Human Hand Detection for Gestures Recognition of A Partner Robot

Setsuo HASHIMOTO*, Naoyuki KUBOTA**, and Fumio KOJIMA***

*Faculty of Economics, Kyoto Gakuen University
1-1 Ootani, Nanryo, Sogabe-cho, Kameoka, Kyoto, 621-8551 Japan
email: setsuo-h@kyotogakuen.ac.jp

**Dept. of System Design, Tokyo Metropolitan University
1-1 Minami-Ohsawa, Hachioji, Tokyo 192-0397, JAPAN

***“Interaction and Intelligence,” PRESTO, Japan Science and Technology Agency (JST)
email: kubota@comp.metro-u.ac.jp

*** Dept. of System Function Science, Graduate School of Science and Technology, Kobe University
1-1 Rokkodai-cho, Nada-ku, Kobe 657-8501, JAPAN
email: kojima@cs.kobe-u.ac.jp

Abstract - This paper proposes a human hand detection method for gesture recognition used for communication of a robot with a human. A human hand motion can be recognized as a different meaning according to a situation in the communication between the human and robot. First, we propose a steady-state genetic algorithm for extracting a time series of human hand position, and next, a gestures recognition method of a robot composed of a spiking neural network and self-organizing map. Finally, we discuss the effectiveness of the proposed method through several experimental results.

I. INTRODUCTION

Recently, various types of image processing methods have been developed as computational capabilities increase [1-4]. Thanks to real-time image processing, we can recognize both human expression according to the motion of body, face and hand. Furthermore, not only image processing, but also speech recognition technology has been improved. The integration of human motion recognition and speech recognition enables a robot to do a lot of things in human-friendly communication. Actually, various kind human-friendly robots has been cheaply produced and sold. However, a human hand motion, or a gesture can be recognized as a different meaning according to a situation in the communication between the human and the robot. In general, the gesture plays an important role in interaction and communication with a human, because the meanings of utterance can be emphasized by using gestures [8,9]. Furthermore, it is very natural and useful for a human to use a gesture in order to give the robot a specific task to share their attention. Therefore, a gesture is regarded as a symbolic action for communicating intention to the other. Accordingly, the robot needs to recognize gestures.

Various types of gesture recognition methods have been proposed so far [5-7]. The research stream can be classified into hand shape recognition and hand motion recognition. Since a hand is a complex object, it is very difficult to detect a hand from an image. Therefore, many researchers have used a simple background, colored gloves, and others. In general, skin color information is used for detecting a hand, must distinguish the hand from its similar color of objects. Therefore, a hierarchical detection method of skin color region extraction, hand extraction, and hand shape recognition has been used in order to reduce computational cost. Hand shape recognition is performed by using hand contour or 3D hand model. On the other hand, hidden Markov Model has been used for extracting gesture sequence as one of hand motion recognition. If knowledge database of gesture sequence patterns is available in hand motion recognition, template candidates used in gesture recognition can be reduced. We should use both of hand shape recognition and hand motion recognition simultaneously. We also proposed alternative method of hand extraction and gesture recognition based on computational intelligence [10,11]. In our method, we used a blue glove to reduce the computational cost.
Therefore, this paper proposes a method of human hand detection based on skin color and edge information. Next, we propose two methods of gesture recognition. The first one is used to extract the posture of the human arm. The position and posture of the human arm includes the meaning to be communicated. We apply a steady-state genetic algorithm [12,13] to identify the posture or configuration of the human arm. The other is used to extract a human hand motion. The dynamics of the human hand motion itself includes the meaning to be communicated. We apply a spiking neural network [16,17] for extracting a human hand motion and a self-organizing map [18,19] for extracting motion patterns. Finally, we show experimental results of gesture recognition for a partner robot, and discuss the effectiveness of the proposed method.

II. A Partner Robot

We developed a partner robot; MOBiMac as shown in Figure 1. Two CPUs are used for PC and robotic behaviors. The robot has two servo motors, eight ultrasonic sensors, and CCD camera. Therefore, the robot can take various actions such as collision avoiding, human approaching, and line tracing. The behavior modes used for this robot are human detection, human communication, behavior learning, behavioral interaction. The communication with a human is performed by the utterance as the result of voice recognition and human motion recognition. The behavior learning includes the reinforcement learning through interaction with the environment, and imitative learning through interaction with the human. The behavioral interaction includes the soccer and games with a human. In the following, we focus on image processing for gesture recognition.

The robot takes an image from the CCD camera, and extracts a human. If the robot detects the human, the robot extracts the motion of the human hand. According to the human hand motion, the robot decides the action outputs. Furthermore, the robot expresses the internal or perceptual state by utterance. A behavior of the robot can be represented using fuzzy rules based on simplified fuzzy inference [14,15]. The logical structure written by fuzzy rules is easy for humans to understand and to design. In general, a fuzzy if-then rule is described as follows,

\[
\text{If } x_1 \text{ is } A_{i,1} \text{ and } \ldots \text{ and } x_m \text{ is } A_{i,m} \\
\text{Then } y_1 \text{ is } w_{i,1} \text{ and } \ldots \text{ and } y_n \text{ is } w_{i,n}
\]

where \(A_{i,j}\) and \(w_{i,k}\) are a symmetric triangular membership function for the \(j\)th input and a singleton for the \(k\)th output of the \(i\)th rule; \(m\) and \(n\) are the numbers of inputs and outputs, respectively. Fuzzy inference is generally described by,

\[
\mu_{w_{i,j}}(x_j) = \begin{cases} 
1 - \frac{|x_j - a_{i,j}|}{b_{i,j}} & |x_j - a_{i,j}| \leq b_{i,j} \\
0 & \text{otherwise}
\end{cases}
\]

\[
\mu_i = \prod_{j=1}^{m} \mu_{w_{i,j}}(x_j)
\]
\[ y_k = \frac{\sum_{i=1}^{g} \mu_i w_{i,k}}{\sum_{i=1}^{g} \mu_i} \]  

(3)

where \(a_{ij}\) and \(b_{ij}\) are the central value and the width of the membership function \(A_{ij}\); \(R\) is the number of rules. Outputs of the robot are motor output levels. Fuzzy controller is used for collision avoidance and target tracing behaviors. The inputs to the fuzzy controller for collision avoidance and target tracing are the measured distance to the obstacle by ultrasonic sensors, and the relative direction and to a target point, respectively. Basically, a target point is generated by using the humans and objects on the image. The gesture recognition is composed of three stages; (1) human hand detection from a single image, (2) human arm posture extraction from a single image, and (3) human hand motion extraction from temporally sequential images (Fig 2). In this paper, we explain about (1) the human hand detection method in the following sections.

III. HUMAN HAND DETECTION

The image of RGB color space is taken by CCD camera. Because the image processing takes much time and computational cost, the full size of image processing to every image is not reasonable. Therefore, we use the reduced size of image to detect a moving object, that is recognizable as a human. First, the robot calculates the center of gravity of the pixels different from the previous image as the differential extraction. The size of image used in the differential extraction is 20×15. The attention range is formed according the center of gravity. Next, the colors corresponding to human hair and skin are extracted by using thresholds. The region including the human face and hair colors are detected by using a steady-state genetic algorithm (SSGA) based on template matching. Figure 3 (a) shows a candidate solution of the template used for detecting a human as the target object. In SSGA, only few existing solutions are replaced with the candidate generated by genetic operators in each generation. In this paper, the worst candidate solution is eliminated (“Delete least fitness” selection), and it is replaced with the candidate solution generated by the crossover and the mutation.

A human hand is detected under the lighting condition of a finger edge. Figure 4 shows a candidate solution for detecting a human hand. A template is composed of numerical parameters of \(g_{i,1}\), \(g_{i,2}\), \(g_{i,3}\), and \(g_{i,4}\). Here, \(g_{i,1}\), \(g_{i,2}\), \(g_{i,3}\), and \(g_{i,4}\) indicate the starting point of the finger edge, the angle, and the length of the edge respectively. The number of individuals is \(G\).

First, the pixel corresponding to an edge is extracted by calculating difference among neighboring pixels as preprocessing. Here the length of edge is gradually extended according to the change of fitness values. The fitness value is calculated by the following equation,

\[ f_{i}^{\text{fit}} = C^\text{sk} + C^\text{ed} + \eta_1 \cdot C^\text{sk} \cdot C^\text{ed} - \eta_2 \cdot C^\text{other} \]  

(6)

where \(C^\text{sk}\) and \(C^\text{ed}\) indicate the number of pixels of the colors corresponding to human hand (skin) and the edge, respectively; \(\eta_1\) and \(\eta_2\) are coefficients. Therefore, this problem result in the maximization problem. We use the elitist crossover and adaptive mutation. The elitist crossover randomly selects one individual and generates an individual by combining genetic information from the randomly selected individual and the best individual. Next, the following adaptive mutation is performed to the generated individual,

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

(5)

where \(f_i\) is the fitness value of the \(i\)th individual; \(f_{\text{max}}\) and \(f_{\text{min}}\) are the maximum and minimum of fitness values in the population; \(N(0,1)\) indicates a normal random value; \(\alpha_j\) and \(\beta_j\) are the coefficient and offset, respectively. In the adaptive mutation, the variance of the normal random number is relatively changed according to the fitness values of the population. The pixel corresponding to an edge is extracted by calculating difference among neighboring pixels.
IV. EXPERIMENTAL RESULT
This section shows several experimental results of behavior learning using the partner robot MOBiMac. The size \((X,Y)\) of an image is \((160, 120)\). The number of individuals in SSGA is 60, and the number of iterations in SSGA is 3000, respectively. Figure 5 show an image processing result of extraction of a human hand. The human hand moves in the original image as shown in Fig.5 (a). By using SSGA, the robot extracts human hand with dynamic motion.

V. SUMMARY
This paper proposes a human hand detection method for gestures recognition of a partner robot based on human arm posture and human hand motion. We applied steady-state genetic algorithm (SSGA) for image processing. Human arm posture is static information, while the human hand motion is dynamic position. These can be used for expressing the human feeling and pointing direction, respectively. Experimental results show that the robot can extract the human hand motion dynamically.

We will develop the method for extracting the meanings of gestures through inter-action with a human. Furthermore, as a future work, we intend to incorporate the obtained action patterns into reactive motions of the robot interacting with human.

REFERENCES
Fig. 5 Image processing results