https://doi.org/10.20965/jrm.2008.p0515

Paper:

Target Identification Through Human Pointing Gesture Based on Human-Adaptive Approach

Yusuke Tamura, Masao Sugi, Tamio Arai, and Jun Ota

Research into Artifact, Center for Engineering (RACE), The University of Tokyo
5-1-5 Kashiwanoha, Kashiwa-shi, Chiba 277-8568, Japan
E-mail: tamura@race.u-tokyo.ac.jp

[Received January 31, 2008; accepted June 24, 2008]

We propose a human-adaptive approach for calculating human pointing targets, integrating (1) calculating the user's subjective pointing direction from finger direction, (2) integrating sensory information obtained from user pointing and contextual information such as user action sequences, and (3) arranging target candidates based on the user's characteristics of pointing and action sequences. The user's subjective pointing direction is approximated by the linear function with the finger direction. Integration of sensory and contextual information using a probabilistic model enables the system to calculate the target accurately. Using a force-directed approach, we obtained good placement in which false estimations are decreased and not moved much from initial placement. Experimental results demonstrate the usefulness of our proposal.

Keywords: pointing, context, epistemic action, human-robot interface

1. Introduction

Over the last decade, several studies have been made on intelligent robotic systems that support everyday living at home or in an office setting [1, 2].

People typically spend significant time at there desks, doing computer work, reading and writing documents, letters, and books, eating lunch, and assembling objects. Therefore supporting individuals who work at desks by using a robotic system could have a great deal of benefit.

The authors proposed the design of an Attentive Workbench (AWB) that helps people work at their desks [3].

AWB has following three key components:

- **EnhancedDesk:** The EnhancedDesk is an augmented desk interface [4]. The user expresses intention by hand gestures, and the system presents information using an LCD projector and plasma display.
- **Self-moving trays:** The self-moving trays, which are driven by a Sawyer-type 2-DOF stepping motor [5], deliver necessary objects and clear unnecessary objects. Each tray has square shape with a side of 80 mm, and is characterized with high speed (0.8 m/s at maximum) and high positioning accuracy (about 40 μm).
- **Estimation of a worker's state based on biomeasurement technologies:** A worker's state can be estimated from heart rate and respiration measured by vital-signs monitors. It applies a method for analyzing respiratory sinus arrhythmia (RSA) with respect to respiratory phase [6].

Considering a robot system such as the AWB that supports people where they live or work, it is inevitable that humans and robot will interact. Especially, in the home or office, an intuitive way of instructing a robotic system will be necessary if they are to be operated by ordinary people.

In this study, we focused on finger pointing, which is deictic and intuitive gesture. We plan to employ such a gesture to give instructions to an intelligent system. Pointing is often used to indicate a specific object or location in interpersonal communication; however, it is not always easy, even for humans, to identify the object at which someone is pointing.

Here, we assume the following environment (**Fig. 1**):

- Many objects are placed on a desk.
- A user sits on a chair at the desk.
- The user points at a target object.
- The target object among many is identified and the system acts as the user indicates.

Pointing has been long and widely studied in the input device context. In what follows, we briefly review related work in pointing gestures.

A pioneering work on the use of pointing gestures as an input device was conducted by Bolt [7]. In a "Put-That-There" system, a user could manipulate objects on a large screen using voice and pointing gestures. In this system, the direction of pointing is measured by a magnetic field sensor attached to a user's wrist and finger.

Tsukada and Yasumura's Ubi-Finger [8], a finger-wearable input device, enables the user to operate home appliances using finger gestures such as pointing.

As these studies require users to attach or wear sensing devices, this can interfere with the performance of other

© Fuji Technology Press Ltd. Creative Commons CC BY-ND: This is an Open Access article distributed under the terms of the Creative Commons Attribution-NoDerivatives 4.0 International License (http://creativecommons.org/licenses/by-nd/4.0/).

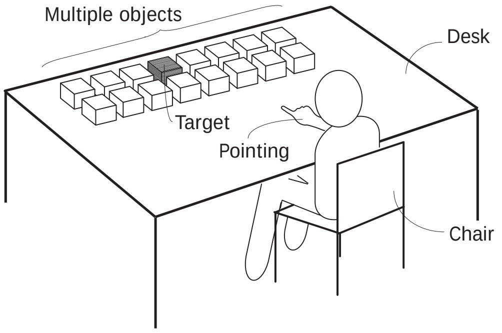


Fig. 1. Schematic view of the assumed environment.

tasks. To resolve such problems, many studies have been undertaken that deal with the recognition of pointing by using image processing.

Cipolla and Hollinghurst present a pointing-based interface for robot guidance [9]. In their study, they defined the direction of pointing as the direction of the index finger. Sato and Sakane, as well as Cipolla and Hollinghurst, presented a pointing-based interface system "Interactive Hand Pointer," which is used for instructing a robot manipulator to conduct a pick-and-place task [10]. The system relied on visual feedback to a user by projecting a mark at the indicated location in the real workspace.

Kahn et al. defined the direction of pointing as the direction from head to hand [11]. In their study, they assumed a virtual cone whose tip is at the hand and the indicated object is in the cone. This definition of the direction of pointing has been used in many other studies [12–14].

Fukumoto et al. assumed that the direction of pointing is determined by a straight line defined by a fingertip and a base point [15]. They reported that the base point differs for operators and for postures of the operator and therefore used a virtual base point calibrated beforehand.

Based on a similar idea, Mashita et al. proposed a pointing gesture model [16]. They reported that the cognitive origin, which is the same concept as the base point [15], lies in the reference plane and its coordinates are determined from the posture of the pointing arm.

Given the definitions of the direction of pointing, these studies are roughly classified into (1) finger direction [7–10], (2) head to hand [11–14], and (3) base point to fingertip [15, 16].

The finger direction approach is not directly applicable in a situation in which target candidates are too far from the user or candidates are too close together. In some studies [9, 10], visual feedback is used to solve this problem; however, visual feedback sometimes degrades pointing performance [17].

Compared to the finger direction approach, the head to hand approach more accurately determines the direction of pointing, but accuracy depends highly on positioning of user or target candidates.

The base point to fingertip approach requires that the position of the base point be known, but this differs with positions of target candidates or postures of user. Thus gesture recognition system must continuously recalibrate the base point and the direction of pointing is based on this calibration.

Regardless of the approach, errors inevitably occurs in pointing recognition. Even in inter-human communication, it is known to be difficult to understand precisely what another person is pointing to [18, 19]. Recognition error was about 8 cm [19]. Considering these facts, integrating the direction of pointing and other information such as verbal instructions [20] could considerably diminish error and increase target recognition.

The objective of this study is to propose a method that can be used to estimate the target from pointing in which multiple objects are close together. In such a situation, an accurate estimation is required.

Considering the related studies mentioned above, an accurate model for the direction of pointing is useful for this objective; however, such a model would be expensive and require a great deal of sensory information, such as the positions of the head, shoulder, elbow, wrist, finger, and eye gaze.

To limit the amount of sensory information required to identify a target, we integrated the following approaches:

1. Estimating the user's subjective pointing direction based on a linear model using finger direction.

2. Integrating sensory information from the user's pointing and contextual information such as action sequence.

3. Arranging target candidates as appropriate according to the user's characteristics.

Approaches 1 and 2 are passive in that the gesture recognition system receives information from the user and adapts internal parameters to the user. Approach 3 is active in that the system rearranges target candidates to redirect the user's pointing.

This paper is organized as follows: Section 2 gives an overview of how the user's subjective pointing direction is estimated. The integration of sensory and contextual information is proposed in Section 3. We describe the method for adaptive placement of the target candidates according to the user's characteristics in Section 4. The experiments for verifying these approaches are described in Section 5. In Section 6, we discuss the advantages and limitations of the method. In Section 7, we conclude the paper and refer to future research.

2. Estimation of the User's Subjective Pointing Direction

2.1. Recognition of Pointing

To recognize the position of a user's hands and fingers, we adopt the following recognition method by Oka et al. [4]. We extract a hand region using an infrared camera and binarize the input image with an appropriate threshold. We search for a fingertip based on its geometrical features using a circular template. The center of the hand

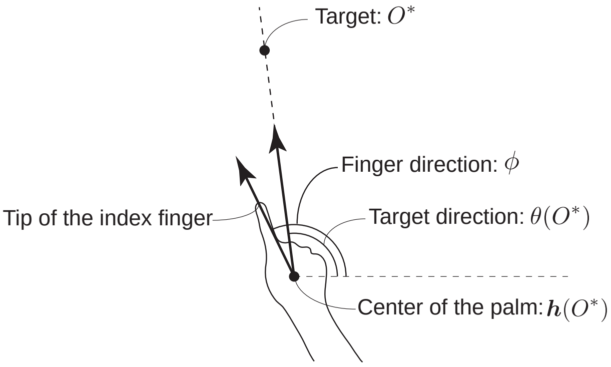

Target: O^*
Finger direction: ϕ
Tip of the index finger
Target direction: $\theta(O^*)$
Center of the palm: $\boldsymbol{h}(O^*)$

Fig. 2. Definition of finger and target direction.

is given as the point at which the distance to the closest region boundary is maximum. This can be obtained by repeatedly applying a morphological erosion operator until the region becomes smaller than a threshold, then calculating the resulting region's center of mass.

In this study, pointing is defined as an action meeting the following conditions:

- The center of the hand is outside a specific work space.
- Only the index finger of a user's dominant hand is observed.
- A user's hand remains almost stationary.

2.2. Model of the User's Subjective Pointing Direction

In order to estimate a user's subjective pointing direction, too much information is required as mentioned in the previous section. Making use of all the information is, however, unreasonable from the viewpoint of the computational cost.

In this study, we try to estimate a user's subjective pointing direction from the direction of the user's index finger, which is acquired with relative ease. Considering the situation in which a sitting user indicates an object, which is on a desk, the cost effectiveness of using three-dimensional data is thought to be low. In this paper, therefore, we use two-dimensional data acquired from an infrared camera from overhead.

As illustrated in **Fig. 2**, finger direction ϕ is defined as the direction from the center of the palm to the fingertip of the index finger, and target direction $\theta(O^*)$ is defined as the direction from the center of the palm to the center of target O^*.

Using coordinates with their origin at the center of the user's palm, we can ignore seating position of a user.

In this study, we assume that ϕ^{subj}, the user's subjective pointing direction, is approximated by a linear function as follows:

$$\phi^{\mathrm{subj}} = a\phi + b \quad \dots\dots\dots\dots\dots \quad (1)$$

where a and b are constants.

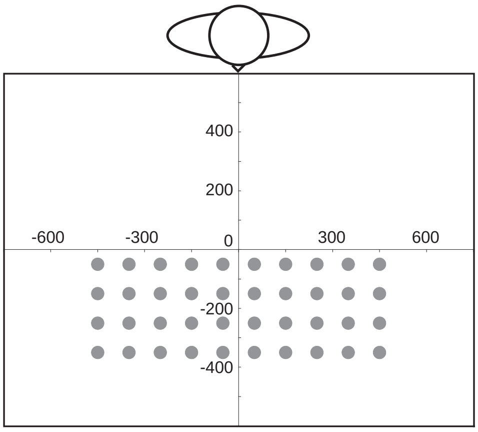


Fig. 3. Arrangement of projected markers.

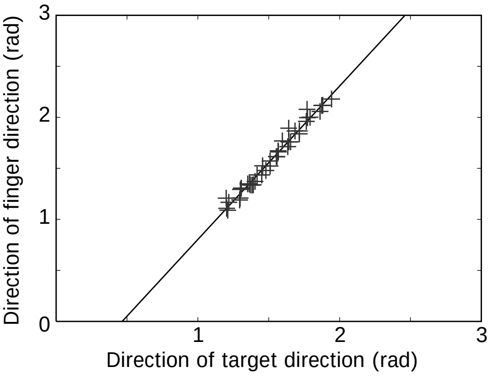


Fig. 4. An example of the relation between the direction of finger and target.

Ideally, the user's subjective pointing direction ϕ^{subj} agrees completely with target direction $\theta(O^*)$.

2.3. Validation of the Proposed Model of the User's Subjective Pointing Direction

To validate the linear model, we studied the relationship between finger direction and target direction via experiments.

Subjects were 7 men and 3 women ($n = 10$) from 21 to 30 years old who pointed at 40 blue markers (a circle with a radius of 20 mm) randomly and consecutively projected onto a desktop at a grid spacing of 100 mm by an LCD projector (**Fig. 3**).

The chair was on an extension of the desk centerline 300 mm away from the desk.

Figure 4 shows an example of the relationship between the direction of finger and target. **Table 1** gives determination coefficients for regression lines for each subject.

As the average of the determination coefficients of the regression lines is quite high at 0.940, the validity of the proposed model of a user's subjective pointing direction is determined. Based on the model, the average pointing direction error is 0.0613 rad.

The average regression line gradient is 1.40, exceeding 1.0, meaning that the user's subjective pointing direction

Table 1. Determination coefficients of regression lines of finger direction (ϕ) on the target direction ($\theta(O^*)$).

Subject	Regression line	Determination coefficient
A	$\phi = 1.504\theta - 0.701$	0.979
B	$\phi = 1.494\theta - 0.713$	0.899
C	$\phi = 1.567\theta - 0.793$	0.975
D	$\phi = 1.353\theta - 0.466$	0.955
E	$\phi = 1.321\theta - 0.334$	0.970
F	$\phi = 1.434\theta - 0.503$	0.974
G	$\phi = 1.412\theta - 0.483$	0.965
H	$\phi = 1.213\theta - 0.268$	0.952
I	$\phi = 1.571\theta - 0.789$	0.891
J	$\phi = 1.168\theta - 0.011$	0.842
Average		0.940

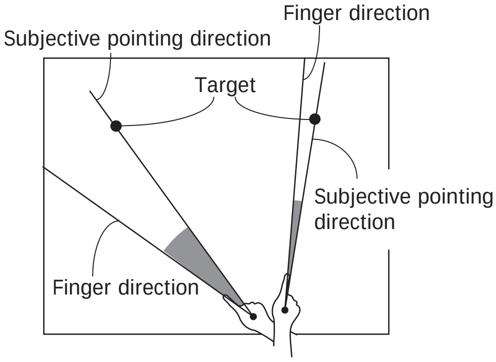


Fig. 5. Difference between the finger and subjective pointing direction.

does not correspond to the finger direction regardless of coordinates.

As illustrated in **Fig. 5**, the difference between the user's subjective pointing direction and the finger direction grows as the target moves to the left when the experimental subject is right-handed.

This is because the right-hander bends the elbow when pointing to an object on the left.

3. Integration of Sensory and Contextual Information

To integrate sensory information obtained from a user's pointing and contextual information from a user's action sequence, we apply a probabilistic model to the user's subjective pointing direction. In concrete terms, we assume that the user's subjective pointing direction ϕ^{subj} follows the wrapped normal distribution [21] with mean $\theta(O^*)$, the direction of target O^*.

We define the conditional probability density function for ϕ^{subj} given target $O^* = O^i$ as follows:

$$\begin{aligned}&\rho(\phi^{\mathrm{subj}}|O^*=O^i)\\&\quad=\frac{1}{\sqrt{2\pi}\sigma}\sum_{m=-\infty}^{\infty}\exp\left\{-\frac{\left(\phi^{\mathrm{subj}}-\theta(O^i)+2\pi m\right)^2}{2\sigma^2}\right\}\end{aligned} \quad \ldots\ldots\ldots\ldots (2)$$

where $\theta(O^i)$ is the direction of the i-th candidate of target O^i and σ^2 is the variance of the pointing direction.

Applying Bayes' rule, we obtain the likelihood that the target is O^i given the user's subjective pointing direction ϕ^{subj} as follows:

$$\rho(O^*=O^i|\phi^{\mathrm{subj}})=\frac{\rho(\phi^{\mathrm{subj}}|O^*=O^i)p(O^*=O^i)}{\rho(\phi^{\mathrm{subj}})} \quad (3)$$

where $p(O^*=O^i)$ is the prior probability in which the target is O^i and $\rho(\phi^{\mathrm{subj}})$ is the prior probability in which the user's subjective pointing direction is ϕ^{subj}. Here, $\rho(\phi^{\mathrm{subj}})$ is assumed to follow a uniform distribution.

In this study, each "action" corresponds to an object, so an "action sequence" becomes a target sequence.

In the target sequence, O_t^*, the target at time step t, depends only on O_{t-1}^*, so conditional probability $p(O_t^* = O^j|O_{t-1}^* = O^i)$ is defined for each combination of O^i and O^j as follows:

$$p(O_t^*=O^j|O_{t-1}^*=O^i)=\frac{N_{ij}+\beta_{ij}}{\sum_{k=1}^{n}(N_{ik}+\beta_{ik})} \quad \ldots (4)$$

where N_{ij} is the number of times that O^j is the target next to O^i, and β_{ij} represents an initial distribution of conditional probability. If there is prior knowledge of the relationship between target candidates, initial β_{ij} is adjusted to represent prior knowledge. Here, however, the initial β_{ij} are equal.

Conditional probabilities are described and retained as a conditional probability table (CPT):

$$\mathrm{CPT}_{O_t^*|O_{t-1}^*}=\begin{pmatrix}p(O^1|O^1) & \cdots & p(O^n|O^1)\\ \vdots & \ddots & \vdots\\ p(O^1|O^n) & \cdots & p(O^n|O^n)\end{pmatrix}. \quad (5)$$

We integrate sensory information from the user's subjective pointing directions and contextual information from the user's action sequences using Bayes' rule.

Given that the target sequence from step 1 to step t is $\mathcal{O}_1^t = \{O_1^*, O_2^*, \ldots, O_t^*\}$ and the sequence of the user's subjective pointing directions from step 1 to step t is $\Phi_1^t = \left\{\phi_1^{\mathrm{subj}}, \phi_2^{\mathrm{subj}}, \ldots, \phi_t^{\mathrm{subj}}\right\}$, $\rho(\mathcal{O}_1^t, \Phi_1^t)$, their joint probability distribution, is calculated as follows:

$$\begin{aligned}\rho(\mathcal{O}_1^t,\Phi_1^t)&=\prod_{\tau=2}^{t}p(O_\tau^*|O_{\tau-1}^*)\prod_{\tau=1}^{t}\rho(\phi_\tau^{\mathrm{subj}}|O_\tau^*)p(O_1^*)\\&=p(O_t^*|O_{t-1}^*)\rho(\phi_t^{\mathrm{subj}}|O_t^*)\rho(\mathcal{O}_1^{t-1},\Phi_1^{t-1})\end{aligned} \quad (6)$$

where $\phi_\tau^{\mathrm{subj}}$ is the user's subjective pointing direction at step τ.

The target at step t is estimated as follows:

$$\hat{O}_t^* = \operatorname*{argmax}_{O^i} p(O^* = O^i | O^*_{t-1}) \rho(\phi_t^{\mathrm{subj}} | O_t^* = O^i). \quad (7)$$

This model assumes that the target at step t can be estimated only from the target at step $t-1$ and the user's subjective pointing direction at step t.

4. Adaptive Placement of Target Candidates

Increasing the difference in directions between each pair of target candidates is useful in reducing the number of errors in target estimation.

From a human interface perspective, however, it is undesirable to drastically change the placement of target candidates just to increase this difference, so it is important to keep positioning of target candidates as close to the initial one as possible.

We placed target candidates meeting the requirements above using a force-directed approach, assuming that target candidates are square.

The force-directed approach was developed during studies of a graph-drawing problem [22, 23] in which a graph was modeled as a spring system whose spring constants are defined based on the relationship of node pairs and minimizing total system energy.

In the graph-drawing problem, node size and shape is ignored, making it difficult to apply to our problem. In the studies of Printed Circuit Board (PCB) floor planning, a system of springs whose nodes are rectangular was considered by Quinn et al. [24]. In their method, the procedure of placing modules has two phases. The first is for determining the relative placement, and the second is for removing the overlaps of modules. In the second phase, it is necessary to make drastic changes in the placement of the modules, which makes it impossible for this method to meet the requirements for our application.

In what follows, we propose a force-directed method that meets the requirements mentioned above.

4.1. Error Reduction in Target Estimation

Our system estimates the target using Eq. (7) based on the product of the following two elements:

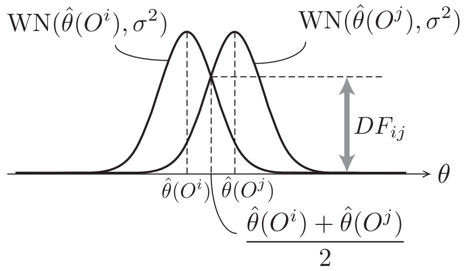
- $p(O_t^* = O^i | O^*_{t-1})$: probability based on a user's action sequence
- $\rho(\phi_t^{\mathrm{subj}} | O_t^* = O^i)$: probability based on the user's pointing direction

We define difficulties to distinguish between the two candidates as follows:

4.1.1. Difficulty from the User Action Sequences

Given that the target in the previous step is O^k, the difficulty in distinguishing between O^i and O^j derived from the user's action sequence is defined as follows:

$$DA_{ij}^{(k)} := \frac{p_{ki} + p_{kj}}{|p_{ki} - p_{kj}|} \quad \ldots\ldots\ldots\ldots (8)$$



$\mathrm{WN}(\hat{\theta}(O^i), \sigma^2)$
$\mathrm{WN}(\hat{\theta}(O^j), \sigma^2)$
DF_{ij}
θ
$\hat{\theta}(O^i)$
$\hat{\theta}(O^j)$
$\frac{\hat{\theta}(O^i) + \hat{\theta}(O^j)}{2}$

Fig. 6. Definition of difficulty derived from the user's pointing directions, DF_{ij}.

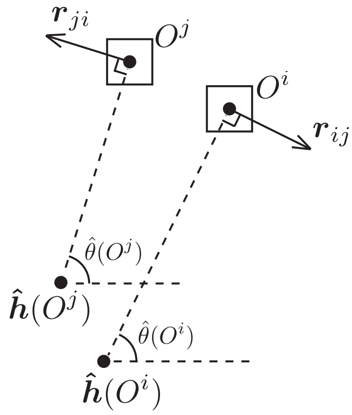


Fig. 7. Repulsion acting between O^i and O^j.

where $p_{ki} = p(O_t^* = O^i | O^*_{t-1} = O^k)$. The denominator of the right side represents the proximity of p_{ki} to p_{kj}, and the numerator means the possibility of being selected as the target.

4.1.2. Difficulty from the User's Pointing Directions

The difficulty in distinguishing between O^i and O^j derived from the user's pointing is defined as follows:

$$DF_{ij} := \frac{1}{\sqrt{2\pi}\sigma} \sum_{m=-\infty}^{\infty} \exp\left\{ -\frac{(\hat{\theta}(O^j) - \hat{\theta}(O^i) + 4\pi m)^2}{8\sigma^2} \right\} \quad \ldots\ldots\ldots\ldots (9)$$

where $\hat{\theta}(O^i)$ and $\hat{\theta}(O^j)$ represent estimated object directions of O^i and O^j based on past data for pointing.

DF_{ij} is defined as the probability density (2) under the condition that the user's subjective pointing direction is the intermediate between the directions of O^i and O^j (**Fig. 6**).

4.1.3. Repulsion Based on Difficulty

When the target of the previous step is O^k, the difficulty of distinguishing between O^i and O^j is calculated from Eqs. (8) and (9) as follows:

$$D_{ij}^{(k)} := DA_{ij}^{(k)} \cdot DF_{ij}. \quad \ldots\ldots\ldots\ldots (10)$$

We assume that repulsion acts between O^i and O^j (**Fig. 7**).

Repulsion is proportional to $D_{ij}^{(k)}$ and acts on O^i from O^j in direction $\boldsymbol{e}_{ij} = (e_{ij}^x, e_{ij}^y)$.

$$\begin{cases} e_{ij}^x = \cos\left(\hat{\theta}(O^i) + \dfrac{\pi}{2}\delta\right) \\ e_{ij}^y = \sin\left(\hat{\theta}(O^i) + \dfrac{\pi}{2}\delta\right) \end{cases} \quad \ldots\ldots. \quad (11)$$

where

$$\delta := \begin{cases} 1 & (\hat{\theta}(O^i) > \hat{\theta}(O^j)) \\ -1 & (\hat{\theta}(O^i) < \hat{\theta}(O^j)). \end{cases} \quad \ldots\ldots (12)$$

When $\hat{\theta}(O^i) = \hat{\theta}(O^j)$, δ is assigned 1 or -1 randomly.

Repulsion increases the difference between the directions of O^i and O^j.

From Eqs. (10), (11), and (12), $\boldsymbol{r}_i$, resulting repulsion acting on O^i is defined as follows:

$$\begin{aligned} \boldsymbol{r}_i &:= \sum_{j \neq i} \boldsymbol{r}_{ij} \\ &= \sum_{j \neq i} cD_{ij}^{(k)} \boldsymbol{e}_{ij}, \quad c = \text{const.} \end{aligned} \quad \ldots\ldots \quad (13)$$

where $\boldsymbol{r}_{ij}$ is repulsion acting on O^i from O^j.

4.2. Dependence on Initial Placement

To maintain the initial placement of target candidates as closely as possible, we assume the spring of natural length 0 between $\boldsymbol{v}_i = (v_i^x, v_i^y)$, the current position of O^i, and $\boldsymbol{v}_i^{\text{init}}$, the initial position of O^i. $\boldsymbol{v}_i^{\text{init}}$ is determined by a user.

According to Hooke's law, $\boldsymbol{s}_i$, restoring force acting on O^i, is directly proportional to its extension as follows:

$$\boldsymbol{s}_i = -k(\boldsymbol{v}_i - \boldsymbol{v}_i^{\text{init}}) \quad \ldots\ldots\ldots\ldots \quad (14)$$

where k is the spring constant.

4.3. Constraints

Since, in this study, the location in which the target candidates can be placed is restricted to the desktop, we define potential function $P(x,y)$ so that candidates do not protrude from the desk as follows:

$$\begin{aligned} P(x,y) := e^{(x-w/2+l)} + e^{(-x-w/2+l)} + e^{(y-h/2+l)} \\ + e^{(-y-h/2+l)} + P_0 \end{aligned} \quad \ldots\ldots\ldots \quad (15)$$

where the desk size is $w \times h$, l is the offset, and P_0 is the constant value meeting the condition that $P(0,0) = 0$ (**Fig. 8**).

Constraint force $\boldsymbol{d}_i = (d_i^x, d_i^y)$ acts on O^i as follows:

$$\begin{cases} d_i^x = -e^{(v_i^x - w/2+l)} + e^{(-v_i^x - w/2+l)} \\ d_i^y = -e^{(v_i^y - h/2+l)} + e^{(-v_i^y - h/2+l)}. \end{cases} \quad \ldots \quad (16)$$

4.4. Changes in Placement of Target Candidates

The result of all forces acting on O^i is calculated as the sum of Eqs. (13), (14), and (16) as follows:

$$\boldsymbol{f}_i = \boldsymbol{r}_i + \boldsymbol{s}_i + \boldsymbol{d}_i. \quad \ldots\ldots\ldots\ldots\ldots \quad (17)$$

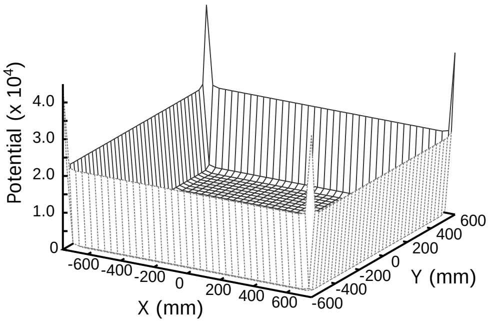
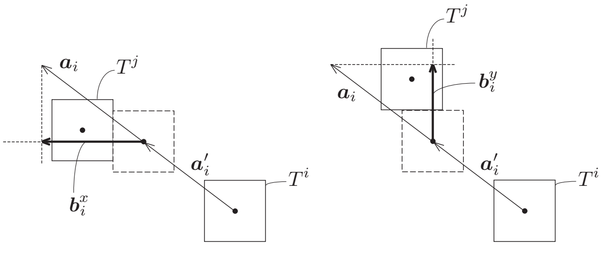


Fig. 8. Potential function derived from the desk size constraint (desk size: $w = 1500$ mm, $h = 1200$ mm).



(a) Collision in the X-axial direction

(b) Collision in the Y-axial direction

Fig. 9. Collision between O^i and O^j.

k/c is the ratio of spring constant (k) to the coefficient of repulsion (c). The greater the k/c, the smaller the movement. The placement of target candidates changes drastically as k/c approaches 0.

The placement of target candidates changes based on this model. O^i movement is based on movement vector $\boldsymbol{a}_i$ defined as follows:

$$\boldsymbol{a}_i = a\boldsymbol{f}_i, \quad a = \text{const.} \quad \ldots\ldots\ldots\ldots \quad (18)$$

O^i moves based on this movement vector $\boldsymbol{a}_i$ from its current position unless it collides with other candidates. Given that all target candidates move once in a cycle, the cycle is repeated until the movement of all candidates converges.

The procedure when O^i collides with O^j during O^i movement is shown below.

1. For each target candidate O^i:

   a. Move O^i until it touches O^j, and this movement vector is defined as $\boldsymbol{a}_i'$.

   b. Define $\boldsymbol{b}_i$ as $\boldsymbol{b}_i := \boldsymbol{a}_i - \boldsymbol{a}_i'$, and preserve the vertical component of it to the contact side as $\boldsymbol{b}_i^p$, where $p \in \{x,y\}$ (**Fig. 9**).

2. For each pair of O^i and O^j:

   a. If O^i touches O^j and $|\boldsymbol{b}_i^p - \boldsymbol{b}_j^p|$ exceeds the threshold, swap the positions of O^i for that of O^j.

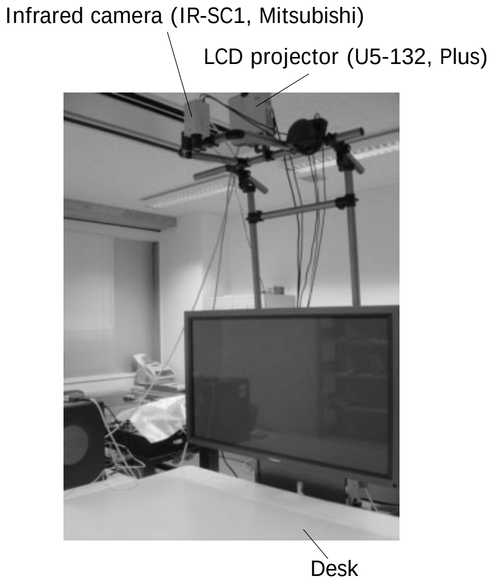

Infrared camera (IR-SC1, Mitsubishi)
LCD projector (U5-132, Plus)
Desk

Fig. 10. Overview of the experimental setup.

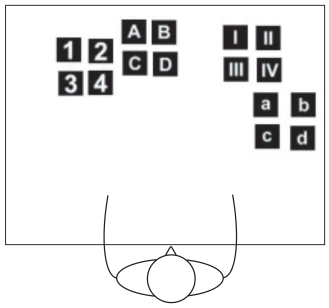


Fig. 11. Initial placement A.

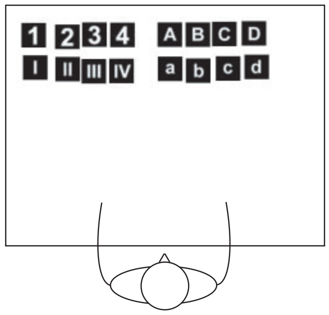


Fig. 12. Initial placement B.

5. Experiments

To verify the usefulness of our proposal, we conducted the experiments described in the sections that follow:

5.1. Experimental Setup

In our experimental setup (**Fig. 10**), the desk is 1000 × 800 mm. An infrared camera (IR-SC1, Mitsubishi) measures the user's palm and fingers, and an LCD projector (U5-132, Plus) projects 16 target candidates onto a desktop. Each target candidate is square, 80 mm on a side. For image processing, we used a Linux PC (Pentium 4, 2.8 GHz) with a fast image-processing system (IP7000BD, Hitachi).

Subject A (male, 24 years old), who took part in the experiment in Section 2.3, took part in this experiment.

5.2. Overview of Experiments

Initially arranging 16 target candidates at discretion, then he pointed at candidates. The candidates are classified into four groups:

- Arabic numerals (1 / 2 / 3 / 4)
- Roman numerals (I / II / III / IV)
- Uppercase letters (A / B / C / D)
- Lowercase letters (a / b / c / d)

These groups correspond one to one with four tasks, with the subject alternating between pointing from A to D, defined as a task, and pointing to A, defined as a subtask.

The subject repeated the four tasks in sets of 10 in random order.

This corresponds to the fact that, in daily life, the order of tasks is not usually fixed, but the order of subtasks within a task is typically fixed. For instance, a coffee drinker usually conducts subtasks in a specific order: making coffee, adding milk, adding sugar, and stirring them in.

The system adapts to the subject to estimate targets based on the procedure in Sections 2 and 3, and arranges target candidates according to the procedure mentioned in Section 4.

Offset in the constraint equation (16) is $l = 8.0$, and the coefficient of movement vector $\boldsymbol{a}_i$ is $a = 0.0001$. The ratio of k, the spring constant, to c, the coefficient of repulsion, is $k/c = 10$. The initial conditional probability table (5) is $\beta_{ij} = 1.0$.

5.3. Experimental Results

We compared the following three:

- **M1:** Use of sensory information alone
- **M2:** Integration of sensory and contextual information
- **M3:** Adaptive placement of target candidates with integrated sensory and contextual information (proposed)

Two types of initial placement, A (**Fig. 11**) and B (**Fig. 12**), are examined.

Figure 13 shows the change in average rate of correct estimation of targets.

When using sensory information alone (M1), the system learned only the characteristics of the user's pointing gesture as the subjective pointing direction, and the rate of correct estimations of targets was approximately 60%,

Table 2. Average ratio of correct estimations of targets in task switching.

Method	Placement A (%)	Placement B (%)	Average (%)
M1	55.0	61.3	58.1
M2	71.3	63.8	67.5
M3	85.0	80.0	82.5

Table 3. Breakdown of misestimates in task switching.

Cause	Ratio (%)
(i) Error in estimation of subjective pointing direction	17.9
(ii) Experimental setting	25.0
(iii) Tray arrangement	57.1

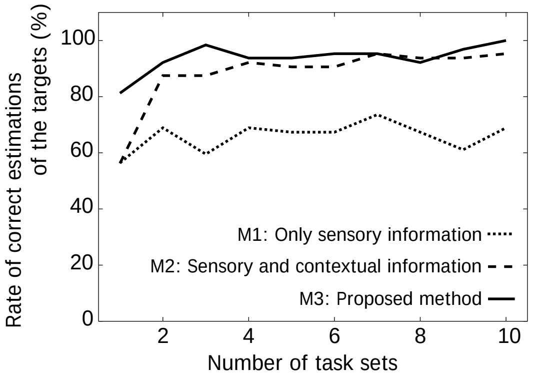


Fig. 13. Change in the rate of correct estimations of targets.

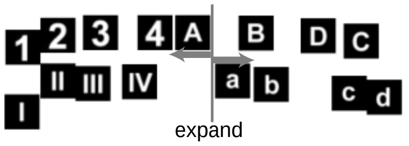


Fig. 14. An example of increasing the angle between two candidates.

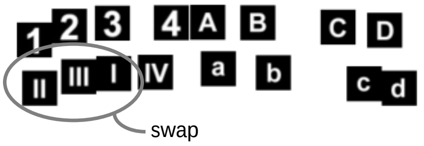


Fig. 15. An example of swapping two candidates.

a low rate due to the placement of target candidates when candidates were placed close to each other.

Compared to the result of M1, the rate of correct estimations of M2 rose to about 90% after the second step. Contextual information contributed to the high rate in each task, and sensory information contributed to it in task switching.

In the second set and after, both M2 and M3 achieved a high rate of correct estimations, and average rate of the correct estimations of M3 in the second set and after (95.3%) was slightly higher than that of M2 (91.8%).

In the first set, the rate of correct estimations of M2 was 56.3%, whereas that of the proposal (M3) was 81.3%. This difference was due to the fact that the system had not yet learned the user's action sequence, and the result of target estimation depends heavily on pointing.

When the system accumulates experience, whether the target candidates are placed adaptively or not, it rarely makes mistakes regarding the estimation in each task. Estimation faults mainly occur at the very moment a task is changed.

For instance, the system can see III as a successor to II with a high level of credibility, but cannot determine the successor to IV, which is the last subtask, with only action sequences. If the initial placement of the target candidates is similar to placement A (**Fig. 11**), the system distinguishes the next target from 1, I, A, and a without difficulty. When initial placement is similar to placement B (**Fig. 12**), however, the system cannot determine whether the subject points at A or a because the values of the posterior probability of A and a are close and the directions of A and a are also very close.

On the other hand, in the proposed method, the system increased the angle between the direction of A and that of a in the third set and after as illustrated in **Fig. 14**, so misestimates rarely occurred.

Furthermore, in initial placement B, the system sometimes swapped the positions of I, II, and III to decrease the difficulty of distinguishing among 1 and I in the fifth set and after as shown in **Fig. 15**.

Average rates of correct estimations of targets in task switching are shown in **Table 2**.

Note that the proposed method improves the rate of correct target estimations, especially in task switching. Nevertheless, misestimates occurred at 17.5% even when the proposed method (M3) was used.

To identify the causes of the misestimates, we examined the obtained data, such as positions of palm and index finger and those of target candidates at each misestimate. Our analysis indicated that the causes were classified in three (**Table 3**).

Twenty percent of the first type (i) is due to faults in the recognition of hands and fingers using the infrared camera. A user's subjective pointing direction is estimated from Eq. (1) using the least-squares method, given a set of 10 pointings. Therefore, when an outlier from fault recognition is detected, the outlier adversely affects the model of a user's subjective pointing direction. To minimize such influences, it may be necessary to increase the number of data used for the least squares.

Misestimates from the rest of (i) occurred at the first set. Because the system had not yet learned the individuality of the user's pointing, and the system could not obtain a good approximation of the subjective pointing direction. This problem can be resolved as learning progresses.

The second type of error (ii) is due to the particularity of the experiment. All misestimates are observed at the first step in each set. In this experiment, prior probabilities of all candidates are equal when each set is finished, so target estimation depends only on the user's subjective pointing direction at the first step in each set. This is rare in real deskwork.

The third type of error (iii) is derived from the layout method proposed in this paper. The specific reason for this is that the extent of the difference between each pair of trays is insufficient. Such misestimates could be lowered by reducing k/c to close to zero.

6. Discussion

Although human pointing depends on complex movement of the shoulder, elbow, and other body parts, as stated in Section 2, the relationship between finger direction and target direction is accurately approximated by a linear function. This result supports the idea that human pointing direction depends on the direction of the bodily member, such as a finger or arm, as well as on a higher level of information processing related to eye gaze and spatial cognition.

Assuming target candidates are placed in front of a user, the theoretically available number of target candidates is about 50 ($\pi/0.0613 = 51.25$). This estimation is based on the assumption that the system uses only sensory information. Therefore, the number may increase depending on the user's action sequences and arrangement of target candidates. On the other hand, assuming that the proposed method applies to the Attentive Workbench, the size and mechanism of the self-moving tray become bottlenecks. Realistically, the applicable number of target candidates is about 20.

Skilled people sometimes perform an *epistemic action* – a physical action to simplify internal problem solving [25, 26]. It can be said that adaptive placement of target candidates in this study is a kind of epistemic action. If the positioning of target candidates is fixed, the system has to make estimations only on the basis of the recognition ability of an infrared camera. The system simplifies the estimation of the target by changing the placement of the target candidates that are parts of the system itself.

In addition, adaptive placement can be seen as the sharing of a load. In other words, a user of this system bears the load as a result of reducing the load on the system by changing the placement of the candidates.

It is therefore important to consider the system advantages and disadvantages to the user when placement of target candidates is changed. If the target candidates move slightly and the phase relation of target candidates does not change, the load on the user is assumed to be slight. However, if the travel distance of the target candidates is long and the phase relation of the candidates changes dramatically, the load on the user can increase too much.

The load grows due to the placement of target candidates and differs between users, so it cannot be simply said that the smaller the k/c, the better the system performance. The system must determine k/c considering the balance of the load between the system and each user.

7. Conclusions

We have proposed a human-adaptive approach for estimating targets of human pointing. We integrated (1) estimating the user's subjective pointing direction, (2) integrating sensory and contextual information, and (3) arranging target candidates based on the user's characteristics. We demonstrated the usefulness of our proposal through target estimation experiments.

The primary challenge facing us is the load on the user. Placement of target candidates likely has some psychological and cognitive effects on the user. We must examine such effects through experiments.

We also plan to apply the proposed method to the Attentive Workbench (AWB), and to use the proposed method for instructing self-moving trays to deliver objects.

Acknowledgements

This work was supported in part by the Ministry of Education, Culture, Sports, Science and Technology, Japan, Grant-in-Aid for COE Research "Information Science and Technology Strategic Core," and Research Fellowships of the Japan Society for the Promotion of Science for Young Scientists (No.18·10474).

References:

[1] T. Sato, Y. Nishida, and H. Mizoguchi, "Robotic room: symbiosis with human through behavior media," Robotics and Autonomous Systems, Vol.18, pp. 185-194, 1996.

[2] R. A. Brooks, "The intelligent room project," Proc. of the 2nd Int. Cognitive Technology Conf., pp. 271-278, 1997.

[3] M. Sugi, Y. Tamura, J. Ota, T. Arai, K. Takamasu, K. Kotani, H. Suzuki, and Y. Sato, "Attentive Workbench: an intelligent production cell supporting human workers," 7th Int. Symposium on Distributed Autonomous Robotic Systems, pp. 441-450, 2004.

[4] K. Oka, Y. Sato, and H. Koike, "Real-time fingertip tracking and gesture recognition," IEEE Computer Graphics and Applications, Vol.22, No.6, pp. 64-71, 2002.

[5] B. A. Sawyer, "Magnetic positioning device," US patent, 3,457,482, 1969.

[6] K. Kotani, I. Hidaka, Y. Yamamoto, and S. Ozono, "Analysis of respiratory sinus arrhythmia with respect to respiratory phase," Method of Information in Medicine, Vol.39, pp. 153-156, 2000.

[7] R. A. Bolt, " "Put-that-there": voice and gesture at the graphics interface," Proc. of the 7th Annual Conf. on Computer Graphics and Interactive Techniques, pp. 262-270, 1980.

[8] K. Tsukada and M. Yasumura, "Ubi-Finger: gesture input device for mobile use," Proc. of the 5th Asia-Pacific Conf. on Computer-Human Interaction, Vol.1, pp. 388-400, 2002.

[9] R. Cipolla and N. J. Hollinghurst, "Human-robot interface by pointing with uncalibrated stereo vision," Image and Vision Computing, Vol.14, pp. 171-178, 1996.

[10] S. Sato and S. Sakane, "A human-robot interface using an interactive hand pointer that projects a mark in the real work space," Proc. of the 2000 IEEE Int. Conf. on Robotics and Automation, pp. 589-595, 2000.

[11] R. E. Kahn, M. J. Swain, P. N. Prokopowicz, and R. J. Firby, "Gesture recognition using the perseus architecture," Proc. of the 1996 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, pp. 734-741, 1996.

[12] M. S. Lee, D. Weinshall, E. Cohen-Solal, A. Colmenarez, and D. Lyons, "A computer vision system for on-screen item selection by finger pointing," Proc. of the 2001 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, Vol.1, pp. 1026-1033, 2001.

[13] M. Hild, M. Hashimoto, and K. Yoshida, "Object recognition via recognition of finger pointing actions," Proc. of the 12th Int. Conf. on Image Analysis and Processing, pp. 88-93, 2003.

[14] C. Colombo, A. D. Bimbo, and A. Valli, "Visual capture and understanding of hand pointing actions in a 3-D environment," IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics, Vol.33, No.4, pp. 677-686, 2003.

[15] M. Fukumoto, Y. Suenaga, and K. Mase, " "Finger-pointer": pointing interface by image processing," Computers & Graphics, Vol.18, No.5, pp. 633-642, 1994.

[16] T. Mashita, Y. Iwai, and M. Yachida, "Detecting indicated object from head-mouted omnidirectional images by pointing gesture," Proc. of SICE Annual Conf., pp. 3230-3235, 2003.

[17] B. A. Po, B. D. Fisher, and K. S. Booth, "Pointing and visual feedback for spatial interaction in large-screen display environments," Proc. of the 3rd Int. Symposium on Smart Graphics 2003, pp. 22-38, 2003.

[18] T. Miyasato and F. Kishino, "An Evaluation of Virtual Space Teleconferencing System Based on Detection of Objects Pointed through a Virtual Space," The IEICE Transactions on Information and Systems, Vol.J80-D-II, No.5, pp. 1221-1230, 1997 (in Japanese).

[19] T. Imai, D. Sekiguchi, N. Kawakami, and S. Tachi, "Measuring Accuracy of Nonverbal Information Perception of Humans: Measurement of Pointing Gesture Perception," Transactions of the Virtual Reality Society of Japan, Vol.9, No.1, pp. 89-96, 2004 (in Japanese).

[20] O. Sugiyama, T. Kanda, M. Imai, H. Ishiguro, and N. Hagita, "Three-layered draw-attention model for humanoid robots with gestures and verbal cues," Proc. of the 2005 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, pp. 2140-2145, 2005.

[21] E. Gumbel, J. Greenwood, and D. Durand, "The circular normal distribution: theory and tables," Journal of the American Statistical Association, Vol.48, No.261, pp. 131-152, 1953.

[22] T. Kamada and S. Kawai, "An algorithm for drawing general undirected graphs," Information Processing Letters, Vol.31, No.1, pp. 7-15, 1989.

[23] T. M. Fruchterman and E. M. Reingold, "Graph drawing by force-directed placement," Software – Practice and Experience, Vol.21, pp. 1129-1161, 1991.

[24] N. R. Quinn and M. A. Breuer, "A forced directed component placement procedure for printed circuit boards," IEEE Transactions on Circuits and Systems, Vol.26, No.6, pp. 377-388, 1979.

[25] D. Kirsh, "The intelligent use of space," Artificial Intelligence, Vol.73, pp. 31-68, 1995.

[26] P. Maglio and D. Kirsh, "Epistemic action increases with skill," Proc. of the 18th Annual Conf. of the Cognitive Science Society, pp. 391-396, 1996.

Name:
Yusuke Tamura

Affiliation:
Postdoctoral Researcher, Research into Artifact, Center for Engineering (RACE), The University of Tokyo

Address:
5-1-5 Kashiwanoha, Kashiwa-shi, Chiba 277-8568, Japan

Brief Biographical History:
2006-2008 JSPS Research Fellow (DC2)
2008 Received Ph.D. from School of Engineering, The University of Tokyo
2008- Postdoctoral Researcher, RACE, The University of Tokyo

Main Works:

- Y. Tamura, M. Sugi, J. Ota, and T. Arai, "Deskwork Support System Based on the Estimation of Human Intentions," Proc. IEEE Int. Workshop on Robot and Human Interactive Communication, pp. 413-418, 2004.
- Y. Tamura, M. Sugi, J. Ota, and T. Arai, "Estimation of User's Intention Inherent in the Movements of Hand and Eyes for the Deskwork Support System," Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, pp. 3709-3714, 2007.

Membership in Academic Societies:

- The Institute of Electrical and Electronics Engineers (IEEE)
- The Robotics Society of Japan (RSJ)

Name:
Masao Sugi

Affiliation:
Project Assistant Professor, IRT Research Initiative, The University of Tokyo

Address:
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

Brief Biographical History:
2003-2007 Project Research Associate, Graduate School of Information Science and Technology, The University of Tokyo
2007- Project Assistant Professor, IRT Research Initiative, The University of Tokyo

Main Works:

- M. Sugi, H. Yuasa, J. Ota, and T. Arai, "Autonomous Distributed Control of Traffic Signals with Cycle Length Control," Trans. of the Society of Instrument and Control Engineers, Vol.E-3, No.1, pp. 8-18, 2004.
- M. Sugi, Y. Maeda, Y. Aiyama, T. Harada, and T. Arai, "A Holonic Architecture for Easy Reconfiguration of Robotic Assembly Systems," IEEE Trans. on Robotics and Automation, Vol.19, No.3, pp. 457-464, 2003.
- M. Sugi, I. Matsumura, Y. Tamura, J. Ota, and T. Arai, "Quantitative Evaluation of Physical Assembly Support in Human Supporting Production System "Attentive Workbench"," Proc. 2008 IEEE Int. Conf. on Robotics and Automation, pp. 3624-3629, 2008.

Membership in Academic Societies:

- The Institute of Electrical and Electronics Engineers (IEEE)
- The Robotic Society of Japan (RSJ)
- The Society of Instrument and Control Engineers (SICE)
- The Japan Society for Precision Engineering (JSPE)

Name:
Tamio Arai

Affiliation:
Professor, Department of Precision Engineering, School of Engineering, The University of Tokyo

Address:
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

Brief Biographical History:

1977 Dr. Engineering, The University of Tokyo
1987-present Professor, Dept. of Precision Engineering, The University of Tokyo

Main Works:

- T. Arai, N. Yamanobe, Y. Maeda, H. Fujii, T. Kato, and T. Sato, "Increasing Efficiency of Force-Controlled Robotic Assembly Design of Damping Control Parameters Considering Cycle Time," Annals of the CIRP, Vol.55, No.1, pp. 7-10, 2006.
- Y. Maeda and T. Arai, "Planning of Graspless Manipulation by a Multifingered Robot Hand," Advanced Robotics, Vol.19, No.5, pp. 501-521, 2005.
- T. Yan, J. Ota, A. Nakamura, T. Arai, and N. Kuwahara, "Development of a Remote Fault Diagnosis System Applicable to Autonomous Mobile Robots," Advanced Robotics, Vol.16, No.7, pp. 573-594, 2002.
- T. Arai, Y. Aiyama, M. Sugi, and J. Ota, "Holonic Assembly System with Plug and Produce," Computer in Industry, Vol.46, pp. 289-299, 2001.

Membership in Academic Societies:

- The Japan Society of Precision Engineering (JSPE)
- The Robotic Society of Japan (RSJ)
- Intelligent Autonomous System Society (IAS)
- The Int. Academy for Production Engineering (CIRP)
- The Institute of Electrical and Electronics Engineers (IEEE)

Name:
Jun Ota

Affiliation:
Associate Professor, Department of Precision Engineering, Graduate School of Engineering, The University of Tokyo

Address:
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

Brief Biographical History:

1989-1991 with Nippon Steel Corporation, Kanagawa, Japan
1991- with the University of Tokyo
1996-1997 Visiting scholar at Stanford University, Stanford, CA, USA

Main Works:

- N. Fujii and J. Ota, "Rearrangement Task by Multiple Mobile Robots with Efficient Calculation of Task Constraints," Advanced Robotics, Vol.22, No.2-3, pp. 191-213, 2008.
- S. Hoshino, J. Ota, A. Shinozaki, and H. Hashimoto, "Improved design methodology for an existing automated transportation system with automated guided vehicles in a seaport container terminal," Advanced Robotics, Vol.21, No.3-4, pp. 371-394, 2007.
- J. Ota, "Multi-agent Robot Systems as Distributed Autonomous Systems," Advanced Engineering Informatics, Vol.20, No.1, pp. 59-70, 2006.

Membership in Academic Societies:

- The Robotics Society of Japan (RSJ)
- The Society of Instrument and Control Engineers (SICE)
- The Japan Society for Precision Engineers (JSPE)
- The Japan Society of Mechanical Engineers (JSME)
- The Institute of Electrical and Electronics Engineers (IEEE)