

Paper:

# Human Head Tracking Based on Particle Swarm Optimization and Genetic Algorithm

Indra Adji Sulistijono<sup>\*,\*\*</sup> and Naoyuki Kubota<sup>\*\*\*,\*\*\*\*</sup>

<sup>\*</sup>Dept. of Mechanical Engineering, Graduate School of Engineering, Tokyo Metropolitan University

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

<sup>\*\*</sup>Electronics Eng. Polytechnic Institute of Surabaya - ITS (EEPIS-ITS), Kampus ITS Sukolilo, Surabaya 60111, Indonesia

E-mail: [indra-adj@ed.tmu.ac.jp](mailto:indra-adj@ed.tmu.ac.jp)

<sup>\*\*\*</sup>Dept. of System Design, Tokyo Metropolitan University

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

<sup>\*\*\*\*</sup>SORST, Japan Science and Technology Agency (JST)

E-mail: [kubota@comp.metro-u.ac.jp](mailto:kubota@comp.metro-u.ac.jp)

[Received January 17, 2007; accepted March 20, 2007]

**This paper compares particle swarm optimization and a genetic algorithm for perception by a partner robot. The robot requires visual perception to interact with human beings. It should basically extract moving objects using visual perception in interaction with human beings. To reduce computational cost and time consumption, we used differential extraction. We propose human head tracking for a partner robot using particle swarm optimization and a genetic algorithm. Experiments involving two maximum iteration numbers show that particle swarm optimization is more effective in solving this problem than genetic algorithm.**

**Keywords:** human head tracking, particle swarm optimization, steady-state genetic algorithm, visual perception, partner robot

## 1. Introduction

Visual perception is important for the partner robots, because it may provide information useful in interaction with human partners. Partner robots entertaining, assisting or otherwise interacting with human users should be able to recognize human beings, interact with them using natural communication, and learn how human partners might want to interact.

Most human detection involves detecting single human or face images, assuming that human beings are localized in individual frames based on human models – shape, contrast, and color. Robots could determine human intentions using built in maps, because tasks depend on environmental conditions. Beyond environmental map learning, robots should determine human gestures and postures in communicating with users. Previous studies have proposed using image processing for robotic visual perception such as differential filters, moving object detection, and pattern recognition [16, 17]. For pattern recognition, robots require patterns or templates, but a human-friendly

robot cannot know patterns or templates beforehand so they must learn patterns and templates while interacting with users. Previous research has had a robot detect a human using self-organizing map (SOM) based on color patterns and location of a human obtained from the SOM nodes [25, 26].

We made a multi purpose partner robot, one purpose being to communicate with a user. To reach this goal, the robot's visual perception focuses on face recognition and human head tracking, i.e., head location and extraction.

Particle swarm optimization (PSO) and genetic algorithm (GA) are two widely used types of evolutionary computation (EC). As are other EC techniques, PSO is a population-based search algorithm and is initialized with a population of random solutions, called particles. Unlike the other evolutionary computation techniques, each PSO particle is associated with a velocity. Particles fly through the search space with velocities dynamically adjusted based on their historical behaviors. Particles tend to fly toward increasingly better search areas during a search [18]. Many research results have been reported [11, 12, 20–22]. GAs are adaptive optimization techniques that simulate the mechanics of genetic evolution and have been applied to a wide variety of problems [13, 14]. We propose using PSO or GA in head tracking, to reduce computational cost. The human search problem is formulated into a non-linear integer optimization problem. Integer variables are parameters that represent a subwindow in an image as a template. The objective function involves maximizing the fitness function. First, the robot extracts a human from an image taken by a built-in charge-coupled device (CCD) camera. Assuming that human beings move, moving objects become candidates for human beings. A differential extraction area is formed to reduce time cost consumption and the search area. The robot searches for head using a human head template. Our results showed that PSO performs better in some cases while GAs perform better in others.

This paper is organized as follows. Section 2 introduces a vision-based partner robot, and proposes human head

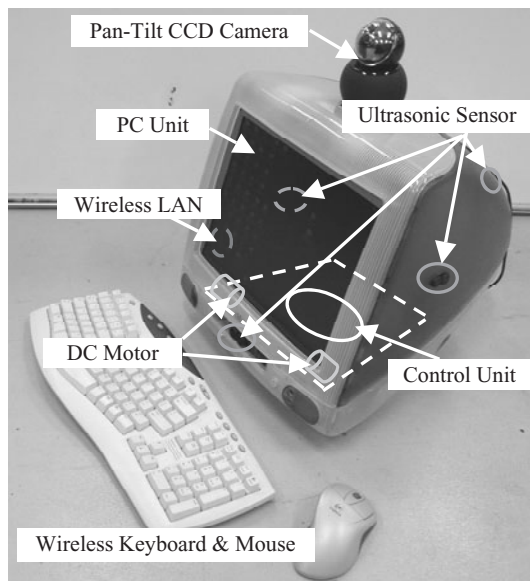


Fig. 1. MOBiMac partner robot.

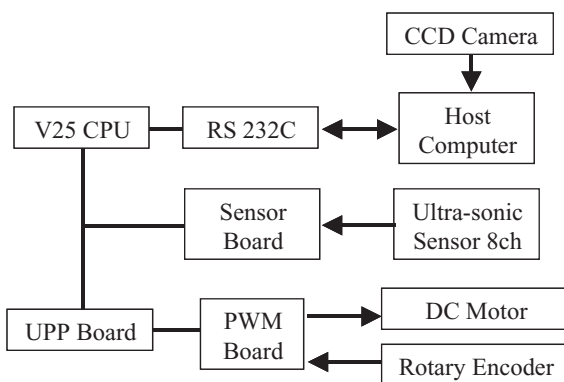


Fig. 2. MOBiMac control architecture.

tracking based on visual perception. Section 3 details experimental results and Section 4 presents conclusions.

## 2. Visual Perception

### 2.1. Partner Robot

We developed a MOBiMac partner robot (Fig. 1) to be used as a personal computer and partner. One of two CPUs is used for PC and the other for robotic behavior. The robot has two servomotors, four ultrasonic sensors, and a CCD camera (Fig. 2). The ultrasonic sensor measures 2000 mm, so the robot captures behavior such as collision avoidance, human approach, and visual tracking.

The robot uses CCD camera to extract a human beings and, it detects them, extracts color patterns in visual tracking (Fig. 3), detailed below.

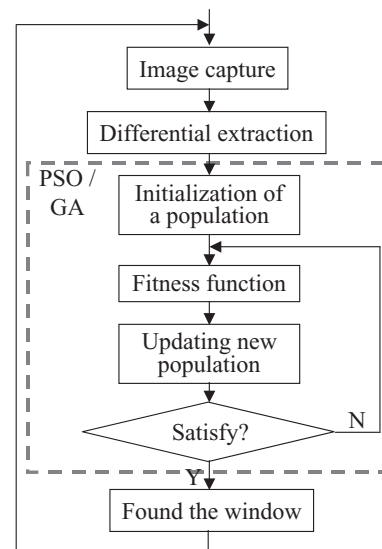


Fig. 3. Visual tracking architecture.

### 2.2. Differential Extraction for Attention Range

Using two continuous images to detect moving objects and set the attention range reduces computational costs and improves the detection accuracy in images with complex backgrounds.

We use differential extraction to form the attention range. Because image processing uses much time and computational cost, using full-sized of image processing is impractical, so we use reduced images. First, the robot calculates the center of gravity of pixels differing from the previous image in differential extraction, using  $40 \times 30$  pixel images. Once a human being is detected, the attention range is formed based on the center of gravity of the pixel of the moving object for horizontal and vertical axes, defining the highest pixel detected as the top range.

Using RGB color space image taken by CCD camera, we extract colors corresponding to hair and face using thresholds.

### 2.3. Particle Swarm Optimization

PSO, developed by Kennedy and Eberhart in 1995, was inspired by the social behavior of a flock of migrating birds trying to reach an unknown destination. In PSO, each solution is a bird in the flock, referred to as a particle. Birds in the population evolve in social behavior and movement toward a destination [11].

Physically, this mimics a flock of birds that mutually communicate as they fly. Each bird looks in a specific direction, and when communicating identify the bird in the best location. Each bird proceeds toward the best bird using a velocity depending on its current position. Each bird then, studies the search space from its new local position, and the process is repeated until the flock reaches a desired destination. Note that this involves both social interaction and intelligence, meaning that birds learn from experience (local search) and from the experience of others around them (global search).