# MULTI-LAYER GROWING NEURAL GAS FOR MOTION ANALYSIS IN EATING SCENES

**TAKENORI OBO[1], ANGGA WAHYU WIBOWO[1], ERI SATO-SHIMOKAWARA[1], KURNIANINGSIH[2], HIROKI SHIBATA[1], ICHIRO KOBAYASHI[3], YIHSIN HO[3]**

[1] Graduate School of Systems Design, Tokyo Metropolitan University, Tokyo, Japan
[2] Department of Electrical Engineering, Politeknik Negeri Semarang, Semarang, Indonesia
[3] Takushoku University, Faculty of Engineering, Tokyo, Japan
E-MAIL: t.obo@tmu.ac.jp, angga-wahyu-wibowo@ed.tmu.ac.jp, eri@tmu.ac.jp, kurnianingsih@polines.ac.id, hshibata@tmu.ac.jp, ikobaya@la.takushoku-u.ac.jp, hyihsin@es.takushoku-u.ac.jp

**Abstract:**

**To facilitate effective international exchange, a precise understanding of differences in behavioral patterns across cultures is essential. This study focuses on mealtime behavior as a culturally significant daily activity that reflects diverse customs and tool usage. Motion data were collected from Japanese and Indonesian university students during lunchtime, with the aim of extracting characteristic behavioral patterns and archiving them as representative examples of eating behavior among students. For this purpose, we employed Growing Neural Gas (GNG), a topological clustering method that dynamically adapts to data structure and enables visualization of motion postures. Furthermore, a Multi-Layer GNG approach is introduced to perform multi-level analysis of behavioral patterns with varying degrees of granularity. The proposed framework contributes to the digital representation of cultural behaviors, supporting experiential learning and promoting cross-cultural understanding.**

**Keywords:**

**Intercultural Understanding; Motion Extraction; Growing Neural Gas;**

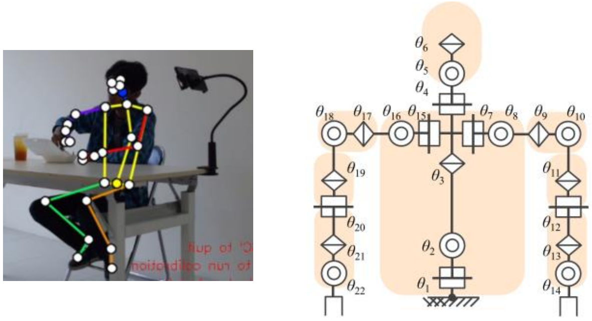
## 1. Introduction

In facilitating effective international exchange, it is essential to develop an accurate understanding of the differences in customs and behavioral patterns across cultures. Regional backgrounds and religious traditions significantly influence individuals' daily activities and the ways in which tools are used. With the advancement of information and communication technology (ICT), the importance of building international relationships that transcend geographical boundaries is expected to increase further. In this context, fostering globally competent individuals, establishing frameworks for the acceptance of foreign human resources, and promoting multicultural coexistence have become pressing societal challenges.

This study focuses on behavioral patterns, which are distinctive and commonly observed behaviors within specific groups or social contexts, as a means to deepen cross-cultural understanding. Observing or experiencing these behavioral patterns in different cultural environments can serve as an effective entry point for intercultural comprehension [1-3]. The aim of this study is to digitally capture, visualize, and share these behavioral patterns in order to support experiential learning and promote intercultural awareness and acceptance.

Among various daily activities, mealtime behavior represents a particularly illustrative setting in which diverse practices and tool usage can be observed. In particular, the way individuals use utensils, such as chopsticks, spoons, forks, plates, bowls, and cups, varies depending on both the shape of the objects and the background of the user, resulting in a wide range of grasping postures. Previous studies have conducted detailed motion analysis of hand movements during Japanese and Western-style meals. By extracting the spatiotemporal relationships between grasping postures and tableware, these studies visualized the differences in eating styles [4]. Furthermore, since eating practices differ considerably across regions and cultures, knowledge of these differences constitutes an essential foundation for fostering intercultural understanding.

In this study, motion data during lunchtime were collected from Japanese and Indonesian university students, with the aim of extracting characteristic postures and movements and archiving them as representative eating behavior patterns among university students. For feature extraction, we applied Growing Neural Gas (GNG), a type of topological clustering method [5]. Unlike conventional topological mapping techniques such as Self-Organizing Maps, GNG flexibly represents spatial features by



(a) (b)

**FIGURE 1.** An example of (a) motion data captured using VisionPose and (b) kinematic model.

dynamically adding and removing nodes and edges that constitute its topological structure, thus adapting to the underlying data distribution. The topological structure generated by GNG is composed of reference vectors derived from motion data, which enables direct visualization of motion postures. Furthermore, this study introduces a Multi-Layer GNG to present multi-level analyses that capture movement features at varying levels of granularity.

The rest of this paper is organized as follows. Section II provides an explanation of the data collection. Section III details the proposed approach for motion extraction. Section IV presents an experimental case study on the extraction of eating behavior patterns among university students. Finally, Section V summarizes the paper while presenting future work.

## 2. Data Collection

In this study, motion data were collected using VisionPose, a markerless motion capture system developed by Next-System Co., Ltd. VisionPose is a software engine capable of three-dimensional skeletal estimation and tracking based on image data obtained from a stereo camera. The acquired skeletal data consist of the three-dimensional coordinates of 30 joint points. An example of the captured motion is shown in Fig. 1(a).

Given that variations in individual body scales can influence motion analysis based on joint coordinates, this study utilizes joint angles estimated from the VisionPose data. Following previous research, joint angle estimation was performed using a genetic algorithm to solve the inverse kinematics optimization problem [6]. As shown in Fig. 1(b), the estimated posture is represented by 22 degrees of freedom (DOF) in the upper limbs. The link lengths of the kinematic model were determined in advance by capturing the subject in a T-pose and computing the Euclidean distances between adjacent joints.

## 3. Multi-Layer Growing Neural Gas

Hierarchical learning structures have been proposed in various topological clustering algorithms, such as Growing Neural Gas (GNG) and Self-Organizing Maps (SOM) [7]. Hierarchical learning approaches can generally be categorized into bottom-up and top-down strategies. This study adopts a bottom-up approach, in which node abstraction is achieved by progressively reducing the number of nodes in higher layers. In this process, the nodes in an upper layer are updated using the reference vectors obtained from the layer directly below. On the other hand, in the top-down approach, the nodes in an upper layer serve as input vectors for training the lower layer, allowing the lower layer to learn local topological structures around the reference vectors extracted in the upper layer.

In the proposed method, the reference vectors in the lowest layer are updated using raw data samples, while each subsequent layer uses the output nodes of the preceding layer as input data. This hierarchical structure enables each higher layer to abstract the topological structure of the layer below while preserving its essential characteristics. To control the granularity of clustering, the maximum number of nodes is progressively reduced at each higher layer, typically by half (e.g., 200 in the lowest layer, 100 in the second layer, and 50 in the third layer). This constraint leads to sparser representations in the upper layers and facilitates the formation of more generalized spatiotemporal feature clusters.

In this study, twenty two dimensional joint angle vectors were calculated from the joint position data obtained using VisionPose, as illustrated in Figure 1(b), and these were used as reference vectors for learning the topological structure based on GNG.

## 4. Experiment

### 4.1. Setting

In this study, motion data during lunchtime were collected from Indonesian and Japanese university students in order to extract characteristic features of their respective eating behaviors.

First, for the Indonesian participants, motion was recorded while eating typical local dishes, namely Nasi Goreng, Mie Goreng, and Nasi Ayam. To extract common motion patterns related to utensil use and eating styles for each dish, participants were instructed to eat Nasi Goreng with a spoon, Mie Goreng with a fork, and Nasi Ayam using their bare hands. The participants consisted of four Indonesian university students, including two male and two

**TABLE 1.** Parameter setting

| Parameters | Values |
|---|---|
| Learning rate for the first winner (linearly decreasing based on the number of learning trials) | 0.05 |
| Learning rate for the second winner | 0.0025 |
| Maximum number of nodes for 1st layer | 20 |
| Maximum number of nodes for 2nd layer | 10 |
| Maximum number of nodes for 3rd layer | 5 |
| Maximum number of nodes for 4th layer | 3 |

female students. Each participant was recorded once for each dish, resulting in three motion recordings per person. The duration of each recording was set to one minute per dish.

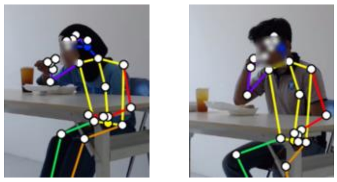
Next, an experiment was conducted with four Japanese university students, all of whom were male. Each participant engaged in three eating sessions: one using chopsticks, one using bare hands, and one using a spoon. Following the same procedure as with the Indonesian participants, each student was recorded once for each eating style, with one minute allocated to each recording.

Motion data for all participants were collected using VisionPose, and joint angles were estimated from the skeletal data. The number of motion data samples collected from both Indonesian and Japanese participants was 18,000. Feature extraction was then performed using the proposed hierarchical Growing Neural Gas method. The unsupervised training was performed for 10 iterations, with each iteration processing the entire dataset as a single batch. The main parameter settings used in the proposed method are summarized in Table 1. All parameter values were determined empirically.

### 4.2. Observed Characteristics of Eating Behavior

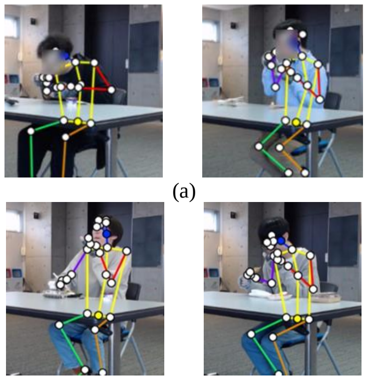
In Indonesia, due to religious and cultural norms, the use of the left hand during meals is generally avoided. Eating is typically performed using only the right hand. Although supportive actions such as holding a container with the left hand may occasionally occur, for example when the lid of a lunch box interferes with eating, the left hand does not come into direct contact with food. As shown in Fig. 2, Indonesian students are observed eating with the left hand resting on the lap, using only the right hand for handling utensils or eating with bare hands.

In contrast, in Japanese eating customs, the use of both hands is socially accepted and is even regarded as proper etiquette in certain situations. For example, when eating from a small bowl, it is considered polite to lift the bowl with one hand while eating with the other. Accordingly, in the recorded motion data, it was observed that when using chopsticks or a spoon, participants consistently supported the container with



(a) (b)

**FIGURE 2.** An example of Indonesian students' motion captured using VisionPose.



(a)

(b)

**FIGURE 3.** Eating motions of Japanese students captured using VisionPose: (a) lifting the container close to the mouth and (b) using the non dominant hand to prevent food from falling or spilling.

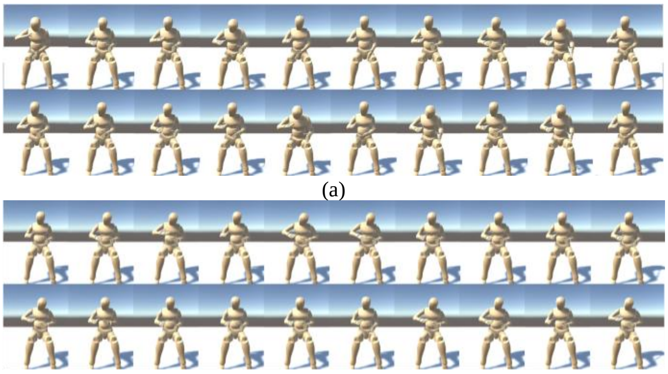
the non dominant hand. For foods that are more likely to spill, participants often lifted the container closer to the mouth to facilitate eating, as shown in Fig. 3(a).

Furthermore, several behaviors were observed among the Japanese participants that were not present in the Indonesian data. These include covering the mouth with the non dominant hand to prevent food from spilling out of the mouth, or using the left hand to catch falling food and prevent it from dropping onto the table (Fig. 3(b)). These behaviors reflect cultural differences in the role and acceptability of hand use during meals.

### 4.3. Motion Analysis with Multi-Layer GNG

The purpose of conducting motion analysis using the Multi Layer GNG algorithm is twofold: first, to evaluate whether representative postures and movements can be extracted from the reference vectors, and second, to examine the adaptability of multi layer structures in capturing features from varying levels of granularity.

Figure 4 presents visualized postures generated using a humanoid model based on the reference vectors obtained from the lowest layer of the Multi Layer GNG. The figure displays 20 postures, corresponding to the upper limit of nodes in the lowest layer. As shown, postures extracted from the Japanese participants consistently depict the left hand positioned near the food container, supporting or stabilizing it in coordination with the right hand (Fig. 4(b)). In contrast, the postures derived from Indonesian participants show more variation in the position of the left hand (Fig. 4(a)), including instances where the hand rests on the lap. This difference reflects cultural tendencies in hand usage during eating. On the other hand, postures involving mouth-covering gestures by Japanese students, as mentioned in the previous section, were not extracted. This limitation is likely due to the granularity of the model, which is influenced by the predefined number of nodes. Furthermore, Figure 5 presents postures visualized based on the reference vectors from the top layer of the Multi Layer GNG. Compared to the lower layers, these postures are more abstract, and it can be observed that there is little variation among the three extracted postures. Nevertheless, differences between the Indonesian and Japanese participants are still evident, particularly in the positioning of the left hand.



(a)

(b)

**FIGURE 4.** Postures visualized from the lowest layer of the Multi Layer GNG: (a) Indonesian participants, (b) Japanese participants.

(a) (b)

**FIGURE 5**. Postures visualized from the top layer of the Multi Layer GNG: (a) Indonesian participants, (b) Japanese participants.

## 5. Conclusion

This study proposed a motion analysis method for eating behavior with the aim of supporting cross-cultural understanding and visualizing differences in cultural habits and movement patterns. Specifically, a Multi Layer GNG algorithm was employed to enable feature extraction at multiple levels of granularity.

Motion data were collected from Japanese and Indonesian university students during lunchtime to examine culturally distinctive behaviors. The experimental results revealed that Indonesian participants ate exclusively with their right hand due to religious reasons, whereas Japanese participants commonly held food containers with one hand while eating with the other. Additionally, certain behaviors were observed only among Japanese participants, such as covering the mouth with the non dominant hand when food was likely to spill, or using the hand to prevent food from falling onto the table. Furthermore, using the Multi Layer GNG approach, representative postures during eating were extracted as reference vectors from the recorded motion data.

As a future direction, we plan to visualize these culturally distinctive movement patterns and postures to develop an archive of standard eating behaviors in different cultures, thereby contributing to systems that promote intercultural understanding.


## Acknowledgements

This work was supported by JSPS KAKENHI Grant Numbers 22KK0162.


## References


[1] D. Archer, "Unspoken diversity: Cultural differences in gestures." Qualitative Sociology, vol. 20 pp.79-105, 1997.

[2] M. LaFrance, and C. Mayo. "Cultural aspects of nonverbal communication." International Journal of Intercultural Relations, vol. 2 (1), pp. 71-89, 1978.

[3] L. M. Forsell and J. A. Åström, "Meanings of hugging: From greeting behavior to touching implications." Comprehensive Psychology, vol. 1, pp.2-17, 2012.

[4] T. Obo et al., "Human Pose Estimation with Multi-Camera Localization Using Multi-Objective Optimization Based on Topological Structured Learning, " Journal of Advanced Computational Intelligence and Intelligent Informatics, vol.27 no.4, pp. 543-553, 2023.

[5] B. Fritzke, "A Growing Neural Gas Network Learns Topologies, "In Proceedings of the 7th International Conference on Neural Information Processing Systems, 1994.

[6] T. Obo, K. Hamada, T. Matsuda and N. Kubota, "Multilayer Topological Clustering for Human Motion Segmentation, " In Proceedings of the 2024 IEEE International Conference on Systems, Man, and Cybernetics (SMC 2024), 2024.

[7] A. Rauber, D. Merkl, and M. Dittenbach. "The growing hierarchical self-organizing map: exploratory analysis of high-dimensional data." IEEE Transactions on Neural Networks, vol. 13, no. 6, pp. 1331-1341, 2002.