https://doi.org/10.20965/jaciii.2019.p0519

Paper:

Human Posture Recognition for Estimation of Human Body Condition

Wei Quan*, Jinseok Woo*, Yuichiro Toda, and Naoyuki Kubota***

*Graduate School of Systems Design, Tokyo Metropolitan University
6-6 Asahigaoka, Hino, Tokyo 191-0055, Japan
E-mail: quan-wei1@ed.tmu.ac.jp, {woojs, kubota}@tmu.ac.jp
**Graduate School of Natural Science and Technology, Okayama University
3-1-1 Tsushima-Naka, Kita, Okayama, Okayama 700-8530, Japan
E-mail: ytoda@okayama-u.ac.jp
[Received November 30, 2018; accepted December 25, 2018]

Human posture recognition has been a popular research topic since the development of the referent fields of human-robot interaction, and simulation operation. Most of these methods are based on supervised learning, and a large amount of training information is required to conduct an ideal assessment. In this study, we propose a solution to this by applying a number of unsupervised learning algorithms based on the forward kinematics model of the human skeleton. Next, we optimize the proposed method by integrating particle swarm optimization (PSO) for optimization. The advantage of the proposed method is no pretraining data is that required for human posture generation and recognition. We validate the method by conducting a series of experiments with human subjects.

Keywords: human posture recognition, growing neural gas, particle swarm optimization, human-robot interaction

1. Introduction

Human posture recognition has been a popular research topic ever since first computer computational recognition method appeared, and many state-of-the-art methods have been proposed by researchers and engineers around the world. It has been applied to various areas, such as human-robot interaction [1–3], operating simulations, and games development.

Human-robot interaction might be the area with the most applications of posture recognition. It makes it possible for humans to communicate with robots not only though cold commands inputted from the keyboard, but also though gesture language understood by computers [1, 4].

At the same time, with the increase of the number of elderly people all over the world, health care issues are getting more and more important. To solve these issues, human motion capture systems are required for the elders who live alone. And elders' state of health can be monitored by determining his or her posture, and an alert can be given in the case that high risk postures, such as fall down, are detected. These series of systems will reduce the burden of human resources while improving the efficiency of posture recognition [5].

Previous human skeleton recognition research was conducted using standard pin-hole digital cameras. However, it is impossible to detect a human's real spatial posture during a period of time owning to the natural weakness of the pinhole camera. The recently developed electronic devices, such as the RGBD camera, stereo camera, and depth sensors, have made it possible to capture objects in a real-world shape. In recent years, the most widely used sensors in the referent field are depth cameras.

In general, depth cameras are divided into two types: stereo cameras and time of flight (TOF) cameras. Stereo cameras capture the detail of depth based on the binocular stereo vision theorem [6, 7], which calculates the distance of points in the real world by the principle of parallax and applies two images for measuring the targets that are captured by two parallel cameras from different positions. The method calculates the positional deviation between the corresponding points of the image for obtaining the three-dimensional geometric information of an object. By combining the two images and observing the differences between them, it is possible to obtain a clear sense of depth, establish the correspondence between the features, and map the same spatial physical point in different images.

By contrast, TOF cameras, also refferred to flight time cameras, obtain the targets' distance by continuously transmitting light pulses to the targets and receiving the light returned from the object. Next, the camera measures the flight (round-trip) time of the light pulse [8–10]. This technique is essentially similar to the principle of a 3D laser sensor; however, while the 3D laser sensor on; y scans the target point-by-point, the TOF cameras also obtain the information of flight time from a 2D area.

In this paper, we propose a framework for recognizing human postures by simulating the human body skeleton and its movements according to the 3-dimensional points

cloud data using a series of unsupervised learning algorithms.

This paper is organized as follows. In Section 2, we review several state-of-the-art methods for human body recognition. In Section 3, we explain the details of the proposed method, and in Section 4, we discuss the experimental results. Finally, concluding remarks are made in the last section.

2. Related Works

Researchers have been working on the referent field for a long time [11–13]. Consequently, various methods have been proposed. In this section, we give a brief review for several typical methods which have been widely accepted and implemented. In [14], Aggarwal developed a taxonomy that divided all the proposed methods into two main groups: single-layer approaches and hierarchical approaches.

In [15], the authors proposed the supervised pose recognition method, which can be considered the most popular method today. The process described in the paper is mainly divided into two steps: first, the body part is marked from a single depth image for segmenting, and then the key nodes are marked. The body joint positioning is then performed, and the marked human body parts are remapped into the three-dimensional space to form a highly reliable spatial position for the key nodes.

The authors in [16] proposed a method for human full-body pose estimation from depth data that can be obtained using TOF cameras and the Kinect device. Their approach consists of robustly detecting anatomical landmarks in the 3D data and fitting a skeleton body model using constrained inverse kinematics. Instead of relying on appearance-based features for interest point detection, which can vary strongly with illumination and pose changes, they built a graph-based representation of the depth data to measure geodesic distances between body parts. As these distances do not change with body movement, it is able to localize anatomical landmarks independently of the pose.

Recently, in [17], the authors proposed a vector-shaped pose descriptor, which allows for the retrieval of similar poses and is easier to use with many machine learning libraries by constructing a feature space for appearances of human poses. This method has improved the limited scope of many methods based on a kinematic or surface mesh model, and performed efficiently in experiment.

Nevertheless, the methods mentioned above are based on supervised learning algorithms, which require a large amount of data for training before they are applied. Differently from the concept of these methods, in this paper, we propose a method for human posture recognition by a series of unsupervised algorithms, which does not require collection of training data.

3. Proposed Method

In this paper, we propose a human posture recognition method based on the concept of unsupervised learning. The framework of this method is shown in **Fig. 1**. The learning process can be generally divided into three steps: preprocessing, growing neural gas (GNG) based rough structure generation, and parameters optimization by particle swarm optimization (PSO). Detail of each step will be given in the following section.

3.1. GNG for Human Structure Construction

The particle points cloud data is computationally costly without peprocessing the data. The GNG is used for representing the points cloud data to a lower density structure, thus we utilized it for the construction of humans' rough structure.

The GNG is a typical self-organizing map (SOM) algorithms for unsupervised learning. It is known that unsupervised learning algorithms are a series of learning methods that work without any prior input data for training and give the desired output. Input data are consecutively represented by SOM in the form of input signals and the SOM changes its topological structure for representing the input data with the self-adaptation mechanism. Next, a growing mechanism is used for gradual adaptation and self-adjustment of size. The growing neural network starts in some minimal state (e.g., with some minimal number of neurons in the network), which is adapted to the input data. Then, it continually grows (increases its size) and adapts again. This cycle is repeated until the desired resolution of the neural network is achieved.

In the GNG learning algorithm, the following notations are used:

w_i: n dimensional vector of a node ($w_i \in Rn$)
G: set of nodes
N_i: set of nodes connected to the i-th node
c: set of edges
$a_{i,j}$: age of the edge between the i-th and the j-th node

The steps of the standard GNG algorithm are as follows:

Step 0. Initialize the network by creating two nodes at random positions, w_{c_1} and w_{c_2} in Rn. Then, set the connection between them.

Step 1. Randomly generate an input data v according to selecting function $p(v)$, which is the probability density function of data v.

Step 2. Select the nearest unit (winner) g_1 and the second-nearest unit g_2 by:

$$g_1 = \underset{i \in G}{\operatorname{argmin}} \parallel v - w_i \parallel \quad (1)$$

$$g_2 = \underset{i \in G/g_1}{\operatorname{argmin}} \parallel v - w_i \parallel \quad (2)$$

Step 3. Generate the connection a connection between g_1 and g_2, is such connection does not exist al-

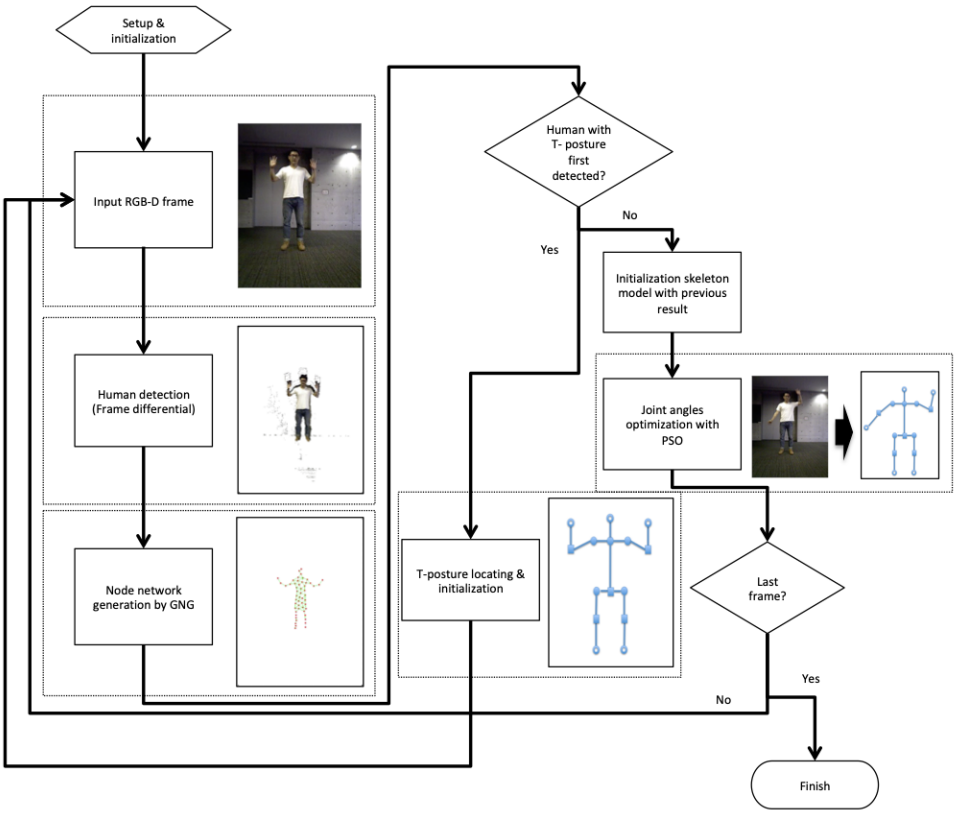

Fig. 1. Framework of proposed method.

ready. Set the age of the connection between g_1 and g_2 to zero:

$$a_{g_1,g_2} = 0 \quad \ldots\ldots\ldots\ldots\ldots \quad (3)$$

Step 4. Add the squared distance between the input data and the winner to a local error variable:

$$E_{g_1} \leftarrow E_{g_1} + \| v - g_2 \|^2 \quad \ldots\ldots\ldots \quad (4)$$

Step 5. Update the reference vectors of the winner and its direct topological neighbors by the learning rate η_1 and η_2 respectively, of the total distance to the input data:

$$w_{g_1} \leftarrow w_{g_1} + \eta_1 \cdot (v - w_{g_1}) \quad \ldots\ldots\ldots \quad (5)$$

$$w_j \leftarrow w_j + \eta_2 \cdot (v - w_j) \qquad \text{if} \quad c_{g_1,j} = 1 \quad (6)$$

Step 6. Increment the age of all edges emanating from s_1:

$$a_{g_1,j} \leftarrow a_{g_1,j} + 1 \qquad \text{if} \quad c_{g_1,j} = 1 \quad \ldots \quad (7)$$

Step 7. Remove edges with an age larger than a predefined threshold. If this results in units having no more emanating edges, remove those units as well.

Step 8. If the error E_q is higher than the predefined threshold, insert a new unit as follows:
Select the unit f with the maximum accumulated error among the neighbors of q.
Add a new unit r to the network and interpolate its reference vector from q and f:

$$W_r = 0.5 \cdot (w_q + w_f) \quad \ldots\ldots\ldots \quad (8)$$

Create a new edge that connects the new unit r with units q and f, and remove the existing edge between q and f.
Decrease the error variables of q and f by a fraction α:

$$E_q \leftarrow E_q - \alpha E_q \quad \ldots\ldots\ldots\ldots \quad (9)$$

$$E_f \leftarrow E_f - \alpha E_f \quad \ldots\ldots\ldots\ldots \quad (10)$$

Interpolate the error variable of r from q and f:

$$E_r = 0.1 \cdot (E_q + E_f) \quad \ldots\ldots\ldots\ldots \quad (11)$$

Step 9. Decrease the error variables of all units:

$$E_i \leftarrow E_i - \beta E_i \quad (\forall i \in G) \quad \ldots\ldots \quad (12)$$

Step 10. Continue with Step 1 if a stopping criterion (e.g., net size or some performance measure) is not yet fulfilled.

The number of point could be reduced largely from the original point cloud. Therefore the computation time would also be cut down.

Nevertheless, the standard GNG does not perform in dynamic environments. Considering this, an improved

GNG, named GNG with utility (GNG-U) was developed by Fritzke [18]. It is only slightly the standard GNG in that it updates not only the local errors but also the utility U_{g_1} with

$$U_{g_1} \leftarrow U_{g_1} + E_{g_2} - E_{g_1} \quad \ldots\ldots\ldots\ldots (13)$$

Then, it removes the node g_i if the following inequality is satisfied:

$$\frac{E_{g_i}}{U_{g_i}} > \gamma \quad \ldots\ldots\ldots\ldots\ldots\ldots (14)$$

where γ is a parameter that controls the number of nodes.

GNG-U can perform dynamic distributions. Based on this, Toda [19] proposed the modified GNG-U (GNG-U2) by introducing the weight vector and has achieved the superior results in 3D structures.

3.2. Human Skeleton Modeling

In this paper, we utilized a simplified kinematic model for representing the human skeleton. The i-th skeleton is constructed as $m \in M$ joints ($M = 15$ in this paper):

$$\begin{aligned} J_i = \{j_i^k\} = \{&center\_shoulder, left\_shoulder, \\ &right\_shoulder, left\_elbow, left\_hand, \\ &right\_elbow, right\_nand, center\_torso, \\ &center\_hip, left\_hip, left\_knee, left\_foot, \\ &right\_hip, right\_knee, right\_foot\} \end{aligned} \quad (15)$$

The positions of joints are shown in **Fig. 2**. Each joint in $j_i^k \in J_i$ is represented by 3D coordinates $(x_{j_i^k}, y_{j_i^k}, x_{j_i^k})$.

Subsequently the length of the links between each joints is described as:

$$L_i = \{l_1, l_2, \ldots, l_n\} \quad \ldots\ldots\ldots\ldots (16)$$

and the spatial coordinates of the skeleton are represented as

$$P_i = \{x_i, y_i, z_i\} \quad \ldots\ldots\ldots\ldots (17)$$

Despite being more convenient than the previous model in terms of computation, it is still unstable for generating the same skeleton for the same person in different frames. This is because there are two parameters in a skeleton model: the angles and length of each pair of joints. We know that the angles of the joints change dynamically-when changing posture, but the length of each pair of joints is always a constant for one person.

Thus, we try to make the preliminary experiment of evaluating the length and angels separately. In this model, there are 16 joints for controlling human posture. We do not calculate the positions of these joints directly. Instead, we generate the lengths and angles for each part as shown in **Fig. 2** (initial step), and we then calculate the position of each joints by forward kinematics.

The first part of initialization will solve the parameters of lengths. Given a special posture (e.g., T posture, as shown in **Fig. 2**), we randomly generate different values, and then choose the best one (because the angles are fixed owing to the fixed posture). In the second part, it randomly generates angles and selects the fittest one according to the fitness function.

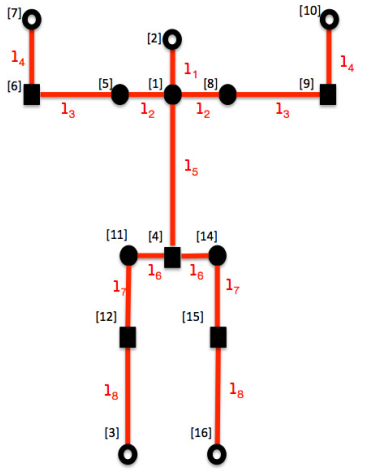

Fig. 2. Indices of human body's joints and links.

To compute the fitness of the skeleton model, we introduced the joints of ribs for auxiliary calculation, even though these rib joints have no meanings and can be hidden for human representation.

Because edged nodes contain much more information than the nodes in the center, as they have more probabilities of representing the limbs of the human, it is necessary to give them a higher weight than other nodes.

To weight them, first we search for their geographic center.

Given a weighted skeleton model, our target is to find the optimized parameters of rotation angles to obtain the fitness skeleton. The proposed method is divided into two parts, i.e., the initial step and the prediction step.

3.3. Denavit-Hartenberg Parameters for Human Skeleton Modeling

Even though the number of nodes is limited, the construction of the skeleton can to achieved. It is difficult to directly model the skeleton. Therefore, we applied a more convenient way for the representation of skeleton parameters.

The Denavit-Hartenberg (DH) parameters representation is the most widely used in kinematics today. It is named after Jacques Denavit and Richard Hartenberg who introduced this representation in 1955 [20, 21]. It calculates the coordinate transformation frame by frame making a list of parameters, with four parameters for each transformation:

1. Rotation angle α about X axis
2. Translation a along X axis
3. Translation d along Z axis
4. Rotation angle θ along Z axis.

Because all of the coordinate systems satisfy the constraint, all of the transformation can be represented by a

Table 1. DH parameters of the human body's joints.

Position	Joint ID	Index	α	θ	a	d
Left arm	3	1	0	0	l_2	0
	4	2	θ_1	0	0	0
		3	$\theta_2+90°$	0	0	0
		4	θ_3	0	l_3	0
	5	5	θ_4	0	l_3	0
Right arm	6	6	0	0	l_2	0
	7	7	$\theta_5+90°$	0	0	0
		8	0	$\theta_6+90°$	0	0
		9	θ_7	0	l_3	0
	8	10	θ_8	0	l_4	0
Head	2	11	$\theta_9-90°$	0	0	0
		12	θ_{10}	$-90°$	l_1	0
Torso	9	13	$\theta_{11}-90°$	0	0	0
		14	θ_{12}	$-90°$	l_5	0
Left leg	10	15	$-90°$	0	l_6	0
	11	16	$\theta_{13}-90°$	0	0	0
		17	0	$\theta_{14}+90°$	0	0
		18	θ_{15}	0	l_7	0
	12	19	θ_{16}	0	l_8	0
Right leg	13	20	$90°$	0	l_6	0
	14	21	$\theta_{17}-90°$	0	0	0
		22	0	$\theta_{18}+90°$	0	0
		23	θ_{19}	0	l_7	0
	15	24	θ_{20}	0	l_8	0

set of quadruple of parameters as

$$T_i = T_{\theta_j} T_{d_j} T_{a_j} T_{\alpha_j} \quad \ldots\ldots\ldots\ldots\ldots\ldots (18)$$

where $T_{\theta_j}, T_{d_j}, T_{a_j}, T_{\alpha_j}, (j \in m)$ represent the rotation matrix of the four steps listed above. The combined matrix T_i is

$$\begin{bmatrix} \cos\theta & -\cos\alpha\sin\theta & \sin\alpha\sin\theta & a\cos\theta \\ \sin\theta & \cos\alpha\cos\theta & -\sin\alpha\cos\theta & a\sin\theta \\ 0 & \sin\alpha & \cos\alpha & d \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (19)$$

We distribute all of the joints according to the parameters in **Table 1**. In this table, joint 1–4, 5–8, 9–12, 13–16 represent the DH parameters for the rotation angles of the left arm, right arm, left leg, and right leg, respectively. The indices of the joints and links are shown in **Fig. 3**.

By applying the DH parameters shown in **Table 1**, it is easy to build up the simulated human skeleton. Based on the DH parameters, the target for determining a human's posture is optimized by rotational angles of each joints.

3.4. PSO for Human Posture Recognition

We apply PSO to optimize all of the rotational angles to properly simulate the human posture. The PSO is a computational method that optimizes a problem by iteratively trying to improve a candidate solution according to given measures of quality. It was originally proposed by Kennedy and Eberhard [22], and has been extended to a series of advanced solutions.

The motivation from the PSO algorithm is inspired by the flocking behavior of birds nature. It contains a set

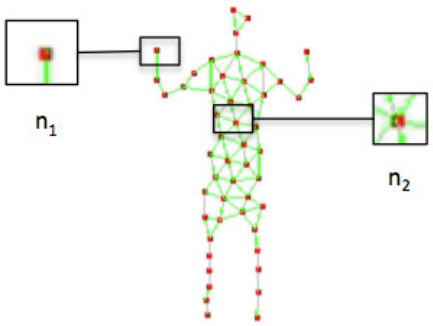

Fig. 3. Illustration of GNG node weights.

of particles, where each particle represents a bird in the flock. In this paper, we assume that each particle represents a skeleton candidate with a different series of rotation angles. The purpose is to iterate all of the particles with their velocities of rotational angle, and selecting the best series with the minimum value, which correspond to that the best skeleton for representing the human posture. The speed of the rotational angle in the PSO algorithm can be represented as

$$v_i^k = v_i^{k-1} + w_1 r_1 (Pbest_i - X_i) + w_2 r_2 (Gbest_i - X_i) \quad \ldots\ldots\ldots\ldots\ldots\ldots (20)$$

where v_i^k is a vector representing the angle velocities of the i-th agent at k-th iteration. v_i^k is controlled by three factors: the global best velocities $Gbest_i$, the personal best velocities $Pbest_i$ and the previous velocities v_i^{k-1}. Thus, the current angles for x_i^k will be calculated as

$$\theta_i^k = \theta_i^{k-1} + v_i^k \quad \ldots\ldots\ldots\ldots\ldots\ldots (21)$$

Different human postures are generated by the angles calculated from the above equation, depending on the previous knowledge and random factors.

Because the GNG nodes have roughly described structures of the human skeleton, it is necessary to optimize all of the parameters of the skeleton model to optimize the rotational angles. In order to calculate the value of the global best and the personal best, it is necessary to propose a proper evaluation function. Here we optimize the best skeleton by searching for the minimum value of the following evaluating function:

$$F(x) = \sum_{i=1}^{M} w_{g_i} d(g_i) \quad \ldots\ldots\ldots\ldots\ldots\ldots (22)$$

where M is the number of GNG nodes that were obtained from the previous step, w_{g_m} is the weight of the m-th GNG node g_m, and the function $d(g_m)$ is represented as

$$d(g_i) = \begin{cases} e^{-\frac{\|g_i - j_n\|^2}{2a^2}} & \text{if } \underset{j_n \in J}{\operatorname{argmin}}\, e^{-\frac{\|g_i - j_n\|^2}{2a^2}} \leq K \\ \tau & \text{otherwise} \end{cases}$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (23)$$

where a is a constant, and τ is a constant threshold with a large value. This evaluation function means that the skeleton with the smallest sum of distances for the total GNG nodes is more likely to be the fittest posture. Considering that each link is represented by a cylinder with a radial threshold of K, the value of $d(i)$ is valuable if and only if the node i is inside this cylindrical space. Nodes that are closer to the axis of the cylinder would have a higher possibility to be part of it.

Despite the advantages of GNG , it is obvious that GNG nodes are generated randomly, which means that each of the node has the same importance for constructing the whole GNG network. However, they should be separated into different levels based on their representations of human body. The reason is that different locations in human body have different importance when representing the human body. For example, the nodes that surround the elbow provide more information than the nodes located in the torso, therefore, these nodes should be given a higher weight when reconstructing the human body.

In this paper, we tried a simple but efficient weighted methods. It is obvious that nodes that are near to the center would have more edges than those located in the edge. Therefore we suppose that the weight of the i-th node is represented as w_i, then the weight is

$$w_{g_i} = \sum_{g_j \in G} D(g_i, g_j) \quad \ldots\ldots\ldots\ldots\ldots\ldots (24)$$

where $D(s_i, s_j)$ denotes

$$D(g_i, g_j) = \begin{cases} 0 & \text{if } c_{g_i,g_j} = 0 \\ 1 & \text{otherwise} \end{cases} \quad \ldots\ldots\ldots (25)$$

This also means that the weight of g_i depends on the number of edges linked to it, as shown in **Fig. 3**. It is obvious from the figure that the n_1 that contains less edges than n_2 provides more significant information for generating the human skeleton.

4. Experiment Setup and Results

For the experiment in this study, we applied ASUS Xtion PRO live depth camera as the frame capturing sensor, shown in **Fig. 4** and **Table 2**. Compared with the other widely used depth cameras, such as Microsoft Kinect, Xtion, it has a low-budget projects or systems.

In the first step, we applied the frame differential algorithm to extract the foreground points, which is regarded as the construction part of human body. Considering that it contains a large amount of noise, we also applied the median filter to reduce them. The result can be seen in **Fig. 5**.

Figure 6 shows the experiment result comparison between the standard GNG and the GNG-U2. It is obvious that the GNG-U2 provides a more stable network from a series of video frames compared with the standard GNG, the number of points have been reduced dramatically but the rough structure of the network remained the same. Next, **Fig. 7** shows that the number of points for processing is reduced dramatically after GNG learning. The left figure shows the point number of the original frame, whereas the right one shows the number of point for the GNG network. It is obvious that the number of points has been dramatically reduced after the GNG learning.

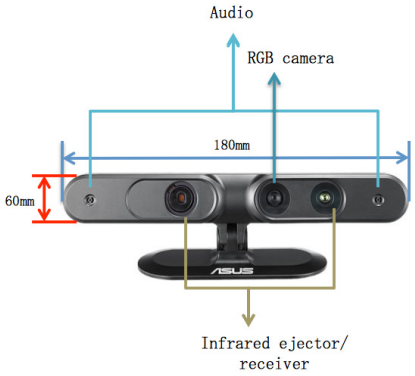

Fig. 4. Profile of the Xtion sensor.

Table 2. Features of the Xtion sensor.

Weight	490g
Interface	USB2.0
Available view angle	70 degree
fps	30 to 60
Resolution	640*480

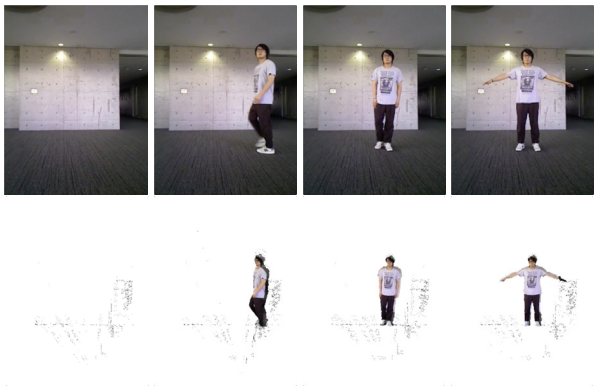

Fig. 5. Performance of foreground extraction.

As mentioned earlier, our solution contains a large number of parameters that need to be optimized, and it is difficult to localize all of them in one step. To overcome this issue, we seek to optimize part of the parameters in the initial step, and make the prediction in the following steps.

To locate the human body as fast as possible, we made the restriction at the initial part, i.e., we defined a special initial posture. In this experiment, we suppose that the human skeleton as a special T-posture because it is easier to determine the length of each link of such posture. In this case, all of the parameters of the rotational angles

Fig. 6. Experiment result between the standard GNG and the GNG-U2 in a series of input frames. The first row represents the original frame, the middle row shows the standard GNG, and the last row shows the GNG-U2.

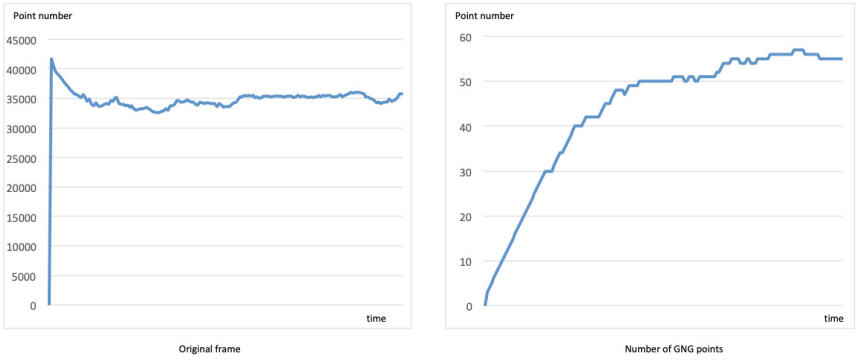

Fig. 7. Comparison of point number for processing for the same given input frames.

are fixed at the initial part, and the optimized parameters are only the spatial coordinates of the skeleton. Once the original coordinates are located, the following predictions will be simpler.

The result of the human posture is shows in **Fig. 8**, where the top row shows the original structure and the bottom row shows the result of human posture recognition.

The computational time in PSO is affected by the number of particles and iterations. We evaluate the run time cost based on three different conditions with different iteration times. **Fig. 9** illustrates the time taken by different particles and iterations. It is obvious that the time rises dramatically if the iterations increase. It is important to choose a proper iteration under different environment.

5. Discussion and Conclusion

In this paper, we proposed an unsupervised human posture recognition method that is different from most of the previous proposed methods. The proposed method contains a series of unsupervised learning algorithms, such as GNG, and PSO. Thus, no pre-training data is required, which is crucial for real world applications. By applying GNG, it deduces the run time cost dramatically compared with tackling the whole point cloud directly. In addition, the PSO made it possible to find the best simulated posture without any training.

Overall, this paper provides a preliminary method for human posture recognition. However, the run time cost for optimization would be a factor that limits the method for real-time implementation. In the next stage, we will focus on reducing the run time cost of optimization step.

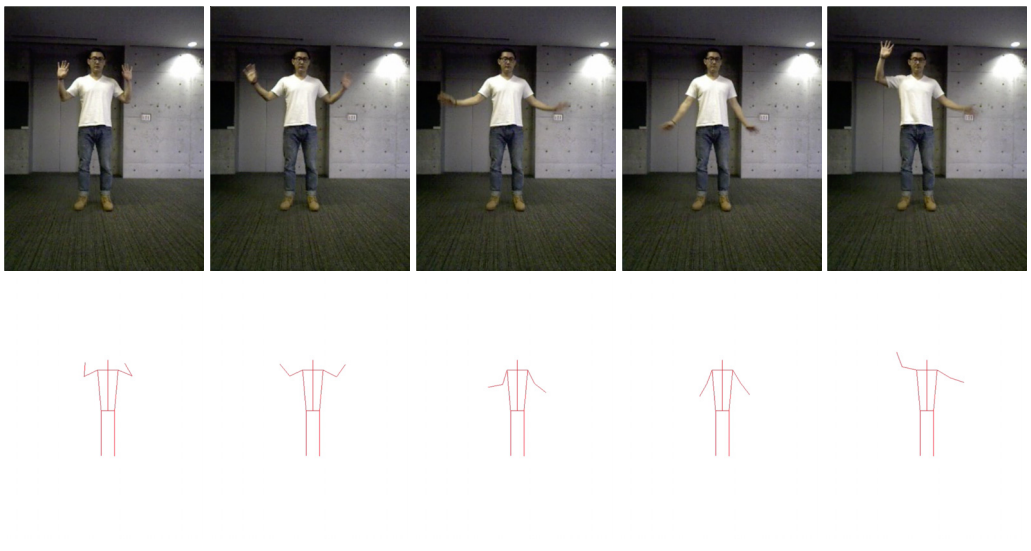

Fig. 8. Experiment results of human posture recognition generated by the proposed method.

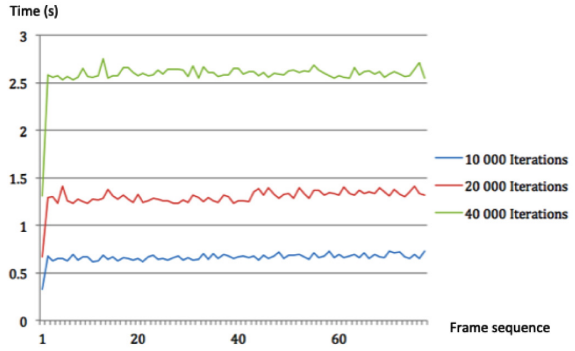

Fig. 9. Experiment result for different iteration times.

References:

[1] Y. Tang, H. Vu, P. Le, D. Masano, O. Thet, C. Fatichah, Z. Liu, M. Yamaguchi, M. Tangel, F. Dong, Y. Yamazaki, and K. Hirota, "Multimodal Gesture Recognition for Mascot Robot System Based on Choquet Integral Using Camera and 3D Accelerometers Fusion," J. Adv. Comput. Intell. Intell. Inform., Vol.15, No.5, pp. 563-572, 2011.

[2] J. J. Cabibihan, W.-C. So, and S. Pramanik, "Human-recognizable robotic gestures," IEEE Trans. on Autonomous Mental Development, Vol.4, Issue 4, pp. 305-314, 2012.

[3] T. Fong, I. Nourbakhsh, and K. Dautenhahn, "A survey of socially interactive robots," Robotics and Autonomous Systems, Vol.42, No.3-4, pp. 143-166, 2003.

[4] Y. Takahashi, K. Yoshida, F. Hibino, and Y. Maeda, "Human Pointing Navigation Interface for Mobile Robot with Spherical Vision System," J. Adv. Comput. Intell. Intell. Inform., Vol.15, No.7, pp. 869-877, 2011.

[5] E. E. Stone and M. Skubic, "Fall detection in homes of older adults using the microsoft kinect," IEEE J. of Biomedical and Health Informatics, Vol.19, Issue 1, pp. 290-301, 2014.

[6] W. Song, Y. Mae, and M. Minami, "Evolutionary Pose Measurement by Stereo Model Matching," J. Adv. Comput. Intell. Intell. Inform., Vol.9, No.2, pp. 150-158, 2005.

[7] M. Ye, Q. Zhang, L. Wang, J. Zhu, R. Yang, and J. Gall, "A survey on Human Motion Analysis from Depth Data," M. Grzegorzek, C. Theobalt, R. Koch, and A. Kolb (Eds.), Time-of-Flight and Depth Imaging: Sensors, Algorithm and Applications, pp. 149-187, Springer, 2013.

[8] C. Lee, H. Song, B. Choi, and Y.-S. Ho, "3D scene capturing using stereoscopic cameras and a time-of-flight camera," IEEE Trans. on Consumer Electron., Vol.57, No.3, pp. 1370-1376, 2011.

[9] J. Jung, J. Y. Lee, Y. Jeong, and I. S. Kweon, "Time-of-flight sensor calibration for a color and depth camera pair," IEEE Trans. Pattern. Anal. Mach. Intell., Vol.37, No.7, pp. 1501-1513, 2015.

[10] S. Foix, G. Alenya, and C. Torras, "Lock-in time-of-flight (ToF) cameras: a survey," IEEE Sens J., Vol.11, No.9, pp. 1917-1926, 2011.

[11] A. Kleinsmith and N. Bianchi-Berthouze, "Affective body expression perception and recognition: A survey," IEEE Trans. Affect. Comput., Vol.4, No.1, pp. 15-33, 2013.

[12] M. Hayase and S. Shimada, "Posture Estimation of Human Body Based on Connection Relations of 3D Ellipsoidal Models," J. Adv. Comput. Intell. Intell. Inform., Vol.14, No.6, pp. 638-644, 2010.

[13] J. Suarez and R. R. Murphy, "Hand gesture recognition with depth images: A review," Proc. 2012 IEEE RO-MAN: The 21st IEEE Int. Symp. on Robot and Human Interactive Communication, pp. 411-417, 2012.

[14] J. K. Aggarwal and M. S. Ryoo, "Human activity analysis: A review," ACM Computing Surveys, Vol.43, No.3, pp. 16:1-16:43, 2011.

[15] J. Shotton et al., "Real-Time Human Pose Recognition in Parts from a Single Depth Image," Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR 2011), pp. 1297-1304, 2011.

[16] L. A. Schwarz, A. Mkhitaryan, D. Mateus, and N. Navab, "Human skeleton tracking from depth data using geodesic distances and optical flow," Image Vision Comput., Vol.30, pp. 217-226, 2012.

[17] M. Stommel, M. Beetz, and W. Xu, "Model-free detection encoding retrieval and visualization of human poses from kinect data," IEEE/ASME Trans. Mechatronics, Vol.20, No.2, pp. 865-875, 2015.

[18] B. Fritzke, "Self-organizing network that can follow non-stationary distributions," Proc. of the Int. Conf. on Artificial Neural Networks '97, pp. 613-618, Springer, 1997.

[19] Y. Toda, H. Yu, Z. Ju, N. Takesue, K. Wada, and N. Kubota, "Real-time 3D point cloud segmentation using growing neural gas with utility," The 9th Int. Conf. on Human System Interaction, pp. 418-422, 2016.

[20] J. Denavit and R. S. Hartenberg, "A kinematic notation for lower-pair mechanisms based on matrices," Trans. ASME J. Appl. Mech, Vol.22, pp. 215-221, 1955.

[21] R. S. Hartenberg and J. Denavit, "Kinematic synthesis of linkages," McGraw-Hill Series in Mechanical Engineering, p. 435, McGraw-Hill, 1965.

[22] J. Kennedy and R. Eberhart, "Particle Swarm Optimization," Proc. of IEEE Int. Conf. on Neural Networks, pp. 1942-1948, 1995.

Name:
Wei Quan

Affiliation:
Graduate School of Systems Design, Tokyo Metropolitan University

Address:
6-6 Asahigaoka, Hino, Tokyo 191-0065, Japan

Brief Biographical History:
2013 Received Master degree from Graduate School of Information, Production and Systems, Waseda University

Main Works:
• W. Quan and N. Kubota, "Evolutionary people tracking for robot partner of information service in public areas," Int. Conf. on Intelligent Robotics and Applications, 2017.

Name:
Jinseok Woo

Affiliation:
Research Assistant Professor, Graduate School of Systems Design, Tokyo Metropolitan University

Address:
6-6 Asahigaoka, Hino, Tokyo 191-0065, Japan

Brief Biographical History:
2011 Received Master of Engineering degree from Graduate School of Systems Design, Tokyo Metropolitan University
2017 Received Ph.D. degree from Graduate School of Systems Design, Tokyo Metropolitan University
2017- Research Assistant Professor at Intelligent Mechanical Systems, Graduate School of Systems Design, Tokyo Metropolitan University

Main Works:
• J. Woo, J. Botzheim, and N. Kubota, "A Socially Interactive in Japanese Robot Partner Using Content-based Conversation System for Information Support," J. Adv. Comput. Intell. Intell. Inform., Vol.22, No.6, pp. 989-997, 2018.
• J. Woo, J. Botzheim, and N. Kubota, "System Integration for Cognitive Model of a Robot Partner," Intelligent Automation & Soft Computing (AUTOSOFT), pp. 1-14, 2017.

Membership in Academic Societies:
• The Institute of Electrical and Electronic Engineers (IEEE)
• The Japanese Association of Research on Care and Welfare (JARCW)

Name:
Yuichiro Toda

Affiliation:
Assistant Professor, Graduate School of Natural Science and Technology, Okayama University

Address:
3-1-1 Tsushima-Naka, Kita, Okayama, Okayama 700-8530, Japan

Brief Biographical History:
2017- Visiting Assistant Professor, Department of Systems Design, Tokyo Metropolitan University
2018- Assistant Professor, Graduate School of Natural Science and Technology, Okayama University

Main Works:
• A. A. Saputra, Y. Toda, J. Botzheim, and N. Kubota, "Neuro-Activity-Based Dynamic Path Planner for 3-D Rough Terrain," IEEE Trans. on Cognitive and Developmental Systems, Vol.10, No.2, pp. 138-150, 2018.
• Y. Toda and N. Kubota, "Evolution Strategy Sampling Consensus for Homography Estimation," J. Adv. Comput. Intell. Intell. Inform., Vol.20, No.5, pp. 788-802, 2016.

Membership in Academic Societies:
• The Institute of Electrical and Electronic Engineers (IEEE)
• The Japan Society of Mechanical Engineers (JSME)

Name:
Naoyuki Kubota

Affiliation:
Professor, Graduate School of Systems Design, Tokyo Metropolitan University

Address:
6-6 Asahigaoka, Hino, Tokyo 191-0065, Japan

Brief Biographical History:
2005- Associate Professor, Department of Systems Design, Tokyo Metropolitan University
2007 Visiting Professor, University of Portsmouth
2009-2012 Visiting Professor, Seoul National University
2012- Professor, Graduate School of Systems Design, Tokyo Metropolitan University

Main Works:
• D. Tang, B. Yusuf, J. Botzheim, N. Kubota, and C. S. Chan, "A novel multimodal communication framework using robot partner for aging population," Expert Systems with Applications, Vol.42, Issue 9, pp. 4540-4555, 2015
• N. Kubota and Y. Toda, "Multi-modal Communication for Human-friendly Robot Partners in Informationally Structured Space," IEEE Trans. on Systems, Man, and Cybernetics – Part C, Vol.42, No.6, pp. 1142-1151, 2012.

Membership in Academic Societies:
• The Institute of Electrical and Electronic Engineers (IEEE)
• Japan Society for Fuzzy Theory and Intelligent Informatics (SOFT)
• The Society of Instrument and Control Engineers (SICE)