Evolutionary Robot Vision and Fuzzy Evaluation for Natural Communication of Partner Robots

Akihiro Yorita Naoyuki Kubota

Graduate School of System Design, Tokyo Metropolitan University,
Hino, Tokyo 191-0065, Japan
e-mail:yorita-akihiro@sd.tmu.ac.jp, kubota@tmu.ac.jp

Abstract— *This paper proposes a method of evolutionary robot vision based on a steady-state genetic algorithm and fuzzy evaluation. In order to improve the communication capability of human-friendly partner robots, the perception of human face should be performed as correctly as possible. First, we discuss the concept of evolutionary robot vision in dynamic environments. Next, we propose growing neural gas for preprocessing as a bottom-up processing, and steady-state genetic algorithm for template matching in human face recognition as a top-down processing. In order to improve the performance of the human face recognition, we use fuzzy evaluation for evaluating the degree of human face. Finally, we show several experimental results and discuss the effectiveness of the proposed method.*

Keywords— Face Recognition, Evolutionary Computation, Robot Vision, Fuzzy Theory, Partner Robots

1 Introduction

Natural communication and interactions among people and robots have been widely researched [1,2]. Furthermore, various types of human-friendly robots such as pet robots, amusement robots, and partner robots have been developed to communicate with people [3,4]. However, there are still many problems to realize natural communication between people and robots. For example, it is difficult for a robot to identify the speaking person and the place interacting with the person. As a result, the content of utterances from the robot might not be suitable for the person and place.

Relevance theory is helpful to discuss the natural communication between a human and robot [21]. Relevance theory is based on a definition of relevance and two principles of the relevance. One is a cognitive principle that human cognition is geared to the maximization of relevance. The other is a communicative principle that utterances create expectations of optimal relevance. The central claim of relevance theory is that the expectations of relevance raised by an utterance are precise enough, and predictable enough to guide the hearer towards the speaker's meaning. Furthermore, in relevance theory, human thought is shared between two people rather than transmitted. Each person has his or her own cognitive environment. An important role of utterances, facial direction, pointing behaviors, and gestures is to make the hearer pay attention to a specific target object or person. As a result, the cognitive environment of the hearer can be enlarged by the utterances or gestures. The cognitive environment shared between two people is called a mutual cognitive environment. Based on the above discussion, a robot also should have such a cognitive environment, and the robot should keep updating the cognitive environment according to the current perception through the interaction with a human in order to realize the natural communication.

Furthermore, the utterance capability of a robot can be applied for preventing dementia of elderly people. Robotic conversation can activate the brain of such elderly people and can improve their concentration and memory abilities. Nursing care for the elderly people can be expected to keep their health by having conversations with robots. However, it is very difficult to continue the meaningful and attractive conversations with robots. Therefore, such a robot requires adaptive perceptual systems to communicate with a human flexibly, and adaptive action systems to learn human behaviors. To realize the learning through interaction with people, we must consider a total architecture of the cognitive development. The cognitive development for robots has been discussed in the fields such as cognitive robotics and embodied cognitive science [8-10]. In the previous research of cognitive robotics, many researchers have proposed the learning methods for the achievement of joint attention, imitative learning, linguistic acquisition from the viewpoints of babies and infants [6,10]. On the other hand, we focus on the refinement of associative memory by using symbolic information used for utterances and patterns based on visual information through interaction with people as cognitive development of robots. We proposed the concept of structured learning and discussed the importance of total architecture of the learning mechanism [30]. However, we did not discuss the performance of the human detection so much. In the previous works, we proposed a simple method of people tracking based on the combination of skin color and hair color, and we have a problem of misdetection of people by objects with similar color combination in the background image. In this paper, we propose a method for detecting a human face based on evolutionary computation and fuzzy evaluation in order to improve the performance of people tracking. The both of evolutionary computation and fuzzy theory are useful and practical in the search under the environment including noise. We proposed the concept of evolutionary robot vision based on the analogy between visual perception and evolutionary search [32]. We apply the concept of evolutionary robot vision and fuzzy evaluation for people tracking.

The paper is organized as follow. In the section 2, we explain the concept of evolutionary robot vision. Next, in the section 3 we propose a growing neural gas for color extraction and a steady-state genetic algorithm with fuzzy inference for face recognition. Section 4 shows several experimental results and discuss the effectiveness of the proposed method.

2 Evolutionary Robot Vision

2.1 Partner robots

We developed two types of partner robots; a mobile PC called MOBiMac and a human-like robot called Hubot in order to realize the social communication with a human [25,26] (Fig.1). Each robot has two CPUs and many sensors such as CCD camera, microphone, and ultrasonic

ISBN: 978-989-95079-6-8

sensors. Therefore, the robots can conduct image processing, voice recognition, target tracing, collision avoidance, map building, and imitative learning.

We have applied steady-state genetic algorithm (SSGA), spiking neural networks (SNN), self-organizing map (SOM), and others for human detection, motion extraction, gesture recognition, and shape recognition based on image processings [25-28]. However, the image processing takes much computational time and cost. Therefore, we discuss the applicability of fuzzy theory and evolutionary computation in robot vision.

2.2. Active Vision for Robots

Computer vision is a research stream on image processing, image understanding, image recovery and others on a computer [9]. The quality of image depends strongly on the lighting condition related with a camera system. Active vision is often used to improve the robustness and flexibility and to eliminate the ill-posed conditions based on the control of camera systems [9]. Robot vision is deeply related with active vision, because a problem on the perception and action of a robot must be solved at the same time. A robot takes actions to perceive the environment when the robot does not know the environment much. Therefore, the robot vision is based on the time-series of image processing, not the processings on a single image. Various technologies for image processings are required for realizing the robot vision, *e.g.*, color processing, target detection, template matching, shape recognition, motion extraction, and optical flow. Recently, evolutionary computation has been applied to improve the performance of image processing. We also have discussed the applicability of the evolutionary computation in robot vision [32-34]. In fact, we proposed a method of people detection, people tracking and gesture recognition.

2.3 Evolutionary Computation for Robot Vision

Evolutionary computation (EC) is a field of simulating evolution on a computer. Evolutionary optimization methods are fundamentally iterative generation and alternation processes of multiple candidate solutions. The optimization is done by the multi-point search operating on a set of individuals, which is called a population. First, we discuss the role of evolutionary search in dynamic environments. Figure 2 shows the temporal patterns of spatial changes in dynamic environments where the vertical axis indicates the state of environmental conditions represented as a value. If the search speed of EC is faster than the changing speed of the environmental conditions, EC can obtain feasible solutions in the facing environmental conditions. However, EC should be adaptive to the facing environmental conditions if the environmental changes can be observed. If the environmental change is very slow and the change is small (Fig.2 (a)), the mutation range should be large according to the amount of the environmental change. Basically, this kind of change can be considered as some noise in a stationary environment. The environment of Fig.2 (b) includes big changes, but the environmental condition is stationary after a big change. If the big change can be observed, most of candidate solutions should be replaced with randomly generated candidate solutions. The environment of Fig.2 (c) is changing non-stationarily with both features of Figs.2 (a) and (b). In general, the change of visual images corresponds to the environment of Fig.2 (c). The visual image of a mobile robot changes according to both the dynamics of environmental changes and the robotic motion. Therefore,

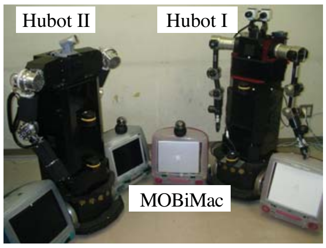


Figure 1: Partner robots; MOBiMac and Hubot

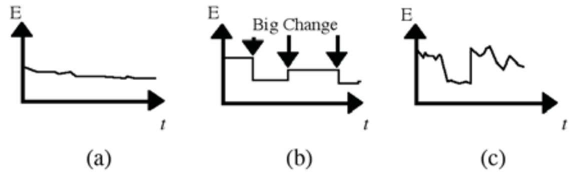
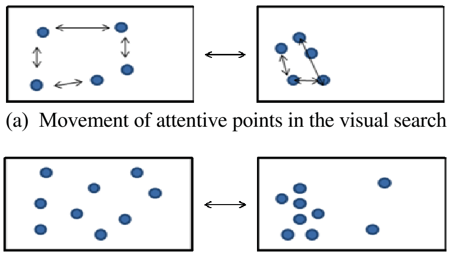


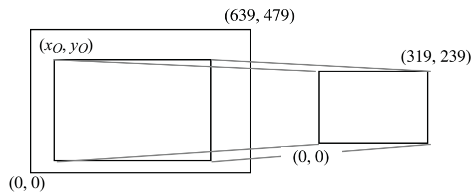
Figure 2: Patterns of changes in dynamic environments



(a) Movement of attentive points in the visual search

(b) Distribution of candidate solutions in evolutionary search

Figure 3: Comparison of visual search and evolutionary search



(a) Original Image (b) Attention range

Figure 4: Preprocessing for human face detection

we discuss visual perception based on active robot vision and EC from the viewpoint of human visual perception.

Figure 3 shows the comparison between the visual search and evolutionary search. The left and right figures show the distribution of search points in the focused search and distributed search, respectively. The visual search controls the searching area based on the fast movement of attentive point. The region of interest (ROI) for the information extraction is deeply related with the search of geometrical features included in the visual target. The search results are reflected to the next visual search. On the other hand, the evolutionary search controls the searching area based on the selection pressure to candidate solutions. If the selection pressure is high, the candidate solutions are centralized toward the better candidate solutions. Otherwise, candidate solutions are globally distributed in the search space. The next search points in the evolutionary search are generated by crossover and mutation. The search in ROI is mainly performed by mutation and local search, while a new search point for ROI is generated by crossover. The degree of interest is calculated by the fitness value. If the fitness value is high, the focused search should be performed. We

ISBN: 978-989-95079-6-8

apply a steady-state genetic algorithm (SSGA) to realize the continuous and real-time search for the robot vision like human visual perception in a dynamic environment, because SSGA can easily obtain feasible solutions through environmental changes with low computational cost.

2.4 Image Processing for Pattern Recognition

Various types of pattern matching methods such as template matching, cellular neural network, neocognitron, and dynamic programming (DP) matching, have been applied for the human detection in image processing. 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 (segmentation), and the aim of the target recognition is to identify the target from classification templates. Furthermore, we can discuss the pattern matching from the viewpoints of the bottom-up processing and top-down processing. The candidate detection is considered as the bottom-up processing based on the color distribution in segments of an image, while the target recognition is considered as top-down processing based on the similarity between the extracted target candidate and classification templates. The synthetic combination of bottom-up processing and top-down processing realizes the efficient and effective search. In this paper, we apply growing neural gas and SSGA for bottom-up processing and top-down processing, respectively.

3. Human Tracking based on Evolutionary Robot Vision

3.1 Growing Neural Gas for Bottom-up Preprocessing

An image is a set of pixels with color information. The segmentation into target candidates based on color distribution is very important to reduce the computational cost of direct template matching. Therefore, we apply an unsupervised clustering method for the segmentation based on color distribution in an image. Growing neural gas (GNG) is a competitive learning network as one of variants of SOM [22] used as a clustering method.

The learning algorithm of GNG is shown as follows.

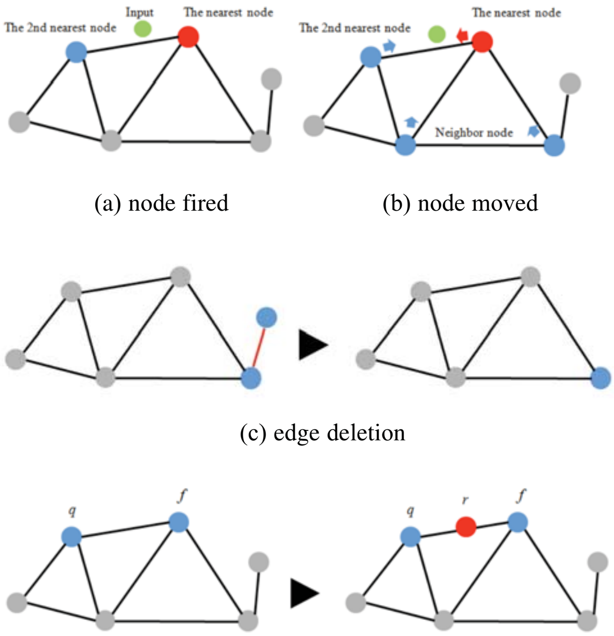
w_i : the nth dimensional vector of a node ($w_i \in \mathbf{R}^{\mathbf{n}}$)
A : a set of nodes
N_i : a set of nodes connected to the ith node
c : a set of edges
$a_{i,j}$: the age of the edge between ith and jth node

Step 0. Generate two units at random position, w_{c1}, w_{c2} in **Rn**. Initialize the connection set.
Step 1. Generate at random an input data v according to $p(v)$.
Step 2. Determine the nearest unit s_1 and the second-nearest unit s_2 by

$$s_1 = \arg\min_{i \in A} \|v - w_i\| \tag{1}$$

$$s_2 = \arg\min_{i \in A \setminus \{s_1\}} \|v - w_i\| \tag{2}$$

where v is composed of the position (x, y) and color information (R, G, B) of color pixel on the image(see Fig.5(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.



(a) node fired (b) node moved

(c) edge deletion

(d) node addition

Figure 5: How to learn GNG nodes and edge

Original Image GNG for Human shape recognition

Figure 6: An example of GNG

$$a_{s_1,s_2} = 0 \tag{3}$$

Step 4. Add the squared distance between the input signal and the winner to a local error variable E_{s1}(see Fig.5(b)).

$$E_{s_1} \leftarrow E_{s_1} + \|v - w_{s_1}\|^2 \tag{4}$$

Step 5. Adapt the reference vectors of the winner and its direct topological neighbors by the learning rate η_1^G and η_2^G, respectively.

$$w_{s_1} \leftarrow w_{s_1} + \eta_1^G \cdot (v - w_{s_1}) \tag{5}$$

$$w_j \leftarrow w_j + \eta_2^G \cdot (v - w_j) \quad if\ c_{s_1,j} = 1 \tag{6}$$

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

$$a_{s_1,j} \leftarrow a_{s_1,j} + 1 \quad if\ c_{s_1,j} = 1 \tag{7}$$

Step 7. Remove edges with the age larger than $a_{\max}$. If units have no more emanating edges after this, remove those units(see 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(see Fig.5(d)).
Step 8-1. Determine the unit q with the maximum accumulated error.

$$q = \arg\max_{i \in A} E_i \tag{8}$$

Step 8-2. Determine the unit f with the maximum accumulated error among the neighbors of q.

$$f = \arg\max_{c \in N_q} E_c \tag{9}$$

Step 8-3. Add a new unit r to the network and interpolate its reference vector from q and f.

ISBN: 978-989-95079-6-8

$$w_r = 0.5 \cdot \left(w_q + w_f\right) \quad (10)$$

Step 8-4. Insert edges connecting the new unit r with units q and f, and remove the original edge between q and f.
Step 8-5. Decrease the error variables of q and f by the discount rate α.

$$E_q \leftarrow (1-\alpha)E_q \text{ and } E_f \leftarrow (1-\alpha)E_f \quad (11)$$

Step 8-6. Interpolate the error variable of r from q and f

$$E_r = 0.5 \cdot \left(E_q + E_f\right) \quad (12)$$

Step 9 Decrease the error variables of all units

$$E_c \leftarrow (1-\beta)E_c \,,\ c \in A \quad (13)$$

Step 10 If a termination condition (e.g., some performance measure) is not yet fulfilled continue with step 2.

Figure 6 shows an example of the learning of GNG. In this way, the color distribution can be extracted from the image by using GNG.

3.2 People Detection and Tracking as Top-down Processing

SSGA simulates a continuous model of the generation, which eliminates and generates a few individuals in a generation (iteration) [13]. The genotype is represented by $g_{i,j}$ (i=1,2,..., G, j=1,2,..., M) and fitness value is represented by f_i. One iteration is composed of selection, crossover, and mutation. The worst candidate solution is eliminated ("Delete least fitness" selection strategy), and is replaced with the candidate solution generated by the crossover and the mutation.

We use the elitist crossover and adaptive mutation [15]. The elitist crossover randomly selects one individual and generates an individual by combining genetic information from the selected individual and the best individual with the crossover probability. If the crossover probability is satisfied, the elitist crossover is performed. Otherwise, a simple crossover is performed between two randomly selected individuals. Next, the following adaptive mutation is performed to the generated individual,

$$g_{i,j} \leftarrow g_{i,j} + \left(\alpha_j \cdot \frac{f_{\max} - f_i}{f_{\max} - f_{\min}} + \beta_j\right) \cdot N\left(0,1\right) \quad (14)$$

where f_i is the fitness value of the ith individual, f_{max} and f_{min} are the maximum and minimum of fitness values in the population; $N(0,1)$ indicates a normal random variable with a mean of zero and a variance of one; α_j and β_j are the coefficients ($0<\alpha_j<1.0$) and offset ($\beta_j>0$), respectively. In the adaptive mutation, the variance of the normal random number is relatively changed according to the fitness values of the population in case of maximization problems.

The robot must recognize a human face from complex background speedily. Therefore, we use SSGA for human detection as one of search methods. The human face candidate positions based on human skin and hair colors are extracted by SSGA with template matching. Figure 7 shows a candidate solution of a template used for detecting a human face. A template is composed of numerical parameters of $g_{i,1}^H$, $g_{i,2}^H$, $g_{i,3}^H$, and $g_{i,4}^H$. The number of individuals is G^H. The initial population of SSGA for human detection at the discrete time step t is updated by using the reference vector of GNG in addition to the candidate solutions obtained at the previous time step t-1. The fitness value of the ith individual is calculated by the following equation,

$$f_i^H = C_{Skin}^H + C_{Hair}^H + \eta_1^H \cdot C_{Skin}^H \cdot C_{Hair}^H - \eta_2^H \cdot C_{Other}^H \quad (15)$$

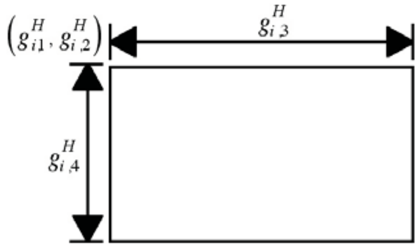
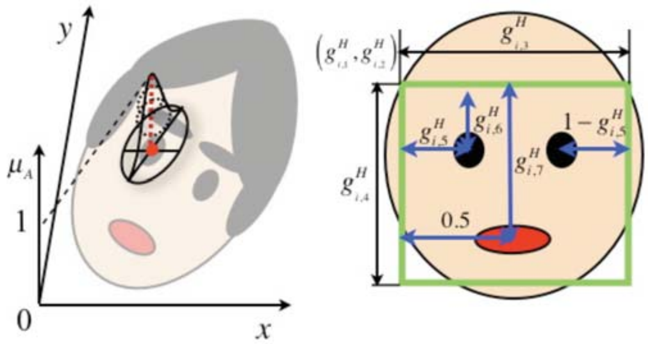


Figure 7: A template used for human detection in SSGA



(a) Fuzzy Evaluation based on Gaussian membership (b) The position of facial landmarks

Figure 8: A template used for face recognition

where C^H_{Skin}, C^H_{Hair} and C^H_{Other} indicate the numbers of pixels of the colors corresponding to human skin, human hair, and other colors, respectively; η_1^H and η_2^H are the coefficients (η_1^H, $\eta_2^H>0$). Therefore, this problem results in the maximization problem. The iteration of SSGA is repeated until the termination condition is satisfied. Here SSGA for the human detection is called SSGA-H.

Since SSGA-H extracts the area of skin colors and hair colors in the human detection, various objects except humans might be detected. Therefore, the human tracking is performed according to the time series position of the ith human candidate ($g_{i,1}^H$, $g_{i,2}^H$) obtained by SSGA-H. The position of the jth human candidate in the human tracking ($X_{k,1}$, $X_{k,2}$) is updated by the nearest human candidate position within the tracking range. In addition, the width and height of the human candidate for the human tracking ($X_{k,3}$, $X_{k,4}$) are updated by the size of the detected human ($g_{i,3}^H$, $g_{i,4}^H$). The update is performed as follows (j=1,2,3,4);

$$X_{k,j} = (1-\lambda)X_{k,j} + \lambda \cdot x_{i,j} \quad (16)$$

Furthermore, the time counter for the reliability of human tracking is used. If the human candidate position in the human tracking is performed, the time counter is incremented. Otherwise, the time counter is decremented. If the time counter is larger than the threshold (HT), the human count is started. Sometimes, several human candidates are close each other, because several human candidates in a single human can be generated by the human detection. Therefore, the removal processing is performed when human candidates are coexisting within the tracking range.

3.3 Fuzzy Evaluation for Face Recognition

The human detection based on color distribution sometime extracts something with similar color distribution. In order to improve the performance of human face recognition, we

ISBN: 978-989-95079-6-8

Table 1: Initial values used in Gaussian membership function

	$a_{h,1}$	$a_{h,2}$	$b_{h,1}$	$b_{h,2}$
Right eye	0.2	0.2	0.4	0.2
Left eye	0.8	0.2	0.4	0.2
Mouth	0.5	0.6	0.4	0.37

Table 2: Search range of $g_{i,5}^H$, $g_{i,6}^H$, $g_{i,7}^H$

	$g_{i,5}^H$	$g_{i,6}^H$	$g_{i,7}^H$
min	0.1	0.15	0.7
max	0.35	0.35	0.9

can use the color information of facial landmarks. The human detection by SSGA can be considered as a coarse search, and the extraction of facial landmarks can be considered as a fine search. We apply fuzzy evaluation for face recognition based on the position of facial landmarks (Fig.8 (a)). We use the position of eyes and mouth. We use

$$\mu_{A(h,i,j)} = \begin{cases} \exp\left(-\dfrac{(x-a_{h,1})^2}{2b_{h,1}{}^2} - \dfrac{(y-a_{h,2})^2}{2b_{h,2}{}^2}\right) & if\ p(i,j) = c_h \\ 0 & otherwise \end{cases} \tag{17}$$

where (x,y) is the normalized position of the pixel (i,j) on the image; $p(i,j)$ is the color ID of a the pixel (i,j) on the image; c_h is the color ID of the hth facial landmark is the center of a facial landmark; $(a_{h,1},a_{h,2})$ and $(b_{h,1},b_{h,2})$ are the normalized position and size of the hth facial landmark in the template candidate extracted by SSGA. Therefore, this value is high if the color pixel corresponding to the facial landmark is near with the center of the facial landmark. Therefore, we can evaluate the degree of existing each facial landmark as follows;

$$f_{Land,h} = \sum_{(i,j)\in g_k} \mu_{A(h,i,j)} \tag{18}$$

where g_k is the template of the kth candidate solution in SSGA; h is facial the landmark ID. Furthermore, we can evaluate the degree of face as follows;

$$f_{Face} = \prod_{h=1}^{H} f_{Land,h} \tag{19}$$

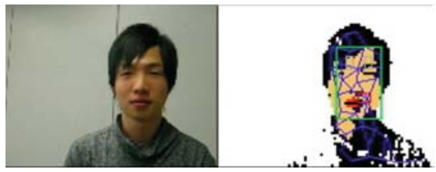
where H is the number of facial landmarks. We use right eye (h=1), left eye (h=2) and mouth (h=3) for the evaluation (H=3). Figure 8 (b) shows the positions of facial landmarks where $(g_{i,5}^H, g_{i,6}^H)$ is the position of the right eye and $(0.5, g_{i,7}^H)$ is the position of the mouth. Therefore, $(a_{1,1},a_{1,2}) = (g_{i,5}^H, g_{i,6}^H)$, $(a_{2,1},a_{2,2}) = (1-g_{i,5}^H, g_{i,6}^H)$, and $(a_{3,1},a_{3,2}) = (0.5, g_{i,7}^H)$. However, the position of each facial landmark is peculiar to a person. Therefore, the position of membership function corresponding to each facial landmark should be updated according to the detected person. As a result, the number of parameters for human detection is 7.

4 Experimental results

This section shows experimental results of human recognition of a partner robot. The maximal number of nodes in GNG is 50. The population size of SSGA is 50. The number of generations (iterations) SSGA in each frame is 200 including the initial evaluations after the frame of image is updated. This value is relatively small comparing with that of the search by standard GA, but the search by SSGA is a time series of continuous search in a dynamic environment including a small change. Table 1 shows the initial values of shape information used in fuzzy evaluation, and Table 2 shows the search range of face parameters used in SSGA.

Original Image Human Tracking

(a) Case 1



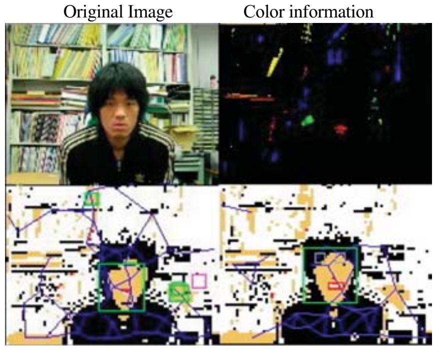
Original Image Human Tracking

(b) Case 2

Original Image Human Tracking

(c) Case 3

Original Image Color information



Templete matching Template with Fuzzy inference

(d) Case 4

Figure 9: Experimental results of face recognition

Table 3: The values of human face obtained in experimental results of Cases 1 and 2

	$g_{i,5}^H$	$g_{i,6}^H$	$g_{i,7}^H$
case 1	0.31	0.19	0.77
case 2	0.26	0.27	0.78

Figure 9 shows experimental results of the proposed method. First of all, we conducted experiments on face recognition of two different people (Case 1 and Case 2). Figure 9.(a) and (b) show snapshots of original image (left), and human-object detection results by SSGA (right) where a

ISBN: 978-989-95079-6-8

green box indicates the extracted human face. Table 3 shows the values of human face obtained in experimental results of Cases 1 and 2. In these results, the proposed method can extract human face with identifying the position of facial landmarks correctly.

Next, we conducted experiments on face recognition where the image includes skin color of object (Case 3) and the image includes the complicated background (Case 4). Figure 9 (c) shows experimental results of the proposed method in Case 3. The proposed method can extract a human face, although the person wears a helmet with having a skin-color book. Figure 9 (d) shows experimental results of the proposed method in Case 4; original image (upper-left), color information (upper-right), template matching without fuzzy evaluation (lower-left), and template matching with fuzzy evaluation (lower-right). The proposed method succeeds to extract a human face correctly. On the other hand, the template matching without fuzzy evaluation fails to extract human face, because some books located in the left of the person on the background image are recognized as a person. In this way, the proposed method can extract a human face in various environmental conditions.

5 Conclusions

This paper proposed a method of human face detection based on the template matching with a steady-state genetic algorithm and fuzzy evaluation. The experimental results show the effectiveness of the proposed method. Membership functions are very useful to evaluate candidate solutions including noisy data. The essence of the proposed method in the flexibility of the search by combining steady-state genetic algorithm and fuzzy evaluation.

As a future work, we will develop a method of human face detection in case of rotation of human face. Furthermore, we apply the proposed method to the associative learning between the perceptual information and symbolic information peculiar with the interacting person.

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] J.A.Anderson, E.Rosenfeld,. *Neurocomputing*, The MIT Press, Cambridge, Massachusetts, US , 1988.

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

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

[8] K.Fukushima 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, 2003.

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

[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 edition, Massachusetts, The MIT Press, 1992.

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

[13] G.Syswerda, G. A study of reproduction in generational and steady-state genetic algorithms, *Genetic Algorithms*, Morgan Kaufmann Publishers Inc., San Mateo, 1991.

[14] S.Watanabe, T.Hiroyasu, M.Miki, LCGA : Local Cultivation Genetic Algorithm for Multi-Objective Optimization Problems, Proceedings of the Genetic and Evolutionary Computation Conference (GECCO2002), pp.702, 2002.

[15] M.Lozano, F.Herrera, N.Krasnogor, D.Molina, Real-Coded Memetic Algorithms with Crossover Hill-Climbing, Evolutionary Computation, Vol. 12, No. 3, Pages 273-302, MIT Press, Cambridge, Massachusetts, US, 2004.

[16] J.Branke, Evolutionary Optimization in Dynamic Environments, Volume 3 of the Book Series on Genetic Algorithms and Evolutionary Computation, Kluwer Academic Pub Published, 2001.

[17] R. C. Eberhart, J. Kennedy, Y. Shi, Swarm Intelligence, Morgan Kaufmann Publisher, San Francisco, 2001.

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

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

[20] Indra A.Sulistijono and N.Kubota, Human Head Tracking Based on Particle Swarm Optimization and Genetic Algorithm, Journal of Advanced Computational Intelligence and intelligent Informatics, Vol.11 No.6, Jul. 2007, pp.681-687, 2007

[21] D.Sperber, D.Wilson, Relevance - Communication and Cognition, Blackwell Publishing Ltd., 1995.

[22] B. Fritzke. A growing neural gas network learns topologies. In G. Tesauro, D. S. Touretzky, and T. K. Leen, editors, Advances in Neural Information Processing Systems 7, pages 625-632. MIT Press, Cambridge MA, 1995.

[23] M.Hynar, M.Burda, J.Sarmanova: Unsupervised clustering with growing self-organizing neural network -- a comparison with non-neural approach. DATESO pp.58-68, 2005

[24] B. Fritzke. Growing cell structures - a self-organizing network for unsupervised and supervised learning. Technical Report ICSTR-93-026, International Computer Science Institute, Berkeley, May 93.

[25] N.Kubota, K.Nishida, Cooperative Perceptual Systems for Partner Robots Based on Sensor Network, International Journal of Computer Science and Network Security, Volume 6, Number 11, pp.19-28, 2006.

[26] N.Kubota, Visual Perception and Reproduction for Imitative Learning of A Partner Robot, WSEAS Transaction on Signal Procesing, Issue 5, Vol. 2, pp. 726-731, 2006.

[27] N.Kubota, Computational Intelligence for Structured Learning of A Partner Robot Based on Imitation, *Information Science*, No. 171, pp.403-429, 2005.

[28] N.Kubota; Y.Nojima, F.Kojima, T.Fukuda,. Multiple Fuzzy State-Value Functions for Human Evaluation through Interactive Trajectory Planning of a Partner Robot, *Soft Computing,* Vol.10, No.10 pp.891-901, 2006.

[29] W. Zhao, R. Chellappa, J. Phillips, and A. Rosenfeld,"Face Recognition: A Literature Survey", ACM Computing Survey, December Issue, pp.399-458, 2003.

[30] N.Kubota, A.Yorita, Structured Learning for Partner Robots based on Natural Communication, Proc. (CD-ROM) of 2008 IEEE Conference on Soft Computing in Industrial Applications (SMCia), pp. 303-308, 2008.

[31] N.Kubota, A.Yorita, M.Satomi, Visual Tracking of A Partner Robot through Interaction with People, Proc. of Joint 4th International Conference on Soft Computing and Intelligent Systems and 9th International Symposium on advanced Intelligent Systems (SCIS & ISIS 2008), pp. 1032-1037, 2008.

[32] N.Kubota, Y.Tomioka, Evolutionary Robot Vision for Human Tracking of Partner Robots in Ambient Intelligence, Proc (CD-ROM) of Congress of Evolutionary Computation, pp. 1491-1496, 2007.

[33] N.Kubota, T.Ohinata, S.Wakisaka, H.Liu, Extraction of Facial Landmarks for a Partner Robot Based on Evolutionary Computation, Proc. of Computation Intelligence, Robotics and Autonomous Systems (CIRAS 2007), pp. 44-49, 2007.

[34] Indra A.Sulistijono, and N.Kubota, Evolutionary Robot Vision for Multiple Human Heads Tracking of A Partner Robot, Proceeding of 16th IEEE International Conference on Robot & Human Interactive Communication (Ro-Man2007), pp. 463-468, 2007.

ISBN: 978-989-95079-6-8