Growing Neural Gas for Intelligent Robot Vision with Range Imaging Camera

Hironobu Sasaki and Toshio Fukuda
Department of Micro-Nano Systems Engineering, Nagoya University
Furo-cho, Chikusa-ku, Nagoya Japan
sasaki@robo.mein.nagoya-u.ac.jp
fukuda@mein.nagoya-u.ac.jp

Masashi Satomi and Naoyuki Kubota
Department of System Design
Tokyo Metropolitan University
6-6 Asahigaoka, Hino, Tokyo Japan
satomi-masashi@sd.tmu.ac.jp
kubota@tmu.ac.jp

***Abstract* - In this paper, we discuss a intelligent robot vision in order to perceive people and the environment. We use the range imaging camera to detect distance and image data. We propose a method for perceive moving target by using growing neural gas and Genetic algorithm. In the experimental results, we show the potency of our method.**

Index Terms - Growing Neural Network, Growing neural gas, range imaging camera, Human tracking, Object tracking

I. INTRODUCTION

Recently, various types of robots have been researched and developed such as human-friendly robots such as pet robots, amusement robots, and partner robots[1-4]. Such a robot requires many intelligent visual perceptions. Visual perception has been discussed from various points of view [5-9]. 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 [35-41]. In this paper, we focus on human tracking. The tracking problem of human or objects is significantly harder than that of a single person or object. The human tracking problem includes two problems of people detection problem in each image and a tracking problem of detected people over time. In previous works, people tracking problems have been mainly solved by appearance-based methods. Kalman filter, particle filters, genetic algorithms, particle swarm optimization, and others [10-21] 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.

Usually, the mobile 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 from the color information. And we have to select information by the task. However, it is difficult to extract the 3 dimensional information from visual image, that is necessary for mobile robot, and we need to put in another sensor to the robot.

In this paper we using range imaging camera sensor to get image data and distance data. And we propose clustering method for 3 dimensional information by growing neural gas. Here we aim multiple human detection, get the shape information, or labeling to the detected object.

II. VISUAL SYSTEM FOR MOBILE ROBOTS

A. Environmental Sensing

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. Here, 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.

B. 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.

III. Object Detection And Classification

A. Digital Image processing

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. In fig.2(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.3). For example, to extract the object area by laplacian filter(Fig.2(b): 4 neighbor, Fig.2(c): 20 neighbor). Here we can see the objects area clearly.

Next, we explanation about time difference filter for radiance value input(Fig.1(a)). Fig.4(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.

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

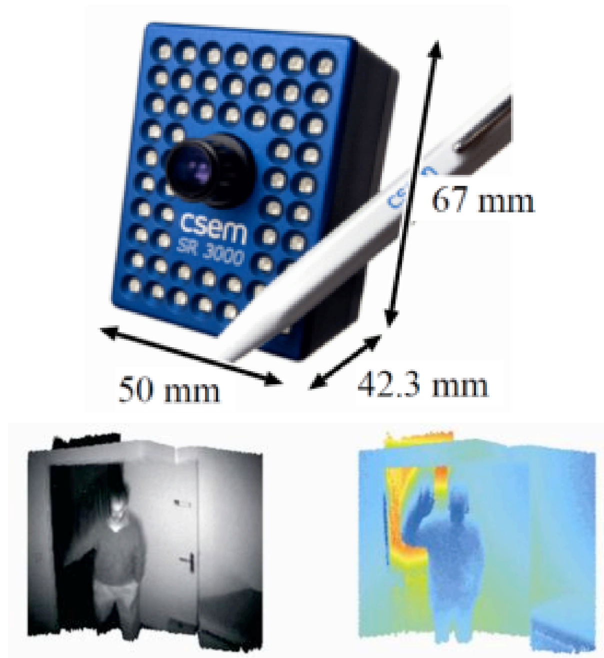
Where, $v(x,y,t)$ indicate current value of time *t*, $b(x,y,t)$ indicate background value of time *t* as long tarm memory, α indicate the learning rate. Here, if the α is large, background value update fast by input value. Fig.4(f) shows the background value of $\alpha=0.1$, Fig.4(g) shows the difference 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.4(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.4h). Here, we can see the 4 people in fig.4(d), and in fig.4(h), the shadow of four people is shown by the different digitized radiance value.

B. Growing Neural Gas

Various types of pattern matching methods such as template matching, cellular neural network, recognition, and dynamic programming (DP) matching [33,34], have been

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 (b) Distance input

Fig. 1 SR3000 (MESA Imaging AG)

(a) (b) (c)

Fig. 2 Image digitization by distance and Laplacian filter

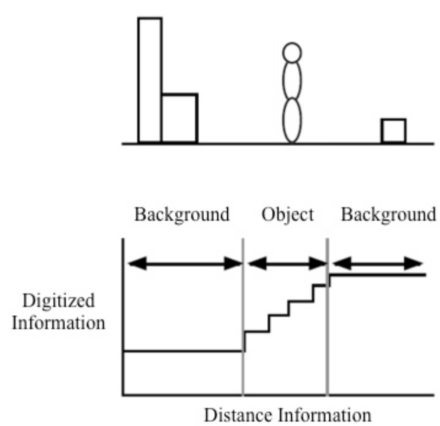


Fig. 3 Information extraction of intent area by the filter

Fig. 4 Digital Image processing for multiple human detection

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 [22-32]. 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 [23,24]. In NG, the number of nodes is fixed beforehand, but the topological structure is updated according to the distribution of sample data [25]. 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 [26,27]. 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 [28,29]. 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.

Here 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 conducted comparison of these methods by using the sample data of three rings. Fig. 5 shows the final stages of SOM, NG, GCS and GNG after 50000 iterations, respectively. Table I 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.

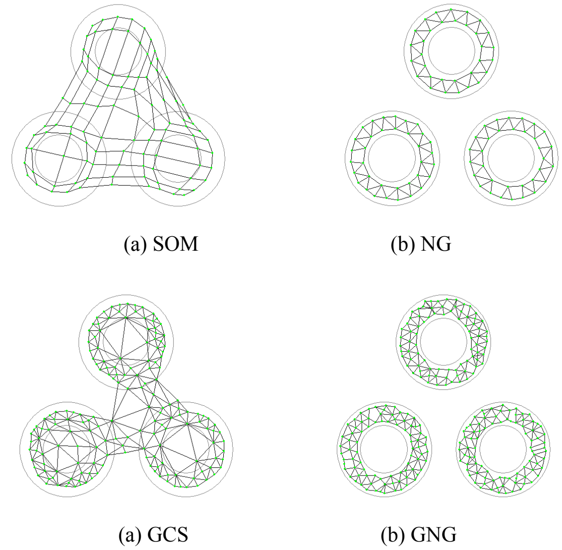


Fig. 5 The comparison among SOM, NG, GCS, GNG

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

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 $age_{(i,j)}$.

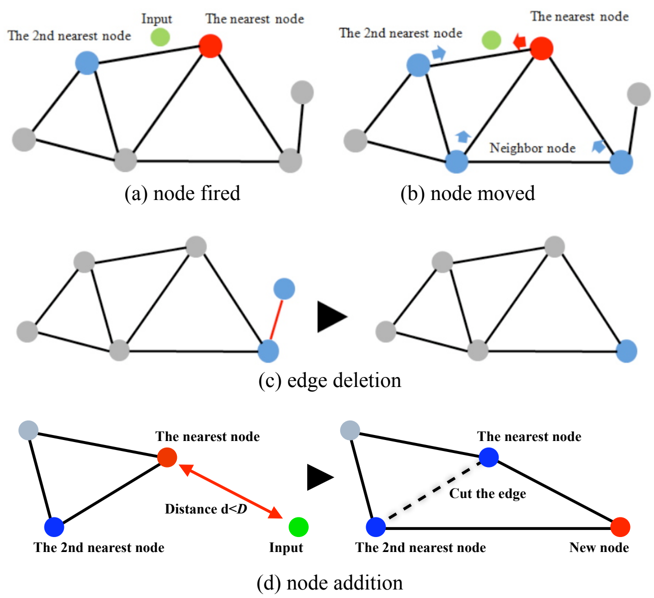


Fig. 5 How to learn GNG nodes and edge

Step 1. *Select two units c1, c2 at random position wc1, wc2 in* ***Rn****. Initialize the connection set.*

Step 2. *Determine the nearest unit s1 and the second-nearest unit s2 according to input signal ξ by*

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

where ξ is composed of the position (x, y) and radiance value on the image (Fig.5(a)).

Step 3. *If a connection between s1 and s2 does not yet exist , create it. Set the age of the connection between s1 and s2 to zero.*

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

Step 4. *Add the squared distance between the input signal and the winner to a local error variable Es1.(Fig. 5(b))*

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

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

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

Step 6. *Increment the age of all edges emanating from s1.*

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

Step 7. *Remove edges with the age larger than amax. If units have no more emanating edges after this, remove those units.(Fig.5(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.5(d))*

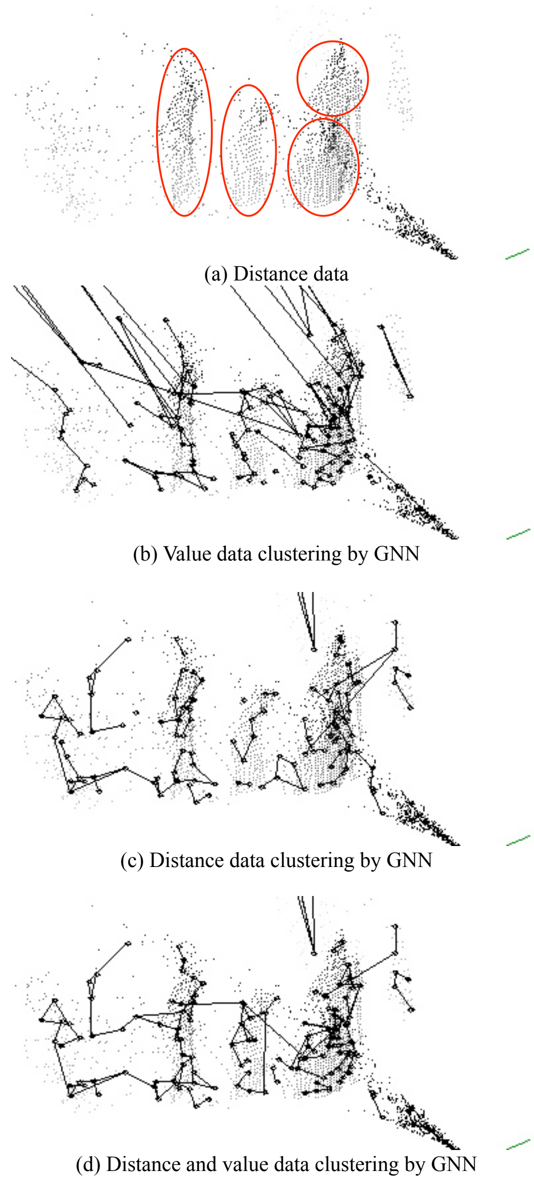


Fig. 6 Sensor fusion for multiple human detection (4 people)

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

IV. Experimental Results

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

A. Sensor fusion of distance data and brightness value

Fig.6 shows the experimental results of sensor fusion with GNN. Fig.6(a) shows distance 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.6(b)). Fig.6(c) shows classificate results of distance data. Human who standing on left side, especially

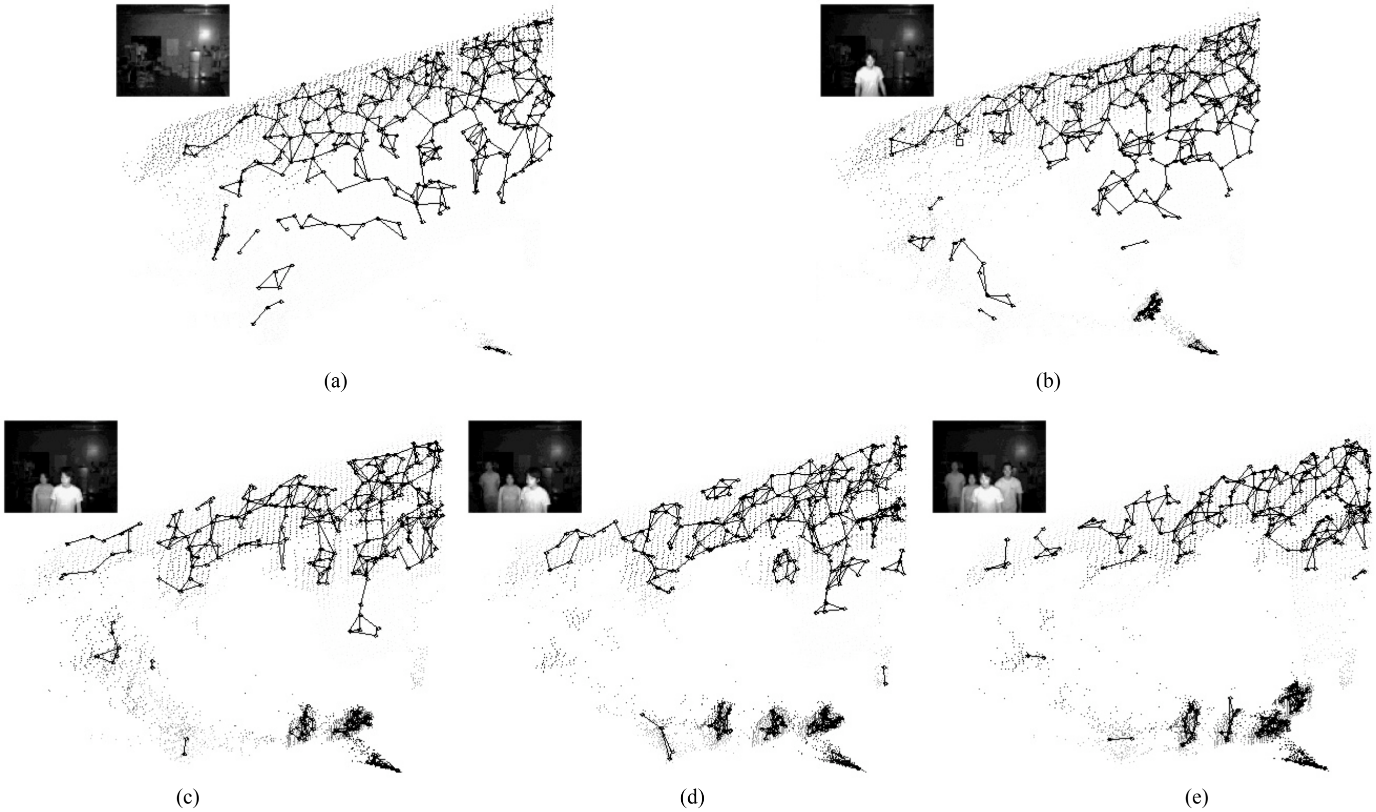


Fig. 7 Multiple Human detection by Growing Neural Network

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.6(d), we can see the human classification is almost done, but it cannot divide to be exact. So we have to improve these algorithm.

B. Multiple human detection by GNG

In fig.7, we can see the experimental results of multiple human classification. Here, we use laplacian filter to brightness value for manage GNG edges. And, we improve the GNG metod for classificate the objects. We set density of nodes is high for near objects and low for far objects. It becomes easy for the objects to classify the human from background. Fig.7(a)-(e) shows the output of GNG when the human was 0, 1, 2, 3, 4. ere GNG can clearly made some clusters as human. Here, GNG can divide multiple human from background, but sometimes the classification is incomplete. And also, there are many irrelative set of nodes with human. So for the future works, we also use brightness value or upper layer of GNG.

V. 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. Also, we want to labeling to the clusters, and utilize the information.

REFERENCES

[1] J.Cassell, Embodied Conversational Agents: Representation and Intelligence in User Interface, AI Magazine, Winter Vol.22, No.3, pp.67-83, 2001.

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

[3] M. Imai, T. Ono, H. Ishiguro: Physical Relation and Expression: Joint Attention for Human-Robot Interaction, IEEE Transactions on Industrial Electronics, Vol.50, No.4, pp.636-643, 2003.

[4] 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.

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

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

[7] C.-H. Chen, L.-F. Pau, and P.-S-P. Wang (1993), Handbook of Pattern Recognition & Computer Vision, World Scientific Publication Co. Pte. Ltd.

[8] 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

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

[10] D. Goldberg, Genetic Algorithms in Search, “Optimization and Machine Learning“, Massachusetts: Addison Wesley Publishing Company Inc., 1989.

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

[12] D. Fogel, “Evolutionary Computation“, New York: IEEE Press, 1995.

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

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

[15] G. G. Syswerda, “A Study of Reproduction in Generational and Steady-state Genetic Algorithms“, Genetic Algorithms, San Mateo: Morgan Kaufmann Publ. Inc., 1991

[16] S. Watanabe, T. Hiroyasu and M. Miki, “LCGA: Local Cultivation Genetic Algorithm for Multi-Objective Optimization Problems", Proc. Genetic and Evolutionary Comp. Conf., pp. 762-763, 2002.

[17] E. Bonabeau, M. Dorigo and G. Theraulaz, “Swarm Intelligence: From Natural to Artificial Systems“, New York: Oxford University Press, 1999.

[18] M. Dorigo and T. Stutzle, “Ant Colony Optimization“, Bradford Books, The MIT Press, 2004.

[19] J. Branke, "Evolutionary Optimization in Dynamic Environments", Vol. 3 of the Book Series on Genetic Algorithms and Evolutionary Computation, Kluwer Academic Publ. 2001.

[20] 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.

[21] M. Lozano, F. Herrera, N. Krasnogor, and D. Molina. "Real-Coded Memetic Algorithms with Crossover Hill-Climbing", Evolutionary Comp., The MIT Press, Cambridge, Massachusetts, USA, vol. 12, no. 3, pp. 273-302, 2004.

[22] A. Khorsi, An Overview of Content-Based Spam Filtering Techniques: Informatica, Vol 31 No 3, pp. 269-277, 2007.

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

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

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

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

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

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

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

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

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

[32] E. Berglund and J. Sitte, The Parameter-Less SOM, Proceedings of the Eight Australian and New Zealand Intelligent Information Systems Conference (ANZIIS 2003), Brian C. Lowell, Duncan A. Campbell, Clinton B. Fookes and Anthony Maeder (Eds), Sydney, Australia , December 2003, pp. 159 – 164, 2003.

[33] T.Fukuda, N.Kubota, An Intelligent Robotic System Based on a Fuzzy Approach, Proceedings of IEEE, 87(9): 1448-1470, 1999.

[34] N.Kubota, K.Tomoda, "Behavior Coordination of A Partner Robot based on Imitation", Proc. of 2nd International Conference on Autonomous Robots and Agents, 2004, pp.164-169.

[35] N. Kubota and K. Nishida, "The Role of Spiking Neurons for Visual Perception of A Partner Robot", Proc. 2006 IEEE World Cong. Comp. Intelligence, pp. 530-537, 2006.

[36] 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.

[37] 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.

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

[39] 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.

[40] I. A. Sulistijono and N. Kubota, “Evolutionary Robot Vision and Particle Swarm Optimization for Multiple Human Heads Tracking of A Partner Robot“, Proc. IEEE Cong. Evolutionary Comp., pp. 1535-1541, 2007.

[41] N.Kubota, T.Ohinata, S.Wakisaka, H.Liu, Extraction of Facial Landmarks for a Partner Robot Based on Evolutionary Computation, Proc. Comp. Intelligence, Robotics and Autonomous Sys., Massey University, Palmerston North, New Zealand, November 28-30, pp. 44-49, 2007.