Motion motif extraction from high-dimensional 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

出版情報：Proceedings of the 12th Korea-Japan Joint Workshop on Frontiers of Computer Vision, pp.92-97, 2006-02
バージョン：
権利関係：

Motion motif extraction from high-dimensional motion information

Yutaka Araki† and Daisaku Arita† and Rin-ichiro Taniguchi†

†: Department of Intelligent systems, Kyushu University,
{araki, arita, rin}@limu.is.kyushu-u.ac.jp

Abstract Recently, there are a lot of researches on virtual environments for distant human communication. Real-time Human Proxy (RHP), which is a concept for such a virtual environment, has been proposed. For realizing natural communication by RHP, it is necessary to recognize human actions essential for human communication. However, it is difficult for system developers to decide which human actions should be recognized. For supporting the decision, we propose a human motion analysis method which automatically extracts frequent motion patterns as human action candidates. And, we show some experimental results for extracting human action.

1 Introduction

With developments in virtual reality techniques, a lot of researches on distant communication via a highly realistic virtual environment become more active. In a virtual environment, there are some CG characters, called avatars, each of which represents a participant of distant communication. There is no limit of the number of participants as well as the real world and each participant can understand positional relationship between avatars to intuitively recognize target avatars with which each participant communicate.

There are some communication systems using a virtual environment such as NICE (Narrative Immersive Constructionist / Collaborative Environments)[1] and Nessie World[2]. However, they have two problems that special input devices are required and their avatars are too simple to make natural communication. These are the bottleneck to achieve as natural communication via a virtual environment as face-to-face communication.

To solve these problems, as a fremework of avatar-based distant communication, we have proposed a concept of Real-time Human Proxy (RHP)[3], which virtualizes the avatar in real-time on which activities of a participant in the real world are projected. To achieve RHP-based communication, it is necessary to acquire information of a participant and recognize human information to generate symbol for operating an avatar. As the first step for realizing RHP-based communication, we are researching for dealing with motion information. To recognize the motions, it is important to decide which symbols are to be recognized as essential motion information for distant communication.

In this paper, we propose a method to extract essential motion patterns from human motion sequences.

2 Real-time Human Proxy (RHP)

2.1 RHP

RHP is proposed as a fremework of avatar-based distant communication, which recognizes essential information of a participant for communication as "symbols" from participant's information such as motions, voices, facial expressions, transfers these symbols among participants, and represents participant's information in a virtual environment based on transferred symbols in real-time.

2.2 Motion information acquisition system

Motion information is acquired by a Motion Capture System (MCS)[4], which can be used as input device for a natural human motion. Since it is not easy to decide which action symbols is to be recognized as essential actions for all situations of distant communication, it is necessary to establish a framework efficiently extracting essential actions from motion information according to situations. Our old approach is a top-down one, in which system constructors manually enumerate action symbols.

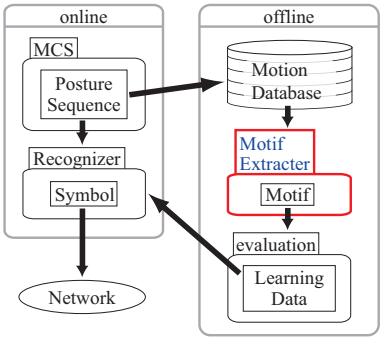


Fig. 1: Processing flow of the motion information acquision system.

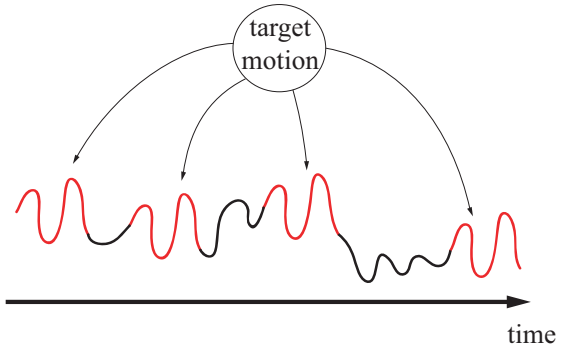


Fig. 2: Motif extraction from 1-dimentional data.

However, this approach is not robust and time-consuming. Therefore in this paper, we will propose a new bottom-up approach, in which a system automatically enumerate frequently occurring motion patterns, called motifs[5], from participant's motion information communicating face-to-face with others acquired in advance and system constructors choose action symbols from motifs. This approach is based on our idea that essential motions are repeatedly occurred as shown in Figure 2. Figure 1 shows the outline of the motion information acquisition system. In the online system, the real-time motion information is acquired by MCS and the recognizer make the symbol using the learning data prepared by the offline system. In the offline system, the acquired motion information is recorded to the motion database. The recorded motion information is analysed by the motif extracter to detect the frequent motion as the motif. To prepare the learning data, the extracted motifs are evaluated so as to select important ones.

3 Motif Extraction

This section explain how to extract motifs from human motion information. To extract motifs, we propose a three-step procedure as follows. The first step is compressing multi-dimensional motion information into lower-dimensional one by Principal Feature Analysis (PFA) [6] for reducing the amount of computation. The second step is labeling time slices of each dimensional motion information according to its value and generating label sequences, the number of which is same as that of the reduced dimensions. The third step is extracting frequently occurring label patterns from label sequences as motifs based on Minimum Description Length (MDL) principle [7] often used for data mining. These steps are explained in the following sections.

3.1 Reducing redundant dimension by PFA

Dimensional reduction of a high dimensional feature data such as a human motion information is a common preprocessing step used for pattern recognition, classification and compression schemes. Principal component analysis (PCA) and independent component analysis (ICA) have been extensively used for dimensional reduction. These methods find a mapping function from the original dimentional space to a lower dimentional space. However, it is disadvantageous in the motif extraction. Since a data in the lower dimensional space can't be decompressed, it is difficult to extract a partial motion from full-body motion. Therefore, in our work, we use PFA, which automatically determine a subset of axes, or features, that represents the original feature space. Human motion information is described as a set of measured positions of human body parts, each of which is composed of three spatial coordinates, (x, y, z). We treat each feature as a single vector $\mathbf{F}_i = [f_{1i}\ f_{2i}\ \cdots\ f_{ni}]^T \in \Re^n$ and all motion information as a matrix $\mathbf{M} = [\mathbf{F}_1\ \mathbf{F}_2\ \cdots\ \mathbf{F}_s] \in \Re^{s\times n}$, where s is a number of features and n is a length of motion information. Principle features are selected by three steps as follows.

First, the eigen vector $\mathbf{v}_i \in \Re^s$ of $\mathbf{M}$ is calculated by following equations:

$$\begin{aligned} C(\mathbf{M}) &= \mathbf{V}\Sigma\mathbf{V}^T \\ \mathbf{V} &= [\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_s] \\ \Sigma &= \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_s) \end{aligned} \tag{1}$$

where $C(\mathbf{M})$ is the covariance matrix, Σ is the diagonal matrix whose diagonal elements are the eigenvalues of $\mathbf{M}$, $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_s$.

Second, let $\mathbf{M}_{pc} \in \Re^{q\times n}$ be the Principal components of $\mathbf{M}$ as equation (2) , $q < s$ be set according to the degree how much variabilities of all features to be retained and $\mathbf{W} = [\mathbf{W}_1\ \mathbf{W}_2\ \cdots\ \mathbf{W}_s] \in \Re^{q\times s}$ be composed by the first q *columns* of the matrix

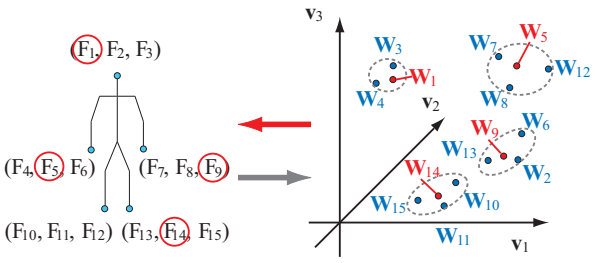


Fig. 3: An example of selecting principal features by PFA. Where the number of features is 15, $q = 3$, $p = 4$, (v_1, v_2, v_3) is principal component coordinate, each $\mathbf{W}_i$ written in red is closest to the mean of the cluster and corresponding $\mathbf{F}_i$ circled in red is principal feature.

$\mathbf{V}^T$,

$$\begin{aligned} \mathbf{M}_{pc} &= \tilde{\mathbf{M}}\mathbf{W}^T \\ \tilde{\mathbf{M}} &= [\tilde{\mathbf{F}}_\mathbf{1}\ \tilde{\mathbf{F}}_\mathbf{2}\ \cdots\ \tilde{\mathbf{F}}_\mathbf{s}] \\ \tilde{\mathbf{F}}_\mathbf{i} &= [\tilde{f}_{1i}\ \tilde{f}_{2i}\ \tilde{f}_{ni}]^T \\ \tilde{f}_{ji} &= f_{ji} - \frac{1}{n}\sum_{k=1}^{n} f_{ki} \end{aligned} \tag{2}$$

cluster each vector $|\mathbf{W}_i|$ to $p \geq q$ clusters by K-Means algorithm using Euclidean distance. Where each vector $\mathbf{W}_i$ represents the projection of $\mathbf{F}_i$ to the lowerthe principal component space.

Last, for each cluster, find the corresponding vector $\mathbf{W}_i$ which is closest to the mean of the cluster and the each corresponding feature $\mathbf{F}_i$ is chosen as one of the important subset features. Therefore, each $\mathbf{F}_i$ can be regarded as the dominant feature of its cluster and has the least redundancy of all combinations of features. The example of this method is shown in Figure 3.

3.2 Transforming time-series data into label sequences

Using the time-series data as an input, it takes too much computation amount to extract motifs from the human motion information. To solve this problem, it is necessary to transform the time-series data $\mathbf{F}_i$ selected by PFA into label sequences by reducing the length of $\mathbf{F}_i$ and quantization. Therefore, we use Symbolic Aggregate approXimation (SAX) [8] which is a symbolic representation method for a time-series data based on Piecewise Aggregate Approximation (PAA) as shown in Figure 4. First, with PAA, a time series data $\mathbf{F}$ of length n, i.e. n-dimensional vector $\mathbf{F}$ can be re-sampled into a w-dimensional vector $\bar{\mathbf{F}} = [\bar{f}_1, \ldots, \bar{f}_w]$. The ith element of $\bar{\mathbf{F}}$ is calculated by the following equation:

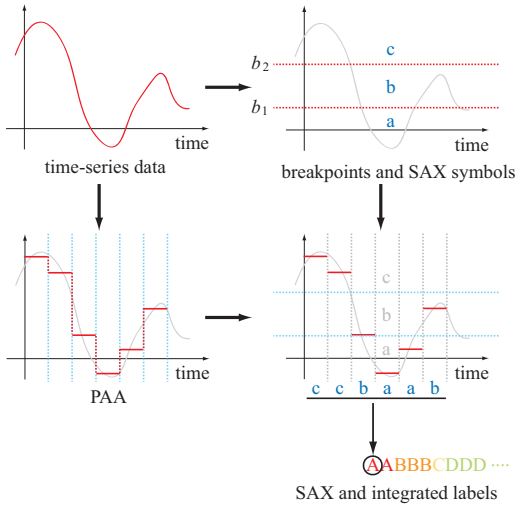


Fig. 4: The process of transforming a time-series data into a label sequence with SAX algorithm.

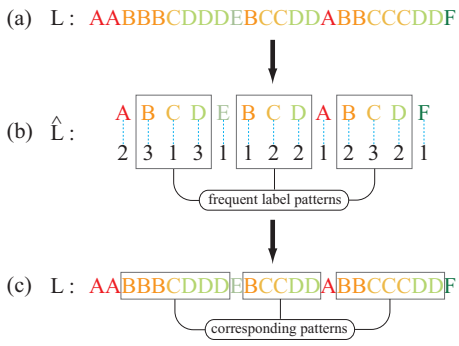


Fig. 5: Modifying the label sequence and finded frequent label pattern: (a) an original label sequence, (b) modified label sequence and frequent label pattern, (c)corresponding label patterns in the original one.

$$\bar{f}_i = \frac{w}{n}\sum_{j=\frac{n}{w}(i-1)+1}^{\frac{n}{w}i} f_j \tag{3}$$

Secondly, $\bar{f}_i$ is re-quantified to a label-based form, SAX symbol, by determing breakpoints. Using the breakpoints, each $\bar{f}_i$ is symbolized to $\hat{f}_i$ as follows:

$$\begin{aligned} \hat{f}_i &= \mathrm{sym}_j,\ iff\ \ b_{j-1} \leq \bar{f}_i \leq b_j \\ &\quad (j = 1, \ldots, a) \end{aligned} \tag{4}$$

Where b_j is a breakpoint and sym_j is a SAX symbol.The breakpoints are determined in order to SAX symbols are generated with same probability assuming that distribution of $\bar{f}_i$ is Gaussian. Thirdly, each series of b SAX symbols

Fig. 6: The recursive motif extraction: (a)an original label sequence, (b)extracting the first motif, (c)extracting the second motif, (d)final segments.

"$\hat{f}_i \ \cdots \ \hat{f}_{i+b-1}$" is assigned a single unique label l_i which is a minimum constituent of a motif. Then, a label sequence $\mathbf{L} = [l_1 \ \cdots \ l_{w-b+1}]$ is constructed. Fourthly, in order to avoid the influence of variations in motion speed, $\mathbf{L}$ is compressed into $\hat{\mathbf{L}}$ by the run-length coding (See Figure 5). $\hat{\mathbf{L}}$ has two kind of information; a label sequence and a numerical sequence which represents the length of the same label.

3.3 Motif extraction based on MDL principle

For example, given a label sequence $\hat{\mathbf{L}}$ as shown in Figure 5(b), "BCD" is to be a motif. However, too many motifs tend to be found in a human motion information. Therefore, it is necessary to evaluate motifs with a certain criterion in order for system constructors to choose essential motions from motifs. To evaluate motifs, we employ the MDL principle. MDL is to find the best model which most efficiently compresses label sequences by means that label patterns are replaced with one unique meta-label. We consider that the best model is proper for the motif.

To evaluate a frequent pattern based on MDL principle, the description length of the label sequence $\hat{\mathbf{L}}$ is needed as the criterion. The equation of the description length DL is defined using the description length of the model h and the label sequence $\hat{\mathbf{L}}$ as follows [9]:

$$DL = DL(\hat{\mathbf{L}}|h) + DL(h) \quad (5)$$

$$DL(\hat{\mathbf{L}}|h) = \sum_i^m \sum_j -w_{ij} \log_2 \frac{w_{ij}}{t_i} \quad (6)$$

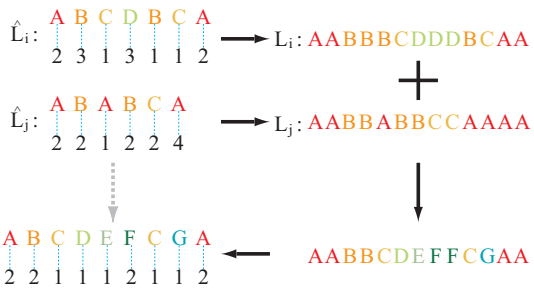


Fig. 7: The example of integrating $\hat{\mathbf{L}}_i$ and $\hat{\mathbf{L}}_j$.

$$DL(h) = \sum_i^m \log_2 t_i + m \log_2 (\sum_i^m t_i) \quad (7)$$

Where the model h is a segmentation of the label sequence $\hat{\mathbf{L}}$, m is the number of segments, w_{ij} is the number of frequency of the jth label in the ith segment, t_i is the length of ith segment. For example using $\hat{\mathbf{L}}$ in Figure 5(b), $h = \{1, 4, 5, 8, 9, 12\}$, $m = 7$, w_{11} is 2, t_1 is 2, w_{21} is 3, w_{22} is 1, w_{23} is 3, t_2 is 7 and so on.

Here, let t_L be the length of the label sequence $\hat{\mathbf{L}}$, the number of all segmentations is $O(2^{t_L})$, which is too much computation amount. Therefore, we propose a new method to solve this problem approximately by a recursive scheme as follows.

First, extract the frequent label pattern as a motif candidate using each size of the search window and choose the best pattern from the all candidates based on equation (5) as the first motif. The this process is show in Figure 6(a). The amount of computation of this process is $O(t_L^3)$.

Second, replace the chosen frequent label with a unique meta-label. If there is no frequent label pattern whose label length is more than 2, the recursive process is end.

Third, iterate the first and second processes and find the $i(\geq 2)$th motif including the labels without the replaced unique meta-labels that is shown in Figure 6(b),(c). When the processes is end, the all segments will be detected as shown in Figure 6(d). Letting t_M be the number of extracted motifs, the amount of computation of this method is $O(t_M t_L^3)$.

3.4 Integrating all features

In the previous sections, we mentioned the method for extracting motifs from one feature of human motion information. To analyze full-body motion information, it is necessary to integrate all features. In this paper, simply, the integrated label sequences are generated from the all combinations of each $\hat{\mathbf{L}}_i$

Table 1: The selected axis and each cluster's elements by PFA.

Cluster	Selected axis	Others
(1)	left hand's x	left elbow's x
(2)	left clavicle's y	head's x head's y ⋮
(3)	right humerus' z	left humerus' z right elbow's x
(4)	right elbow's y	left hand's y left elbow's y right hand's y
(5)	right elbow's z	left elbow's z left hand's z right hand's x right hand's z

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 7. Where l_{it} correspond to tth element of $\mathbf{L}_i$.

4 Experimental results

We demonstrate the motif extraction method using motion data from Carnegie Mellon University's Graphics Lab motion-capture database (http://mocap.cs.cmu.edu/). 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), (B) and (C) shown in Figure 8(a), (b) and (c), each of which is repeated three times and connected by standing posture shown in Figure 8(d). The motion (A) is waving hands up and down, (B) is putting hands on the shoulders and (C) is thrusting hands forward. 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 9.

Five features are selected by PFA shown in Figure 9 and the clustering result is shown in table 1. 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. As a result, the $(32-1)\times 8$ motifs are extracted, where 32 is the number of combinations of the five selected features and -1 means eliminating the combination that no features are selected. For example, three motifs extracted from the integrated feature of (4) and (5) are shown in Figure 10. 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 8(b),(c),(e),(f). However, the fifth motif is corresponding a part of the motion (A) shown in Figure 8(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 corresponding to one iteration.

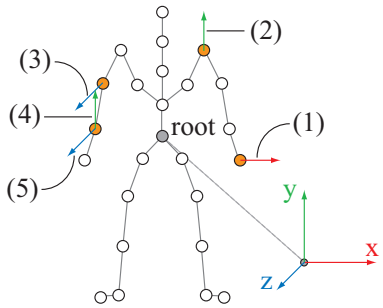
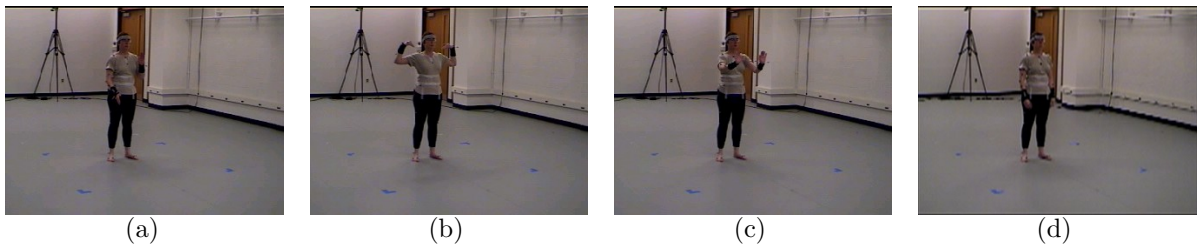


Fig. 9: Selected features and the coordinates: (1)left hand's x-axis, (2)left clavicle's y-axis, (3)right humerus z-axis, (4)right elbow's y-axis, (5)right elbow's z-axis.

5 Conclusion

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 whose dimensions are reduced by PFA. The experimental results show that the our method is effective to extract all motifs. However, there are the two problems. One is that the our method is difficult to extract an iterative motion. Another is that the number of combinations of selected features may become very large in case of extracting motifs from natural human motion information instead of artificial one used for our experiments.

Acknowledgment 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.



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

Fig. 8: The human motion scene: (a)waving hands up and down, (b) putting hands on the shoulders, (c)thrusting hands forward, (d)standing.

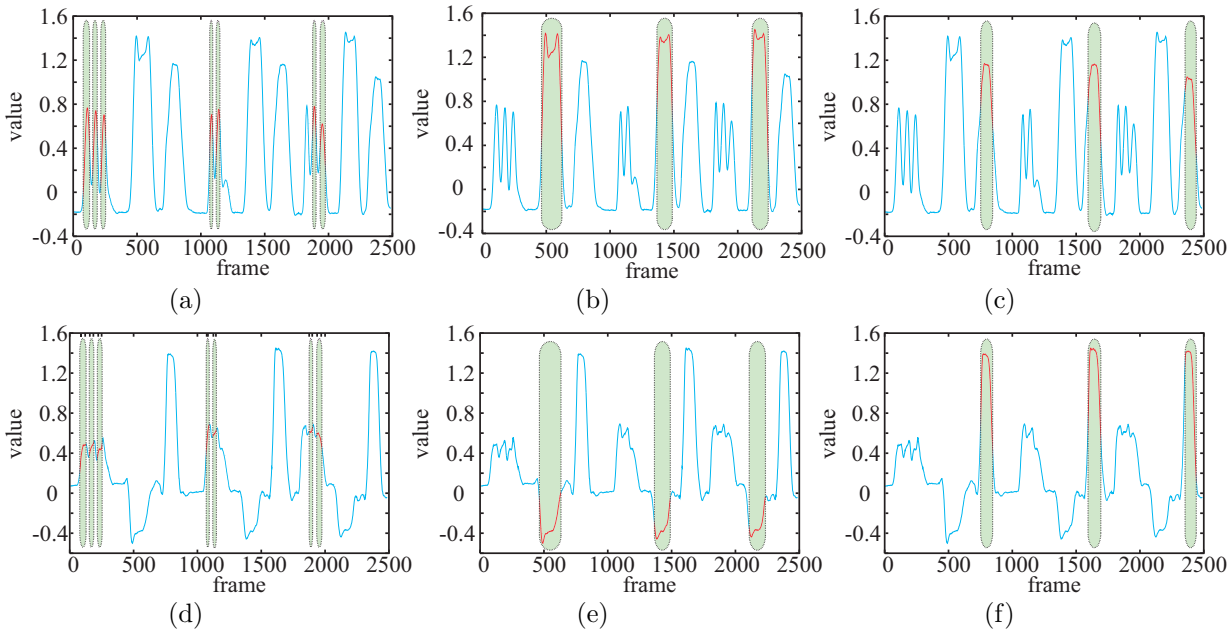


Fig. 10: The extracted 3rd(center), 5th(left) and 7th(right) motifs from an integrated feature: (a),(b),(c) illustrate y-position of the right elbow, (d),(e),(f) illustrate z-position of the right elbow.

References

[1] M. Roussos, A.E. Johnson, J. Leigh, C.A. Vasilakis, C.R. Barnes, T.G. Moher, "NICE: combining constructionism, narrative and collaboration in a virtual learning environment," ACM SIGGRAPH Computer Graphics, Volum 31, Issue 3, pp.62–63, 1997.

[2] P. Jeffrey, A, McGrath, "Sharing serendipity in the workplace," Proc. of the third international conference on Collaborative virtual environments, pp.173–179, 2000.

[3] D. Arita, H. Yoshimatsu, D. Hayama, M. Kunita, R. Taniguchi, "Real-time Human Proxy: An Avatar-based Interaction System," CD-ROM Proc. of International Conference on Multimedia and Expo, 2004.

[4] N. Date, H. Yoshimoto, D. Arita, R. Taniguchi, "Real-time Human Motion Sensing based on Vision-based Inverse Kinematics for Interactive Applications," ICPR'04, vol.3, pp. 318–321, 2004.

[5] 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.

[6] Chohen I., Tian Q., Zhou X. S., and Huang T. S. "Feature selection using principal feature analysis," ICIP'02, 2002.

[7] P.D. Grunwald, "A tutorial introduction to the minimum description length principle," MIT Press, April 2005.

[8] J. Lin, E. Keogh, S. Lonardi, B. Chiu, "A Symbolic Representation of Time Series, with Implications for Streaming Algorithms," DMKD03, pp.2–11, 2003

[9] TANAKA.Y., IWAMOTO.K., & UEHARA.K., "Discovery of Time-Series Motif from Multi-Dimensional Data Based on MDL Principle," Machine Learning Vol.58 pp.269-300, 2005.