Active Sensing and Information Structuring for Intelligent Robot Vision with Range Imaging Sensor

Hironobu Sasaki[1], Tomoya Fukukawa[1], Chen Fei[1], Sun Baiqing[1], Toshio Fukuda[1]

1 Department of Micro-Nano Systems Engineering, Nagoya University

Furo-cho, Chikusa-ku, Nagoya, 464-8603, Japan.

Abstract: Recently, various types of robots have been researched and developed for supporting our life. Also, the perceptual system for the robot is researched. Visual perception includes a lot of valuable information and it is useful for all intelligent robot system. In this paper, we discuss intelligent robot vision in order to active sensing and information structuring for human-robot collaboration system. There are many visual sensors and we use the range-imaging camera to detect distance and image data. We propose a method for active sensing and target information structuring by using growing neural Network and Genetic algorithm. In the experimental results, we show the potency of our method.

1. INTRODUCTION

Visual information include a lot of important information for human and animals. Therefore vision system such as digital image processing have been researched. Also, various types of robots have been researched and developed such as human-friendly robots, pet robots, amusement robots, and partner robots. Such a robot requires many intelligent visual perceptions. Visual perception has been discussed from various points of view. The robot vision is based on the time-series of image processing, not the processing on a single image. Various technologies for image processing are required for realizing the robot vision, e.g., color processing, target detection, template matching, shape recognition, motion extraction, and optical flow. Furthermore, multiple objects tracking and people tracking should be done in the robot vision. We have applied spiking neural networks, cellular neural networks, self-organizing map, and others for human detection, motion extraction, and shape recognition.

In this paper, we focus on multiple object detection and multiple human detection. Kalman filter, particle filters, genetic algorithms, particle swarm optimization, and others have been applied in appearance-based methods. Furthermore, dynamic model of human movement is also applied to improve the accuracy of people tracking. These methods try to detect the features of human appearance, and to trace them over time, but there are problems on variability of appearance features and computational cost in the real-time people tracking.

As for the visual perception, we can divide by the sensor and the processing function. Because the variety of vision sensor is developed, the camera sensor becomes popular. Usually, the robot have some vision sensor for sensing ambient environmental information. Generally, in the vision information processing, static image processing is useful for robot vision, and we can process sequential image to processing videos nowadays. In the visual image, there are

so many useful information to extract. And we have to select and fusing the visual information by the task.

The final goal of this research is generate 3-D template to create object model database of unknown environment automatically. Therefore, we should develop the strategy such as 1;extract the ROI(Region Of Interest) from the environment, 2;extract the object and background information separately from the ROI, 3;generate the 3-diminutional template model, 4;3-diminutional template matching, 5;develop 3-diminutional object database automatically. Here we focus on generating the three dimensional template model. In this paper we propose clustering method for a lot of information by growing neural gas. Here we aim multiple object detection and human detection, get the shape information, or labeling to the detected object.

2. VISUAL SYSTEM FOR MOBILE ROBOTS

Recently, there are many robots such as moving around us. Such robots should collecting external world information. And the robot have to equip many kind of sensors. There are a lot of sensors to get environmental information, and it have been researching. In the field of intelligent robotics, we usually uses vision sensors. The vision sensor can get varieties of information from the environment, and it is useful for such kind of robot. Human and other animate beings using a big part of the brain for the visual processing. We can analogize that vision information is very important for our life.

Robot vision is composed of "Processing", "Classification", and "Perception". "Processing" include the image emphasis or image filtering. "Classification" include the character recognition or labeling etc... "Perception" include the image understanding or information extraction etc... In this paper we explain about "Processing" and "Classification".

2.1. Environmental Sensing

The static image processing is useful for robot vision, such as edge detection by laplacian filter, noise reduction by gaussian filter, and so on. And we can process sequential image to processing videos, such as by moving object detection by using difference picture or optical flow. However the vision sensors are easily influenced of environment light condition. And it is difficult to understanding physical relationship between objects. It is able to get 3 dimensional information by using multiple

vision sensors, but it is difficult to setting cameras with accurate positional relationship.

On the other hand, we can get the distance data by ultra sonic sensor or laser range finder, and so on. However, these sensors only can get a point information or linear information. So if we wont to get the 3 dimensional distance data, we have to move or rotate these sensors. In general, these kind of sensors are uses for mapping or localize ones position.
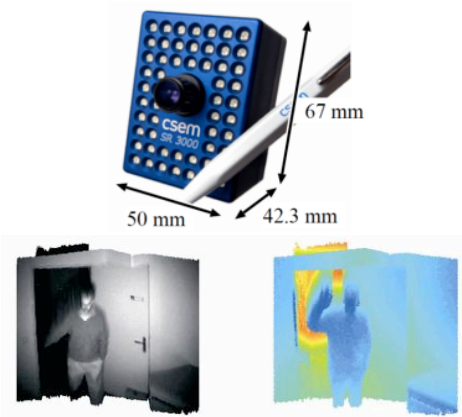
2.2. Range Imaging Camera

In the case of gathering image data and 3 dimensional surface data simultaneously, we generally uses cameras and range sensors. However it is difficult to synchronize or we have to compute differences of physical relationship between sensors. In this paper we use SR-3000 range imaging camera (Table.1, Fig.1). Range imaging camera can gather image data and 3 dimensional surface data simultaneously. The SR-3000 is an optical imaging system which offers real time 3D image data. It has infrared camera so the gathering data is almost free of the influence of environment light, but alternatively, it can't gather color image.

By using this sensor, we don't have to prepare the transform formula for sensor fusion of image and distance data. And we can clustering these data for detecting objects, reintegrate the 3 dimensional image by radiance value, or labeling to the objects.

3. OBJECT DETECTION AND CLASSIFICATION

3.1. Digital Image processing

Here, we explanation about time difference filter for radiance value input(Fig.1(a)). Fig.2(e) shows the difference of value input between time *"t-1"* and *"t"*. Here, we can see the moving objects. But if the target objects aren't moving, it is difficult to extract the target. Therefor, we divide the target by using difference between background and current input value. Here, the background data is made by formula bellow, as a long tarm memory.

$$b(x,y,t+1) = (1-\alpha) \cdot b(x,y,t) + \alpha \cdot v(x,y,t) \quad (1)$$

Where, $v(x,y,t)$ indicate carrent value of time t, $b(x,y,t)$ indicate background value of time t as long tarm memory, α indicate the learning rate. Here, if the α is large, background value update fast by input value. Fig.2(f) shows the background value of $\alpha = 0.1$, Fig.2(g) shows the difference between background and current input value. In general, the difference area increases by the learning rate decreases. However, if the object motion is first, object passed area also detected as target, so trade-off analysis is necessary. Next, the rectangular space in fig.2(g) shows the barycenter of extracted difference between background and current input value. This indicate that the moving objects (=targets) are included in this area, and the system should pay attention to this area. And, filtering the distance data by using the target area distance information(Fig.2h). Here, we can see the 4

Table. 1 Specifications of the SR3000 (MESA Imaging AG)

Pixel Array Size	176 x 144 (QCIF)
Field of View	47.5 x 39.6 degrees
Non-ambiguous range	7.5 meters
Distance Resolution	1% of range, typical
Frame Rate	25 fps, typical

(a) Value input   (a) Distance input
Fig. 1 SR3000 (MESA Imaging AG)

(a) Distance input   (b) 4 neighbor radicalization   (c) 20 neighbor laplacian

(d) Value input   (e) Difference

(f) Long-term memory   (g) Difference on (f) And center of gravity   (h) Filtered distance output

Fig. 2 Digital Image processing for multiple human detection

people in fig.2(d), and in fig.2(h), the shadow of four people is shown by the different digitized radiance value.

Here, the camera image include a copious information, so the digitization method has some availableness. However, averagely digitization is sometimes un-useful in case of many objects are observable. Because important information is not uniformly distributed in the space but biased. Therefore, we

can say that better information can be collected by focused digitalization.

In fig.3(a), we digitization the distance data with many objects in the room into 25 stages. Here, we cannot discover useful information from the digitization image. Therefor, pre image processing, such as granularity digitization is found to be useful(Fig.4). For example, to extract the object area by laplacian filter(Fig.3(b): 4 neighbor, Fig.3(c): 20 neighbor). Here we can see the objects area clearly.

3.2. Growing Neural Gas

Various types of pattern matching methods such as template matching, cellular neural network, recognition, and dynamic programming (DP) matching, have been applied for human detection problems. In general, pattern matching is composed of two steps of target detection and target recognition. The aim of target detection is to extract a target candidate from an image, and the aim of the target recognition is to identify the target from classification candidates. In this paper, we focus on the target detection, because the main aim of this paper is to discuss on growing neural gas.

Unsupervised learning is performed by using only data without any teaching signals [12-20]. Self-organized map (SOM), neural gas (NG), growing cell structures (GCS), and growing neural gas (GNG) are well known as unsupervised learning methods. Basically, these methods use the competitive learning. The number of nodes and the topological structure of the network in SOM are designed beforehand [12,13]. In NG, the number of nodes is fixed beforehand, but the topological structure is updated according to the distribution of sample data [14]. On the other hand, GCS and GNG can dynamically change the topological structure based on the adjacent relation (edge) referring to the ignition frequency of the adjacent node according to the error index. However, GNG does not delete nodes and edges, while GNG can delete nodes and edges based on the concept of ages [15,16]. Furthermore, GCS must consist of k-dimensional simplices whereby k is a positive integer chosen in advance. The initial configuration of each network is a k-dimensional simplex, e.g., a line is used for k=1, a triangle for k=2, and a tetrahedron for k=3 [17,18]. GCS has applied to construct 3D surface models by triangulation based on 2-dimensional simplex. However, because the GCS does not delete nodes and edges, the number of nodes and edges is over increasing. Furthermore, GCS cannot divide the sample data into several segments. Fig.5 shows how to cluster the data by GNG. GNG cluster the data by nodes and edges. GNG is topological clustering method, and it cluster the data with the shape (Fig5.(4)).

Table.2 shows preliminary simulation results of comparison among SOM, NG, GCS, GNG. In the preliminary simulation, the number of nodes in two-dimensional SOM is 100 (10x10), and the maximal number of NG node is 100. The parameters used in GCS simulations were: λ=200, η^{G}_1=0.04, η^{G}_2=0.001, α=1.0, β=0.0005, in GNG simulations were: λ=200, η^{G}_1=0.05, η^{G}_2=0.001, α=0.5, and β=0.0005. We used data distribution of three rings. We
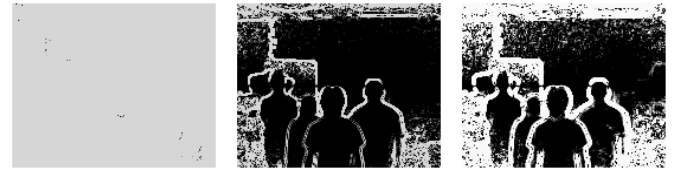

(a) (b) (c)
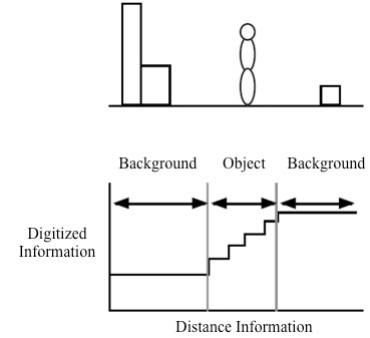
Fig. 3 Image digitization by distance and Laplacian filter

Fig. 4 Information extraction of intent area by the filter

(1)Node initialization   (2)Learn nodes and edges

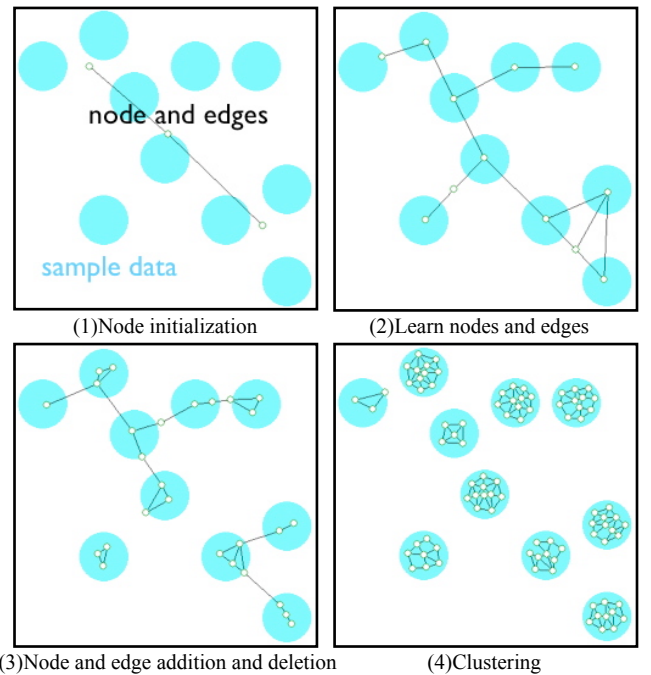(3)Node and edge addition and deletion   (4)Clustering

Fig. 5 Clustering by Growing Neural Gas

conducted comparison of these methods by using the sample data of three rings. Table.2 shows the comparison of evaluation values among SOM, NG, GCS, and GNG. The computational time of GCS is the shortest because the original GCS does not delete nodes and edges. NG needs much more computational time than others because the algorithm of sorting nodes is adopted every iteration. Furthermore, GCS successfully perform triangulation.

GNG can dynamically change the adjacent relation (edge) referring to the ignition frequency of the adjacent

node. We apply GNG for unsupervised clustering of the distribution of radiance value and the distance data. The learning algorithm of GNG is shown as follows. The nth dimensional reference vector of the ith node is w_i; a set of nodes is A; a set of nodes connected to node i is N_i; a set of edges is C; and the age of the edge between the ith and jth node is $age_{(i,j)}$.

Step 1. *Select two units c_1, c_2 at random position w_{c1}, w_{c2} in $\mathbf{R}_n$. Initialize the connection set.*

Step 2. *Determine the nearest unit s_1 and the second-nearest unit s_2 according to input signal ξ by*

$$s_1 = \arg \min_{c \in A} \|\xi - w_c\| \quad \text{and} \quad s_2 = \arg \min_{c \in A\setminus\{s_1\}} \|\xi - w_c\| \quad (2)$$

where ξ is composed of the position (x, y) and radiance value on the image (Fig.6(a)).

Step 3. *If a connection between s_1 and s_2 does not yet exist , create it. Set the age of the connection between s_1 and s_2 to zero.*

$$age_{(s1,s2)} = 0 \quad (3)$$

Step 4. *Add the squared distance between the input signal and the winner to a local error variable E_{s1}.(Fig.6 (b))*

$$E_{s1} \leftarrow E_{s1} + \|\xi - w_{s1}\|^2 \quad (4)$$

Step 5. *Adapt the reference vectors of the winner and its direct topological neighbors by the learning rate ε_b and ε_n, respectively.*

$$\nabla w_s = \varepsilon_b (\xi - w_s) \quad \nabla w_n = \varepsilon_n (\xi - w_n) \quad (5)$$

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

$$age_{s1} \leftarrow age_{s1} + 1 \quad (6)$$

Step 7. *Remove edges with the age larger than a_{max}. If units have no more emanating edges after this, remove those units.(Fig.6(c))*

Step 8. *If the number of input signals generated so far is an integer multiple of a parameter λ, insert a new unit as follows.(Fig.6(d))*

In addition, the node is generated based on the distance of the intent objects. Paying attention to the object is useful for the reduction of nodes that indicate the useless obstacles. And also, we can separate off the adjacent object by paying attention to the object area. We set the node additional space threshold by the position of intent objects. In future we will be using normal distribution or gaussian membership function for node additional probability.

In this way, the radiance value distribution can be extracted from the image by using GNG.

3.3. Multipoint observation

Here, we have one more problem to generate 3-D model. If we use a fixed camera, we couldn't generate the

Table. 2 Comparison of evaluation values

	Calculation Cost (ms)	Nodes	Edges	Deleted Edges
SOM	2100	100	180	0
NG	8600	100	193	280
GCS	1200	169	501	0
GNG	1900	168	369	392

(a) node fired

(b) node moved

(c) edge deletion

(d) node addition

Fig. 6 How to learn GNG nodes and edge

(a) Multiple sensors

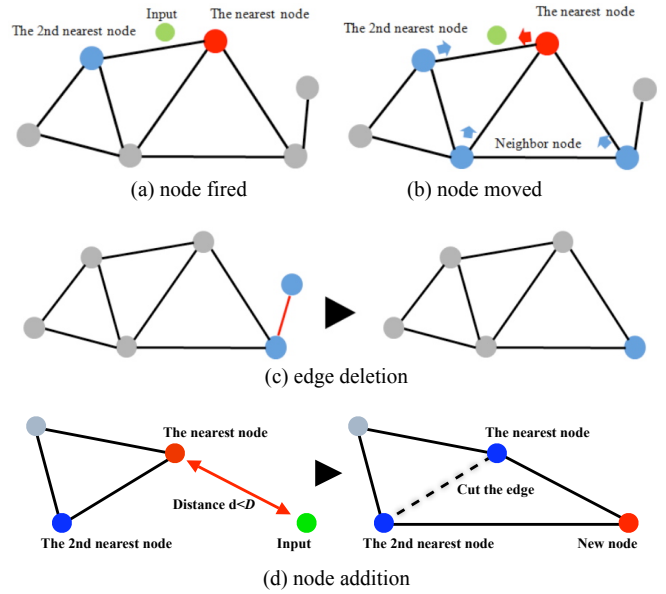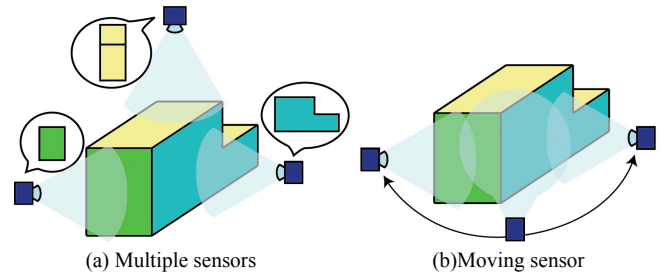(b)Moving sensor

Fig. 7 Multipoint observation of the sensor

model. The system should get the other observing point of data. For example "side view", "top view", "back view". One of the solution for this problem is to install multiple sensor in the environment. Because these sensors position were fixed, system is easy to integrate the information. But it takes much more equipment cost. Moreover, the system doesn't have the robustness. Another solution is to move around the sensors for get multipoint information. Here, we have to estimate the position of the sensor in each time. But the equipment cost will be lower than the described above. And also, the sensing range/area is larger than the above, so we can say this the system has much robustness. In this way, detailed 3-D information should be generated from multiple observation.
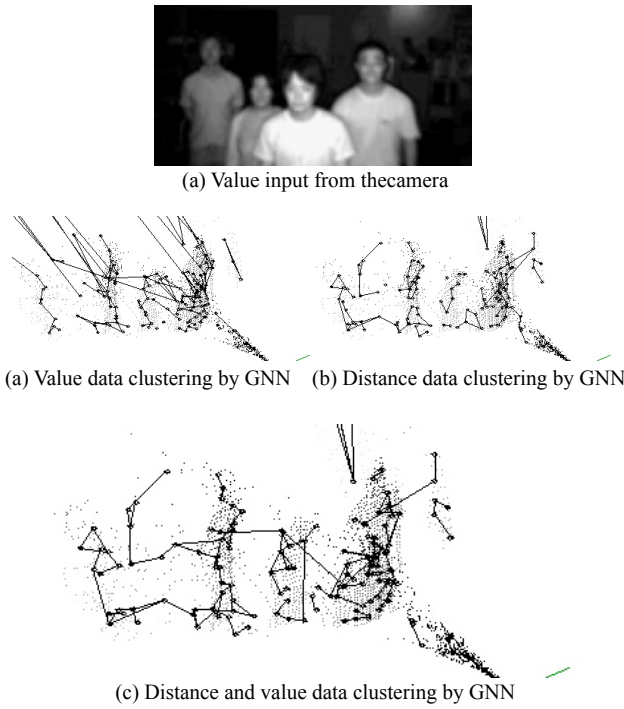
(a) Value input from thecamera

(a) Value data clustering by GNN   (b) Distance data clustering by GNN

(c) Distance and value data clustering by GNN

Fig. 7 Sensor fusion for multiple human detection (4 people)

Fig. 8 Movement of SR3000 with arm robot

Fig. 9 Consecutive data clustering

4. Experimental Results

In the experimental results, we show the results of classification for detecting multiple human by using GNG for the range imaging camera data.

4.1. Sensor fusion of distance data and brightness value

Fig.7 shows the experimental results of sensor fusion with GNN, and it shows sensor data include 4 people. Because peoples aren't standing straight as a line but overlapping each other, it is difficult to classificate only using brightness value(Fig.7(b)). Fig.7(c) shows classificate results of distance data. Human who standing on left side, especially inside, are expressed by a number of node set. But some of the nodes are mutually connected. This cosed by similarity of distance and brightness value. So we have to refine the target domain and have to radicalization or accentuation of distance data. And also we have to use sensor fusion by using various sensor information. Though in fig.7(d), we can see the human classification is almost done, but it cannot divide to be exact. So we have to improve these algorithm.

4.2. Multipoint observation

Detailed 3-D information should be generated from multiple observation. By moving sensor, we implemented the active sensing method for separation from background and the information structuring of the object. The range-imaging camera was installed in the point of the arm robot (Fig.8), and the information gathering was done while rotating surroundings of the object. The sensor works with the arm
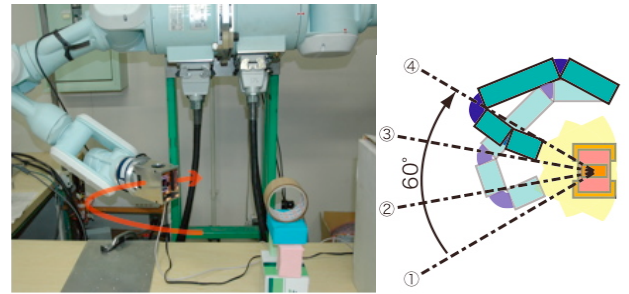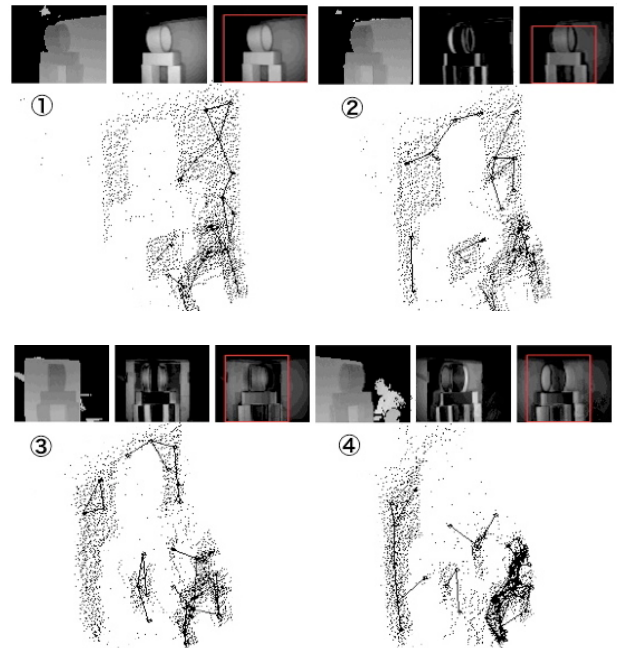
robot by tracks of circle that centers on the object to gather the consecutive view point information. In the experimental results, we show the results of classification for detecting object by using GNG for the range imaging camera data. The generation of the node and the edge was limited according to the output result of the digital image processing with concept of ROI. Fig VII shows the result of sensor information at 0 degrees, 20 degrees, 40 degrees, and 60 degrees, and clustering result of GNG. We can see that the object is separated from background as the calculation advances. However, we couldn't get the accurate shape of the object. Because of a problem that information on the edge is omitted during clustering, we couldn't get the accurate shape of the object.

5. Summary

As mentioned before, range imaging camera and GNG is useful for multiple human detection. Range imaging camera can easily get a lot of information. However, proposal algorithm is unsatisfactory. So in the future tasks, we have to

propose more effective algoritm such as difference filter, long-term memory filter, radicalization.

The research of the part of "Perception" in robot vision is still insufficient. The individual identification of the object, detection and identification of the movement, grouping of the objects are valuable for the intelligent robot vision. Here, we can assume that the labeling for the recognized object is useful for "Perception". And proposed method will be helpful for labeling the clusters, and utilize the information.

In this research, we aimed at the construction of the intelligent robot vision. We using GNG for self-organized information of the object. And we discussed about the usage of the structured information. We discussed about the effectiveness of structurizing method for three dimension information by GNG. Three dimension information on the object was structurized by range-imaging camera with consecutive information. In the experiment result, the boundary of the background and the object was able to be clarified.As future tasks, we improve GNG to preserve more detailed three dimension shape. Also, we should structurize the information on the object in detail. To improve the utilization efficiency of information, we make some groups of the information by applying upper and lower level of the GNN layer.

As mentioned before, range imaging camera and GNG is useful for active sensing. Range imaging camera can easily get a lot of information. However, proposal algorithm is unsatisfactory. So in the future tasks, we have to propose more effective algorithm such as difference filter, long-term memory filter, radicalization.

The research of the part of "Perception" in robot vision is still insufficient. The individual identification of the object, detection and identification of the movement, grouping of the objects are valuable for the intelligent robot vision. Here, we can assume that the labeling for the recognized object is useful for "Perception". And proposed method will be helpful for labeling the clusters, and utilize the information.

References

[1]   Y. Nakauchi, R. Simmons, A Social Robot that Stands in Line, Journal of Autonomous Robots, Vol. 12, No. 3, pp.313-324, 2002.

[2]   H.Ishiguro, M.Shiomi, T.Kanda, D.Eaton, and N. Hagita, Field Experiment in a Science Museum with communication robots and a ubiquitous sensor network, Proc. of Workshop on Network Robot System at ICRA2005, 2005.

[3]   Anderson, J. A. & Rosenfeld, E. (1988). Neurocomputing, The MIT Press, Cambridge, Massachusetts, US

[4]. David Marr (1982), Vision, W. H. Freeman, San Francisco.

[5]   K.Fukushima (2003). Neural network model restoring partly occluded patterns, Knowledge-Based Intelligent Information and Engineering Systems, (7th International Conference, KES 2003), Part II, eds: V. Palade, R. J. Howlett, L. Jain, Berlin - Heidelberg: Springer-Verlag, pp. 131-138

[6]   H. H. Bulthoff, S..W. Lee, T.A. Poggio, & C. Wallraven (2002), Biologically Motivated Computer Vision, Springer-Verlag.

[7]   J. H. Holland, "Adaptation in Natural and Artificial Systems", First MIT Press Ed., Massachusetts: The MIT Press, 1992.

[8]   D. Fogel, "Evolutionary Computation", New York: IEEE Press, 1995.

[9]   R. C. Eberhart, J. Kennedy, and Y. Shi, "Swarm Intelligence", San Francisco: Morgan Kaufmann Publ., 2001.

[10] J. Kennedy and R. Eberhart, "Particle Swarm Optimization" Proc. IEEE Int. Conf. Neural Netw., Perth, Australia, pp. 1942-1945, 1995.

[11] J. Huidong, L. K. Sak, and W. M. Leung, "Genetic-guided Model-based Clustering Algorithms" Proc. 2001 Int. Conf. on Art. Intelligence, vol. 2, pp. 653-659, 2001.

[12] T. Kohonen: Self-Organizing Maps; Springer, 2000.

[13] T. Kohonen: Self-Organization and Associative Memory; Springer-Verlag, 1984.

[14] T. M. Martinetz and K. J. Schulten: A "neural-gas" network learns topologies; Artificial Neural Networks, Vol. 1, pp. 397-402, 1991.

[15] B. Fritzke: A growing neural gas network learns topologies; Advances in Neural Information Processing Systems, Vol. 7, pp. 625-632, 1995.

[16] B. Fritzke: Growing self-organizing networks – why?; European Symposium on Artificial Neural Networks, pp. 61-72, 1996.

[17] B. Fritzke: Unsupervised clustering with growing cell structures; Neural Networks, Vol. 2, pp 531–536, 1991

[18] B. Fritzke: Growing cell structures – a self organizing network in k dimensions; Artificial Neural Networks, Vol. 2, No.2, pp. 1051-1056, 1994.

[19] K.A.J. Doherty, R.G. Adams, N. Davey: Hierarchical Growing Neural Gas; Adaptive and Natural Computing Algorithms, pp. 140-143, 2005.

[20] C. Fyfe: Two topographic maps for data visualization; Data Mining and Knowledge Discovery, Vol. 14, No. 2, pp. 207-224, 2007.

[21] N. Kubota and K. Nishida, "Cooperative Perceptual Systems for Partner Robots Based on Sensor Network", Int. J. Comp. Sci. and Netw. Security, Vol. 6, No. 11, pp. 19-28, 2006.

[22] N. Kubota, "Visual Perception and Reproduction for Imitative Learning of A Partner Robot", WSEAS Trans. Signal Processing, vol. 2, no.5, pp. 726-731, 2006.

[23] N. Kubota, "Computational Intelligence for Structured Learning of A Partner Robot Based on Imitation", Info. Sci., vol. 171, no. 4, pp. 403-429, 2005.

[24] I. A. Sulistijono and N. Kubota, "Particle Swarm Intelligence Robot Vision for Multiple Human Tracking of A Partner Robot", Proc. Society of Instrument and Control Eng. Annual Conf., pp. 604-608, 2007.