Real-time human proxy: an avatar-based interaction system, *CD-ROM Proc. of Int. Conf. on Multimedia and Expo*, 2004.
2. A. Mori, S. Uchida, R. Kurazume, R. Taniguchi, T. Hasegawa, H. Sakoe, Early recognition of gestures, *Proc. of 11th Korea-Japan Joint Workshop on Frontiers of Computer Vision*, pp.80–85, 2005.
3. J. Lin, E. Keogh, S. Lonardi, P. Patel, Finding motifs in time series, *Proc. of the 2nd Workshop on Temporal Data Mining*, pp.53–68, 2002.
4. I. Chohen, Q. Tian, X. S. Zhou, T. S. Huang, Feature selection using principal feature analysis, Submitted to Int. Conf. on Image Processing '02, http://citeseer.ist.psu.edu/cohen02feature.html.
5. P. Grünwald, A tutorial introduction to the minimum description length principle, In *Advances in Minimum Description Length: Theory and Applications (edited by P. Grünwald, I. J. Myung, M. Pitt)*, MIT Press, 2005.
6. Y. Tanaka, K. Iwamoto, K. Uehara, Discovery of time-series motif from multi-dimensional data based on MDL principle, *Machine Learning*, vol.58, pp.269–300, 2005.
7. J. Lin, E. Keogh, S. Lonardi, B. Chiu, A symbolic representation of time series with implications for streaming algorithms, *Proc. of 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*, pp.2–11, 2003.