

Research Article

Estimation Algorithm of Machine Operational Intention by Bayes Filtering with Self-Organizing Map

Satoshi Suzuki¹ and Fumio Harashima²

¹ School of Science and Technology for Future Life, Department of Robotics and Mechatronics, Tokyo Denki University, 2-2 Kanda-Nishiki-cho, Chiyoda-ku, Tokyo 101-8457, Japan

² Tokyo Metropolitan University, 1-1 Minami-Osawa, Hachioji-Shi, Tokyo 192-0397, Japan

Correspondence should be addressed to Satoshi Suzuki, ssuzuki@fr.dendai.ac.jp

Received 6 June 2011; Revised 31 August 2011; Accepted 9 September 2011

Academic Editor: Holger Kenn

Copyright © 2012 S. Suzuki and F. Harashima. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We present an intention estimator algorithm that can deal with dynamic change of the environment in a man-machine system and will be able to be utilized for an autarkical human-assisting system. In the algorithm, state transition relation of intentions is formed using a self-organizing map (SOM) from the measured data of the operation and environmental variables with the reference intention sequence. The operational intention modes are identified by stochastic computation using a Bayesian particle filter with the trained SOM. This method enables to omit the troublesome process to specify types of information which should be used to build the estimator. Applying the proposed method to the remote operation task, the estimator's behavior was analyzed, the pros and cons of the method were investigated, and ways for the improvement were discussed. As a result, it was confirmed that the estimator can identify the intention modes at 44–94 percent concordance ratios against normal intention modes whose periods can be found by about 70 percent of members of human analysts. On the other hand, it was found that human analysts' discrimination which was used as canonical data for validation differed depending on difference of intention modes. Specifically, an investigation of intentions pattern discriminated by eight analysts showed that the estimator could not identify the same modes that human analysts could not discriminate. And, in the analysis of the multiple different intentions, it was found that the estimator could identify the same type of intention modes to human-discriminated ones as well as 62–73 percent when the first and second dominant intention modes were considered.

1. Introduction

Estimation of human intention is quite practical for various applications such as assistance software [1], prediction of users' requests on the internet [2], and marketing [3]. In the robotics fields, realization of the intention estimator is desired especially for power assist systems [4] and for cooperative robots [5] since users' intentions are significant for their control. A function to estimate users' internal status and their intention is embedded in social robots [6] or interactive human-friendly robots [7, 8]. Thus, such technology is becoming a requisite technology to realize advanced human-computer interaction. Interaction techniques used in their applications are, however, designed on a case-by-case basis since practical design methodology for a general artificial

system has not yet been established. Especially, function of existing intention estimators were determined at the design phase against a limited circumstance of encounter and target user group; hence, there is no design method for an intention estimator which can deal with dynamic change of user's environment and of user's individual characteristics such as skill and experience. In some cases, we user needs familiarization with the use of the estimator's function, for instance we implicitly select adequate candidates of keywords when we use some data-search system. Due to this issue, there are many gadgets requiring user efforts in the case of not only intention estimators but also of other machines although a machine has been developed to support humans basically (remember how long and hard training to get a car license was in order to drive a vehicle that was

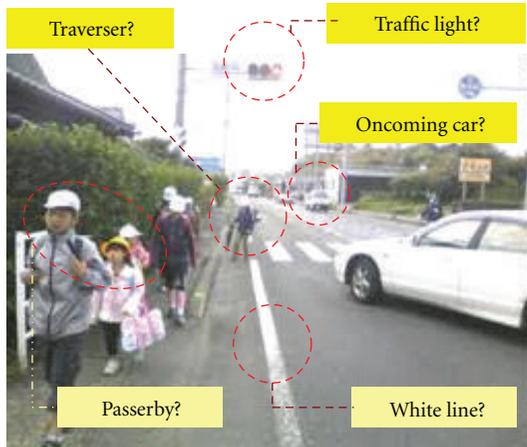


FIGURE 1: Example of difficult situation to specify environmental factor to estimate intentions: on the crossroad.

developed to make human mobility convenient). Against this background, new need of mechatronics supporting operational skill of user, that is a human adaptive mechatronics (HAM), was pointed out [9, 10]. Even to realize HAM, a function of an intention-estimation is desired to enhance the human-computer interaction. Specifications of the intention-estimation for HAM are the following.

- (S1) Adaptivity to natural circumstance which is always varying.
- (S2) Real time processing to cope with events in real world.
- (S3) Ease of designing not depending on ability of application developers.

Below, reasons of these specifications will be explained with discussion on research background and existing methods.

1.1. Existing Methods and the Issues

1.1.1. Difficulty to Specify Adequate Environmental Status. Estimation of intentions at general human-machine system is more difficult than above-mentioned successful examples since it is difficult to identify the information types which are utilized for the user's decision making. Further, it is hard for a machine to measure all types of environmental information that was perceived by a human. And, the machine recognition of events' causality is also hard since many factors and events affect each other. We can understand these difficulties easily when considering a driving situation at a crossroad, as shown in Figure 1. Since a driver has to perceive and recognize many factors—traffic light, oncoming cars, traversers, passersby, and their white line on the road—his/her attention to them should be always changing. Therefore, it is difficult for a machine to detect what the driver is paying attention to. For all these reasons, adequate selection of these environmental elements/status for an intention estimator requires trial and errors and an experience by the developers. If this troublesome work, that

is, selection of environmental variables, is automatically performed, the above-mentioned specifications (S1) and (S3) will be achieved.

1.1.2. Real time Processing. Estimation of timings of events/factors is significant for an operation assistance. Similarly, the intention-estimation has to be executed in real time. However, a sensitive human model that can estimate operator's action/intention under time-varying circumstance is not established yet. Generally, a discrete event modeling is proper to describe human cognitive behavior/action, and several effective methods such as GOMS [11] and Therblig [12] were developed. It is, however, unexpectedly difficult to embed time factors into a frame of these discrete models. As attempts to solve this issue, many researchers proposed a wide variety of hybrid systems. As an example, the stochastic switched ARX model was presented [13]. The model consists of continuous-time linear subsystems and the switching probabilities of the subsystems, are trained using an EM algorithm. The other is a hybrid system consisting of a finite state automaton and multiple linear dynamical systems [14]. Although effectiveness of these methods was demonstrated, the applicable conditions are limited, that is, the number of the discrete modes were relatively small. Because the number of network links increases drastically as the nodes increase, and then computation also increases. The countermeasure is to decrease the links by considering the context of events since the context of action plays a significant role in the decision of users' intentions. For this reason, event-driven models based on graph theories, such as hidden Markov model [15] and fuzzy automata [16], are used frequently. Building of these graph networks needs, however, experiences; hence, this issue comes down to the issue of the developer side as pointed out in (S3).

Moreover, as a research on behavior analysis of humans, activity recognition (AR) is known [17]. AR aims to recognize the actions from a series of observations like a lifelog. AR was performed by classifying measured data of human behavior, and the change/pattern of the data is analyzed mainly. On the other hand, intentions of action are so-called internal status of the human, and the intention does not always appear to the outside as behavior. Hence, methods for AR cannot be applied to the intention-estimation problem for human-support machines.

1.1.3. Usability without Dependency on Developer's Skill. Performance of algorithm depends on ability of the system developer. This fact is not unique to a design of an intention estimator; however, there are several awkward processes even to the skilled developer. For instance, if the aforementioned graph theory approaches are adopted, the system designer has to make the network structures by sufficiently considering relations of transition states and of events causality, that is, the developer has to know sufficiently the task contents. When we use "traditional metrics, an application developer would get conflicting information from the frame and the event level analysis [18]," in short, an experience is demanded to resolve it.

In summary, a design of intention-estimation using AR, Therblig, or graph theories entails human judgement and experience of developers; hence, these are not feasible for an automatic supporting system that has to execute computation without interposition of human developers online.

1.1.4. Merit and Demerit of Related Mathematical Tools. Since the estimation of intentions basically can be traced to a classification problem depending on environment and events, development of various clustering methods such as a support-vector machine and a k-nearest neighbor method makes it possible to realize applications of the intention-estimation. To resolve issues mentioned in Section 1.1.1, if many sensors were used to measure the environmental status, effectiveness of the above-mentioned liner classifiers are reduced due to the increase of the number of variables [19] and the intention estimator by these techniques would not work well. On the other hand, since other intention estimators based on graph theories assume Gaussian distribution for their probability computation, these methods cannot deal with transition of human intention changing under non-Gaussian probability. To resolve such issues, Bayesian approaches are well leveraged in recent years [19]. Bayesian techniques are utilized in wide areas, especially for the information technology field [20]. Since the Bayes approach is faithful to measured data, it is expected that intentions estimation is possible regardless of designers' ability and preoccupation. Although Bayesian method is a powerful tool, huge amount of computation is a similar problem if the number of elements increases.

1.2. Countermeasure and the Approach. From the discussion above, an intention-estimation algorithm that satisfies specifications (S1) and (S3) requires the following functions: a component analysis function which enables to extract dominant factors from multivariable information and a clustering function which can identify information relating to intentions. Therefore, the present authors focused on SOM that can perform the compression and identification of information. SOM is a type of artificial neural network to discriminate multidimensional data [21], and it models the cerebral cortex in the human brain. SOM is adequate for data processing of large dimensional data since the SOM technique can compress multidimensional information into lower (two-) dimensional map by keeping original topological information. Similar types of nodes gather close to each other on the map, and different types are assigned there apart from each other. The Bayesian probability embedded in the SOM is expected to be utilized to predict transition of status. Utilizing such property of SOM, the present authors presented automatic method to identify machine operation [22]. By expanding the method, a basic idea of *SOM-Bayes intention estimator* was also proposed [23, 24]. The merit given by SOM is labor saving for the developer; it is not necessary to specify the type of information utilized in the operator because selection of the informations is performed automatically through the SOM computation. This merits are helpful as solutions for specifications (S1) and (S3).

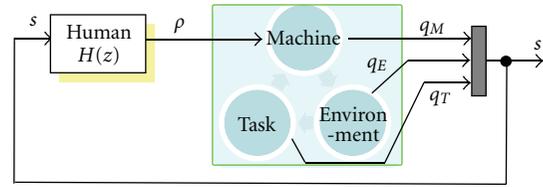


FIGURE 2: System architecture of human-machine system.

1.3. Purpose. In this paper, the details of algorithm of the SOM-Bayes intention estimator presented in [23, 24] are explained, and characteristics, design method, issues, and benefits are discussed comprehensively. Moreover, additional analyses of individual differences in human discrimination about intention modes and of multiple different intention are reported.

An organization of this paper is as follows: Section 2 explains an idea of the SOM-Bayes intention estimator and the algorithm. In Section 3, a particle filtering algorithm to realize the intention estimator is presented. In Section 4, a remote operation experiment system to apply the estimator and preparations to use it are explained. Section 5 shows analyses of the applied example and discusses the results. An improvement of the estimator and the related analysis are mentioned there. Section 6 presents several analyses concerning the estimator's behavior. Lastly Section 7 contains a conclusion and discussion.

2. Algorithm of SOM-Bayes Intention Estimator

As a preparation to design the intention estimator, a human-machine system structure is considered, and an internal status of the human model is assumed as intentions. Then, the design issues are formulated as an observer design problem that estimates the internal status. Elements in the human-machine system are a human (operator), a machine (to be manipulated), an environment, and the work task, as shown in Figure 2. This scheme is interpreted as the following: *human* operates *machine*, motion of the *machine* affects to *environment*, status of *task* is changed by the *environment*, and the change affects the status of *machine* again [25]. Describing these three types of statuses as the machine status (*M*-status, q_M), the environment status (*E*-status, q_E), and the task substance (*T*-status, q_T), a human during machine operation can be defined as an information-processing system, $\rho = H(z) \cdot s$, where $s := [q_M^T, q_E^T, q_T^T]^T$ is an information to be recognized by a human, ρ is an output of operation commands to the machine, and z is intentions of the human. Since it is impossible to describe the function H by using algebraic equations, H is required to be expressed using some mathematical model that can give numerical solution. Therefore, H is approximated by a mapping relation corresponding to transition property of H as an alternative way. Under this framework, the present authors devised a computation of belief $\text{bel}(z)$ of intentions z from measurable data s and ρ using Bayesian estimation method in order to obtain numerical solution

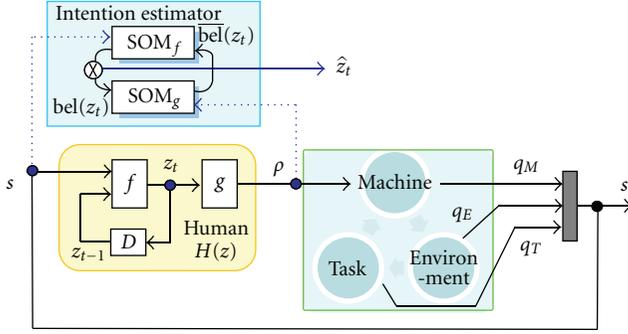


FIGURE 3: Block diagram involving a human model and the intention estimator in human-machine system.

from the function H . This is a computation method to infer probability of reasons using information of measured events. Probability of events is predicted based on the prior distribution, and the prediction is modified using the postmeasurement probability. Repeating this process, an internal status of the dynamical system, that is, intentions in this case, is estimated. Applying the Bayesian estimation approach to the computation of H , the computation of $H(z) \cdot s \rightarrow \rho$ is formulated as follows:

$$z_t = f(z_{t-1}, s_t), \quad (1)$$

$$\rho_t = g(z_t), \quad (2)$$

where the subscript $t \in \mathcal{N}$ is a time counter, the function f is a state transition function describing change of the intention z , and the function g corresponds to selection of operational commands based on the intention z . Based on a concept of “spotlight of selective attention” in global workspace theory (GWT) [26], so-called *spotlight models*, mathematical expression z is defined as a vector $z \in [0, 1]^{n_z}$ of which element corresponds to one intention strength of one operation action (n_z is a size of vector z). According to the *spotlight models*, several types of consciousness exist simultaneously inside human brain, and one of them floats from unconsciousness level as a conscious awareness. Since there are several operational modes in case of a general machine operation, the vector-form expression of intentions can fit a concept of *spotlight models*.

The probabilistic distribution $\text{bel}(z)$ is estimated using a technique of the Bayes filtering with (1) and (2). Although other intentions not involving the machine operation exist inside the human brain simultaneously, only intentions determining the operation are considered since the information which relates to the operation is solely treated; hence, such internal status is called simply as “intention.”

Basic algorithm for estimation of $\text{bel}(z)$ using Bayes filtering is explained below.

Algorithm 1 (Bayes filter). Consider the following:

$$\overline{\text{bel}}(z_t) = \int p(z_t | s_t, z_{t-1}) \cdot \text{bel}(z_{t-1}) dz_{t-1}, \quad (3)$$

$$\text{bel}(z_t) = \eta \cdot p(\rho_t | z_t) \cdot \overline{\text{bel}}(z_t). \quad (4)$$

This algorithm is defined with iterative equations that are computed from time (index) $t = 1$ to the final time T . $p(z_t | s_t, z_{t-1})$ corresponds to a probabilistic distribution of a transition of intention from z_{t-1} to z_t given input s_t . $p(\rho_t | z_t)$ is the conditional probabilistic distribution of judgment that outputs ρ_t if the intention z_t happens to be true. Equation (3) is a *prediction* to obtain a belief $\overline{\text{bel}}(z_t)$ at the time of t . Equation (4) is called *measurement update* and adjusts the prediction $\overline{\text{bel}}(z_t)$ by considering the probability $p(\rho_t | z_t)$. Via this update, a new belief $\text{bel}(z_t)$ at time t is obtained. η is a so-called Bayes normalization constant. In the proposed approach, mapping relations shown below are acquired by utilizing a mapping relation of the SOM:

$$z_t \leftarrow \text{SOM}_f(z_{t-1}, s_t), \quad (5)$$

$$\rho_t \leftarrow \text{SOM}_g(z_t), \quad (6)$$

where arrows in the above equations represent a mapping relation that gives variables written in the left side using SOM_* by using variables written inside parentheses. The mapping relation of SOM_f and SOM_g are acquired offline at the training phase, and these are used as static mapping functions at the computation in (5) and (6). In short, SOM_f for the state transition function f is trained using the input vector sequence including the training time series data of $\{s'\}$ and $\{z'\}$, and then the SOM reference vectors ${}^f\xi_i$ ($i = 1, \dots, L_f$) are obtained, where L_f is the number of all nodes in SOM_f . Similarly, training the SOM_g for the measurement function g by using sequences $\{\rho'\}$, and $\{z'\}$, the other SOM reference vectors ${}^g\xi_i$ ($i = 1, \dots, L_g$) are obtained, where L_g is the number of all nodes in SOM_g . Here, $\{s'\}$ and $\{\rho'\}$ are made from the experimental logging data, and $\{z'\}$ is prepared by an analyst watching the record video of the expert’s operation. Details of preparation of $\{s'\}$, $\{\rho'\}$, and $\{z'\}$ will be explained in Section 4.3.

Figure 3 shows a block diagram describing an operator model and the intention estimator. As shown at the upper left area in the figure, prediction $\overline{\text{bel}}(z)$ is computed through SOM_f using input s , the prediction is updated through SOM_g by referring ρ , and $\text{bel}(z)$ was obtained. Finally estimated intention \hat{z} is derived from $\text{bel}(z)$. Details of the computation concerning (5) and (6) will be described at Algorithm 2 in Section 3.

3. Implementation by Particle Filter Algorithm

Bayesian computation described by (3) and (4) is implemented using the particle filtering technique. Since particle filtering expresses any shape of probabilistic distribution using multiple particles, it is widely used as a general probabilistic computation tools which can deal with non Gaussian distribution, for instance in robotics field [27]. Particle filtering is also for solution of issues discussed in Section 1.1.4. First of all, conceptual diagram of the SOM-Bayes filtering using particle filtering technique is illustrated in Figure 4. The illustration explains that this algorithm begins at the left block of the figure ((I) Prediction phase) by substituting input vectors, and that the process is succeeded

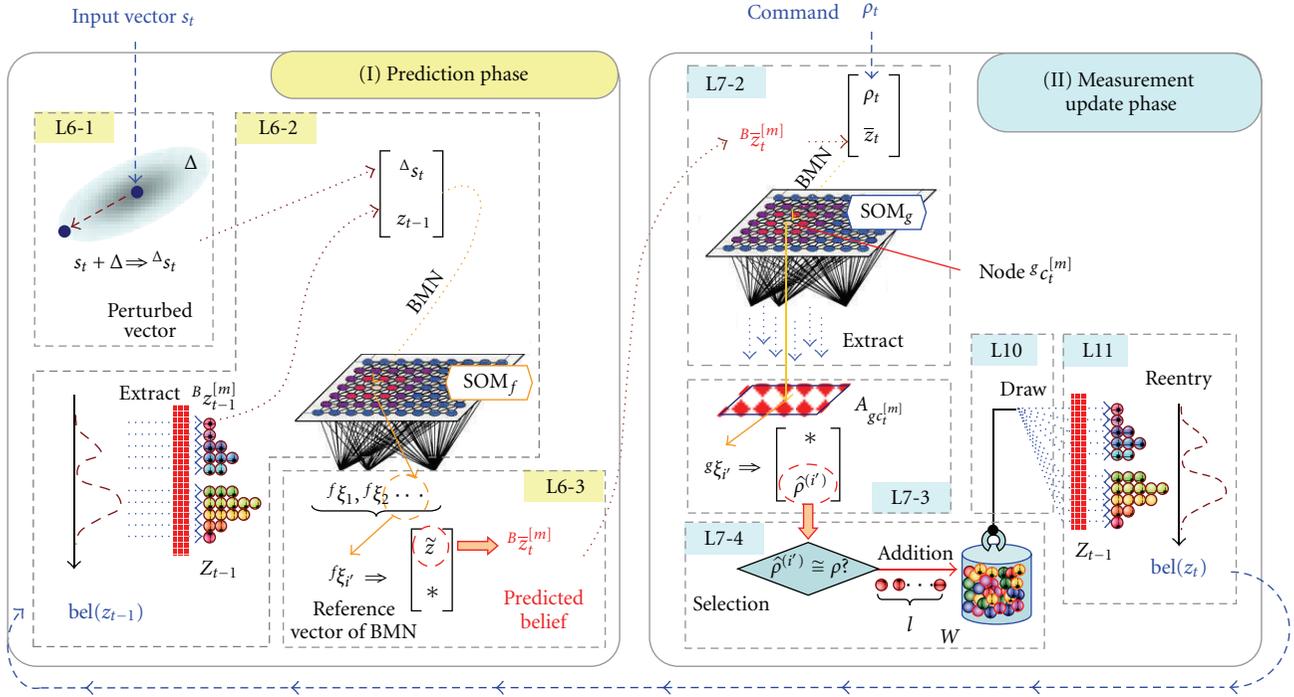


FIGURE 4: Algorithm of the SOM-Bayes estimator by particle filtering.

as “L6-1 \rightarrow L6-2 \rightarrow \dots ” and predicted belief is computed. In the right block of “(II) Measurement update phase”, $\text{bel}(z_t)$ is computed from the predicted belief, and $\text{bel}(z_t)$ is returned to the phase I at the next sampling time. The details of the processing is explained below by referring to the illustration shown in Figure 4. Labels like L6-1 in the figure indicate the corresponding line in latter Algorithm 2; for instance, L6-1 means a substep 1 of the sixth line of the pseudocode.

Assuming the number of particles is M , the m ($= 1, \dots, M$)th particle of the belief $\text{bel}(z_t)$ at time t is described as $B_{z_t}^{[m]}$. As a preparation to generate their particles, standard derivations σ_i ($i = 1, \dots, n_s$) of the sequence data $\{s\} \in R^{n_s \times T}$ in each element are computed, where n_s is the size of vector s . SOMs (which are the reference vectors $f\xi, g\xi$) defined by (5) and (6) are prepared at the SOM training step before execution of Algorithm 2.

Computation steps of the SOM particle filtering algorithm are shown by the following pseudocode. The algorithm consists of two phases: phase I for predictive computation and phase II for measurement updating. Below, notation Z_t is used to express a set of particles as $Z_t := \{B_{z_t}^{[1]}, B_{z_t}^{[2]}, \dots, B_{z_t}^{[M]}\}$.

Algorithm 2 (SOM particle filter).

- (1) initialize Z_1 , prepare $\{s\}$ and $\{\rho\}$
 - (2) for $t = 2$ to T do
 - (3) $Z_t = \emptyset$
- phase I
- (4) for $m = 1$ to M do

- (5) extract $B_{z_{t-1}}^{[m]}$ from Z_{t-1}
 - (6) generate $B_{z_t}^{[m]} \sim p(z_t | s_t, B_{z_{t-1}}^{[m]})$ by SOM_f
 - (7) $w_t^{[m]} := p(\rho_t | B_{z_t}^{[m]})$ using SOM_g
 - (8) endfor m
- phase II
- (9) for $i = 1$ to M do
 - (10) draw m' with probability $\propto w_t^{[m]}$
 - (11) add $B_{z_t}^{[m']}$ to Z_t
 - (12) endfor i
 - (13) endfor t .

First of the phase I is preparation of a temporal input vector which is perturbed from actual input data s_t by random value. Second, a best-matching-node (BMN) that is closest to a combination of the temporal input Δs_t and old particle $B_{z_{t-1}}^{[m]}$ is searched from SOM_f . And the predicted particle $B_{z_t}^{[m]}$ at time t is extracted from component of the reference vector of the found BMN (this computation corresponds to line 6). This perturbation technique is popular as ordinary particle filtering approach. Measurement probability $w_t^{[m]}$ corresponding to these predicted particles is computed using SOM_g for next resampling process on after-mentioned phase II (line 7). At the phase II, a particle number m' is chosen in proportion to the measurement probability $w_t^{[m]}$ of each particle (line 10), and the predicted $B_{z_t}^{[m']}$ which is indicated by the chosen number m' is picked up into a new set Z_t as next time particles (line 11). Then, new M particles are resampled

according to the measurement probability $w_t^{[m]}$. Repeating phase I and phase II till the final time T , $\{\text{bel}(z_t)\}$, that is, a time series $\{Z_t\}$ of the belief of the intention z , are obtained. Sixth line of the pseudocode in Algorithm 1 corresponds to prediction defined by (3). Seventh line is a probability computation of second term in RHS of (4), that is, $p(\rho_t | z_t)$. Tenth and eleventh lines correspond to a probabilistic selection of same second term in RHS of (4), and the repeated computation formed by ninth and twelfth lines plays a role of Bayes normalization described at the first term in (4). The details of main parts in Algorithm 2 are explained below.

Note that pseudocode written in Algorithm 2 does not always correspond to the following computation step by step since the aim of description of Algorithm 2 is to explain semantic principle of the Bayes filtering.

Line 6: Prediction.

L6-1: Preparation of Perturbation Input. A random sample point, Δs_t , that obeys a standard deviation σ_i around s_t is computed for all M particles. Specifically, using a random value Δ , the sample point is computed as

$$\begin{aligned} \Delta s_t^{[m]} &= s_t + \Delta, \quad m = 1, \dots, M \\ \Delta &:= \text{rand}([-1, 1], \sigma), \end{aligned} \quad (7)$$

where $\text{rand}([-1, 1], \sigma)$ is a function that yields pseudorandom value in the range of $[-1, 1]$ under standard deviation σ and it is computed by a method presented in [28].

L6-2: Search of Most Likelihood Node. Using an old particle $B_{z_{t-1}}^{[m]}$, a perturbed input vector $\Delta s_t^{[m]}$, and reference vectors $^f \xi$ of the SOM $_f$, the BMN of a particle m (i.e., $^f c_t^{[m]} \in \{1, \dots, L_f\}$) is searched as

$$^f c_t^{[m]} := \arg \min_i \left\{ \left\| \begin{bmatrix} \Delta s_t^{[m]} \\ B_{z_{t-1}}^{[m]} \end{bmatrix} - ^f \xi_i \right\| \right\}. \quad (8)$$

L6-3: Extraction of Prediction State. A candidate of an intention involved in the particle that corresponds to prediction $\overline{\text{bel}}(z_t)$ is extracted from a reference vector of the BMN predicted at the previous L6-2 step:

$$B_{z_t}^{[m]} \Leftarrow ^f \xi_{i(n_s + 1 : n_s + n_z)}, \quad i = ^f c_t^{[m]}, \quad (9)$$

where an operation described by parentheses in the above RHS indicates an extraction of elements of the vector components.

Line 7: Computation of Measurement Probability. A BMN that is closest to combination of the measured command ρ and the predicted belief $B_{z_t}^{[m]}$ is searched newly on the SOM $_g$. Then, reference vectors belonging to certain area around the new BMN are investigated. The number of nodes whose reference vectors correspond to the measured command ρ_t appears to be proportional to the postmeasurement probability; hence, the measurement probability is computed

from the number of such nodes. For the resampling process at after-mentioned phase II, an information of such nodes is registered into a roulette array W . Numbering of particles is recorded in the array, and the number of the numbering is determined in proportion to the amount of the corresponding nodes. The following are the details.

L7-1: Initialization. Reset the roulette array as $W = \emptyset$.

L7-2: Search of Most Likely Node. Using the reference vectors $^g \xi$ of SOM $_g$, a node $^g c_t^{[m]}$ that is most close to the measured ρ_t and predicted status $B_{z_t}^{[m]}$ is found by

$$^g c_t^{[m]} := \arg \min_i \left\{ \left\| \begin{bmatrix} \rho_t \\ B_{z_t}^{[m]} \end{bmatrix} - ^g \xi_i \right\| \right\}. \quad (10)$$

L7-3: Investigation of Area Around the Most Likely Node.

Computing a coordinate value (u_c, v_c) of the node $^g c_t^{[m]}$ on the SOM $_g$ plane map, reference vectors $^g \xi_{i'}$ of nodes that locate inside a square-like area $A_{g c_t^{[m]}}$ are investigated, where the length of side and the center of $A_{g c_t^{[m]}}$ are $(2L_r + 1)$ and (u_c, v_c) , respectively. Extracting from the vector $\xi_{i'}$ an element that corresponds to operation command, its element is described as $\hat{\rho}^{(i')}$, that is,

$$\hat{\rho}^{(i')} \Leftarrow ^g \xi_{i'}(1 : n_\rho), \quad \forall i' \in A_{g c_t^{[m]}}. \quad (11)$$

Here, L_r is a parameter and n_ρ is a size of vector ρ .

L7-4: Registration to the Roulette Array. The number of $\hat{\rho}^{(i')} (\in \mathcal{R}, \sum i' \leq (2L_r + 1)^2)$ which can be rounded to an integer of $\rho_t (\in I)$ is counted, and the number is described as l . Next, "the particle number, m " is registered into the array W l times additionally,

$$W \Leftarrow W + \{m\}^l. \quad (12)$$

Line 10: Draw. Since the numbering of particle that holds higher measurement probability has been registered in W more times, such particles are resampled again with high ratio in a random drawing. Hence, generating a random integer r within a range of $\{1, \dots, n_W\}$, a number m' that was registered on the r th element of the array W is drawn, where n_W is a length of W ,

$$m' = W(r), \quad r = \text{rand}(\{1, \dots, n_W\}). \quad (13)$$

Line 11: Reentry. m' th particle is re-registered as one of new particles for next step as

$$B_{z_t}^{[m]} \Leftarrow B_{z_t}^{[m']}. \quad (14)$$

Algorithm 2 yields M particles, that is, $\{B_{z_t}^{[1]}, \dots, B_{z_t}^{[M]}\}$, every iteration time. Since the belief is expressed by a distribution of those M particles, an estimated intention, say \hat{z} , is represented by averaging these M particles as

$$\hat{z}_t = \frac{1}{M} \sum_{m=1}^M B_{z_t}^{[m]}. \quad (15)$$



FIGURE 5: (a) Radio-controlled construction equipment on the work area and (b) the operation console.

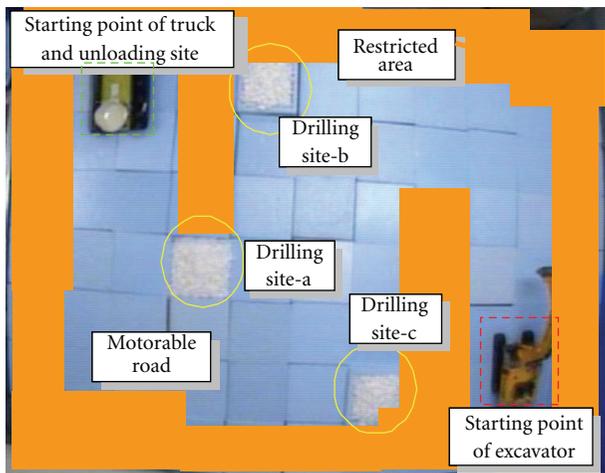


FIGURE 6: Overview of the work area.

4. Application to the Remote Operation Task

4.1. Experimental Setup. An experimental system with radio-controlled model construction equipments [29] was utilized to verify an effectiveness of the proposed intention-estimation algorithm. The purpose of the operation is a basic soil excavation work, as shown in Figure 5(a). Wireless cameras on an excavator and a truck captured video images, and displayed them on monitors for the operator, as shown in Figure 5(b). Both excavator and truck had crawler transporter systems. A bucket arm that consisted of a three-link mechanism is mounted on the superstructure of an excavator, and the mechanism are manipulated using a console system which is similar to the JIS- (Japanese Industrial Standards-) type cross-lever system. Figure 6 shows a top view of the work area. The field size is $3.3\text{ m} \times 2.4\text{ m}$ and

consists of a motorable road, restricted areas, three drilling sites, and one unloading site. Different sample pieces are put at three drilling sites. The excavator and truck were put at their starting points at the beginning of trial.

One operator manipulated both the excavator and the truck on his/her judgement and was required to perform and optimise task scheduling. The requirements are as follows.

- (i) Digging different sample pieces from three drilling sites.
- (ii) Digging operation at each site is permitted only once.
- (iii) Only one type of sample was permitted to be loaded into the truck at a time.
- (iv) Shorten the total trial time.
- (v) Collect as many samples as possible.

Standard task procedure is as follows: move to the drilling site, collect sample pieces with the excavator, load the pieces on the truck bed, and carry them to the unload site by the truck. The task procedure, however, can be chosen freely since the operator has many choices in sequence to visit three sites and in layout of positioning for digging and loading. Hence, the operator pays efforts to master the machine operation and to optimise whole task scheduling by trial and error considering the above-mentioned requirements.

To measure the positions of the excavator and truck for the intention estimator, the two equipments were observed by a camera attached to the ceiling through an infrared filter. The excavator and truck carried three and two infrared LED markers, respectively, and their positions and directions were computed using the detected positions of the LEDs. An angle of the excavator's boom and a rotation of the superstructure were measured by potentiometers. The measured signals were transmitted by wireless and recorded. Data acquisition was performed by the LabVIEW measurement system that consisted of the image acquisition, timing I/O (for digital

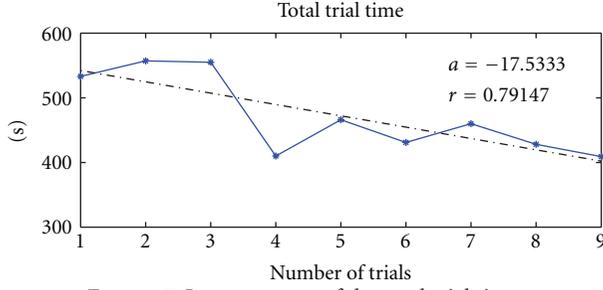


FIGURE 7: Improvement of the total trial time.

input data), and multifunction (for analog signal) modules. After capturing the video image from the ceiling camera, the images of LED markers were extracted through the video processing, and their coordinate values were obtained by centroid computation. Operations of switches on the console were also recorded via the multifunction and the I/O modules in the Labview system. The sampling frequencies of the video images and analog signals were 30 fps and 1 kHz, respectively.

4.2. Experimental Result. Written consent and ethical approval of one participant aged 21 yrs were obtained before experiment. As a training, three trials a day were conducted to the participant for three days; hence, total nine trials were repeated to improve the participant's operational performance. Figure 7 shows the improvement of total trial time. The gradient coefficient and correlation factor of the regression line for the total time are -17.5 and 0.79 , respectively. Since the correlation factor is sufficiently large and the tendency of monotonic decrease was confirmed, it can be thought that the participant improved the skill best at the last ninth trial. Therefore, operation data and recorded movie at the ninth trial were used for the later analysis and construction of the SOM Bayes estimator.

4.3. Preparation for the SOM-Bayes Intention Estimator. The first preparation is to convert experimental logging data into time series sequences of s' and ρ' . Crawler velocities for the excavator and the truck were controlled by two sliders with hands. The velocity commands were converted into the crawler operation mode κ_c by checking the velocities of both sides of crawlers [29].

$$\kappa_c = \begin{cases} 0: & \text{stop} \\ 1: & \text{forward} \\ 2: & \text{left-forward-turn} \\ 3: & \text{left pinwheel} \\ 4: & \text{right-backward-turn} \\ 5: & \text{backward} \\ 6: & \text{left-backward-turn} \\ 7: & \text{right pinwheel} \\ 8: & \text{right-forward-turn.} \end{cases} \quad (16)$$

TABLE 1: Classified operational intention modes, variables, and labels.

Variable	Label	Meanings
$(1)z \dots (3)z$	T/A-*	Truck's approach to the drilling site*
$(4)z \dots (6)z$	E/A-*	Excavator's approach to the drilling site*
$(7)z \dots (9)z$	T/P-*	Truck's positioning around the drilling site*
$(10)z \dots (12)z$	E/P-*	Excavator's positioning around the drilling site*
$(13)z$	E/D	Excavator's digging
$(14)z$	E/L	Excavator's loading of the payload
$(15)z$	T/TU	Transport of the payload by track and unloading

(*a, b, c: identifier of the drilling site).

The commands for the bucket and the superstructure consist of three modes: the superstructure rotation ${}^e\kappa_r$, the arm ${}^e\kappa_a$, and the bucket ${}^e\kappa_b$. The operation modes were determined as follows:

$$\begin{aligned} {}^e\kappa_r &:= \{0 : \text{stop}, 1 : \text{left rotation}, 2 : \text{right rotation}\}, \\ {}^e\kappa_a &:= \{0 : \text{stop}, 1 : \text{arm bend}, 2 : \text{arm stretch}\}, \quad (17) \\ {}^e\kappa_b &:= \{0 : \text{stop}, 1 : \text{boom up}, 2 : \text{boom down}\}. \end{aligned}$$

Difference of these operation groups appeared empirically to be utilized for inference of the operational intentions; hence, an operation command ρ is defined by a vector as follows:

$$\rho := \left[{}^t\kappa_c \quad {}^e\kappa_c \quad {}^e\kappa_h \right]^T \in I^3. \quad (18)$$

Vectors of the machine and environmental status (i.e., q_M and q_E) were chosen by considering position, posture, and geographical relation of the drilling sites and the equipments. The task status q_T was defined as $q_T := \{0: \text{no payload}, 1: \text{payload on bucket}, 2: \text{payload on truck bed}\}$ using the payload status. Refer to [30] for details about selection of these status. Time series sequence $\{s\}$ were obtained by combining q_M , q_E , and q_T , and the size of vector s became as $n_s = 28$.

The second preparation is to obtain the normative intention sequence z' discriminated by human analysts. Types of the remote operation for construction equipments and the definitions of elements $(1)z, \dots, (15)z$ in intention vectors z are summarized in Table 1. The operation modes are classified into three groups: approaching $(1)z, \dots, (6)z$, positioning $(7)z, \dots, (12)z$, and special operations (digging, loading, and transporting; $(13)z, \dots, (15)z$). Sequence of reference intentions $\{z'\}$ was made through video analysis by a participant analyst who did not know the remote-operation experiment at all in order to get rid of any preconceived ideas.



FIGURE 8: Sample footage for discrimination analysis: (a) view of the work area, (b) view of console desk.

Procedure to obtain the human analyst's intention z' is as follows.

- (1) The purpose of task, motion of equipments, geographical relationship of the work area, and the console layout are explained to the participant (below, analyst).
- (2) The analyst understands the intention modes summarized in Table 1.
- (3) The analyst discerns the operator's intention mode by watching the video. At that time, the analyst checks scene per frame using video editing software.
- (4) Type of mode and the frame index of the found intention are recorded.
- (5) After the analyst has finished decoding, the found frame indices are converted into time scale. Then, the sequence $\{z'\}$ is made by putting "1" to the corresponding element z'_i at the found timing.

Footage of the work area and operator from ceiling cameras were recorded into one screen image using a multi-viewer device (Figure 8 shows a part of the images), and the analyst discriminated the operator's intention by checking the motion of equipments and body action of the operator. Discrimination data by multiple analysts that will be mentioned later in Section 6.4 was obtained similarly.

The third preparation is a training of SOM to obtain the reference vectors. The training was performed using SOM_PAK [31]. Input vectors for SOM_f and SOM_g are $\{[z'^T, s'^T]^T\}$ and $\{[z'^T, \rho'^T]^T\}$, respectively. Each component of the input vectors for the training was normalized into the $[-1, 1]$ range by using the maximum and minimum values of the time sequence of the input vectors. Since it is preferable for the horizontal and vertical sizes of the rectangular map of SOM to be chosen in proportion to the ratio of two square roots of the first and second maximum eigenvalues of the covariance matrix of input vector sequence [32], the sizes of the SOM lattice ($u_{\text{dim}}, v_{\text{dim}}$) were decided as (50, 25) and (80, 20) for SOM_f and SOM_g , respectively. Hence, the number of nodes L_f and L_g are 1250 and 1600, respectively. A bubble-type neighborhood kernel function was chosen

for updating the reference vectors at the training. On the learning process, a fine-tuned computation was performed after the rough-tuned one was computed. The learning rate and learning length were specified as 0.05, 2000 and 0.02, 1.5 million, respectively, so as to meet such requirement that the learning length is more than 500-times the number of nodes [21].

The last fourth preparation is for the particle filtering algorithm. Initial state of particles for the Bayes filtering were specified as $Bz_t^{[m]}|_{t=1} := [0 \ 0 \ 0 \ 1 \ 0 \ \dots \ 0]^T$ for all $m (= 1, \dots, M)$ since it was obvious that an initial operation in the experiment was E/A-a (i.e., ${}^{(4)}z = 1$). The number of particles was decided as $M = 1000$. L_r for the area $A_{g_c^{[m]}}$ was specified as 11.

5. Results of Estimation

5.1. Preliminary Verification. For later discussion, the intention discerned by a human analyst is described by \tilde{z} , and the estimated intention computed by the SOM-Bayes estimator is described as \hat{z} . The result of estimation is shown in Figure 9. Lines show the transitions of intention modes from T/A-a (${}^{(1)}z$) to T/TU (${}^{(15)}z$). The same intention modes of \tilde{z} and \hat{z} are plotted by the blue and red lines at the same vertical position, respectively. Figure 9 indicates that \hat{z} overlaps with \tilde{z} well concerning T/A-b (${}^{(2)}z$) - E/A-c (${}^{(6)}z$) as labeled with *Good*. There are, however, several insufficient results: *Weak-estimation* (E/D, E/L), *Delayed-detection* (T/P-c, E/P-c), *Much long duration* (T/TU, T/P-b), and estimation; *Failure* (T/A-a, T/P-a, E/P-a, a part of E/D and E/L; Similar labels are written in the figure). For quantitative assessment of \tilde{z} and \hat{z} , the following concordance ratios were computed,

$${}^{(k)}R_{\text{tim}} := \frac{\sum_{\text{all } t'} {}^{(k)}\tilde{z}}{\sum_{\text{all } t'} {}^{(k)}\hat{z}} \quad t' = \{t \mid {}^{(k)}\hat{z}_t > 0.3\}, \quad (19)$$

$${}^{(k)}R_{\text{str}} := \frac{\sum_{\text{all } t'} {}^{(k)}\hat{z}}{\sum_{\text{all } t'} {}^{(k)}\tilde{z}} \quad t' = \{t \mid {}^{(k)}\tilde{z}_t = 1\}, \quad (20)$$

where $k = 1, \dots, 15$, and the constant 0.3 used in above equation is a threshold parameter (this value was decided

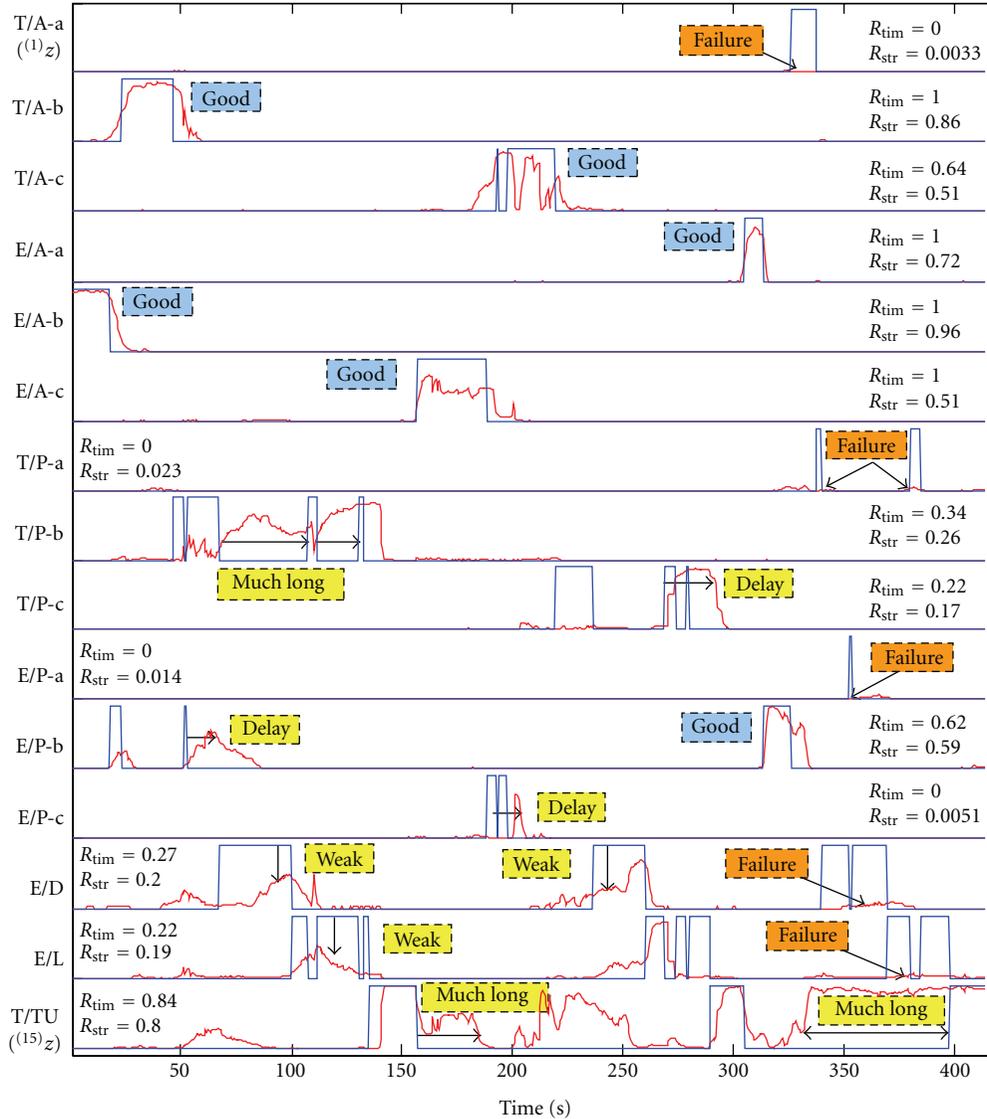


FIGURE 9: Transitions of intentions: red lines were identified by the SOM-Bayes estimator using vector p . Blue lines were discerned by human analyst.

subjectively. However, there is no problem since these ratios are used for relative comparison among $(^1)z \dots (^{15})z$. R_{tim} is a ratio of time that the analyst's \hat{z} matches to the time period of \hat{z} which is identified by the estimator. R_{str} shows how strongly the estimator can identify the intention level while analyst found same type of intention. In the latter discussion, these ratios are called “time-concordance ratio” and “strength-concordance ratio,” respectively. Values of these ratios are also written in each graph as shown in Figure 9. Average of R_{tim} and R_{str} for five *Good* modes (T/A-b - E/A-c) are 0.93 (range: 0.64–1) and 0.71 (range: 0.51–0.96), respectively; hence, high concordance was confirmed since these values are close to 1. On the other hand, other insufficient eight modes except T/TU (*Much long* duration) and E/P-b (a part of this mode is *Good*) show the small values as 0.13 (range: 0–0.34) and 0.11 (range: 0–0.26) concerning R_{tim} and R_{str} , respectively.

To clear the reason of the insufficient estimation, the trained SOM was investigated. Investigating elements corresponding to intention vectors in the reference vector (ξ) attached to each node on the SOM plane map, the type of intention mode included in each reference vector is discriminated. Then clusters in the SOM map are visualized as a colored map according to the discriminated modes on nodes. Figures 10 and 11 show the colored maps of SOM_f and SOM_g , respectively. Labels described in Table 1 are written in each cluster according to the discriminated modes. For multiple clusters indicating the same mode, subindex was written as “E/L-1, E/L-2”, and the dotted lines were drawn between the same clusters on the map. Figure 10 for the SOM_f map shows that obvious clusters are found; however, there is cluttered area in the left side on the SOM_g map as shown in Figure 11. This fact indicates that the classification in SOM_g was not performed

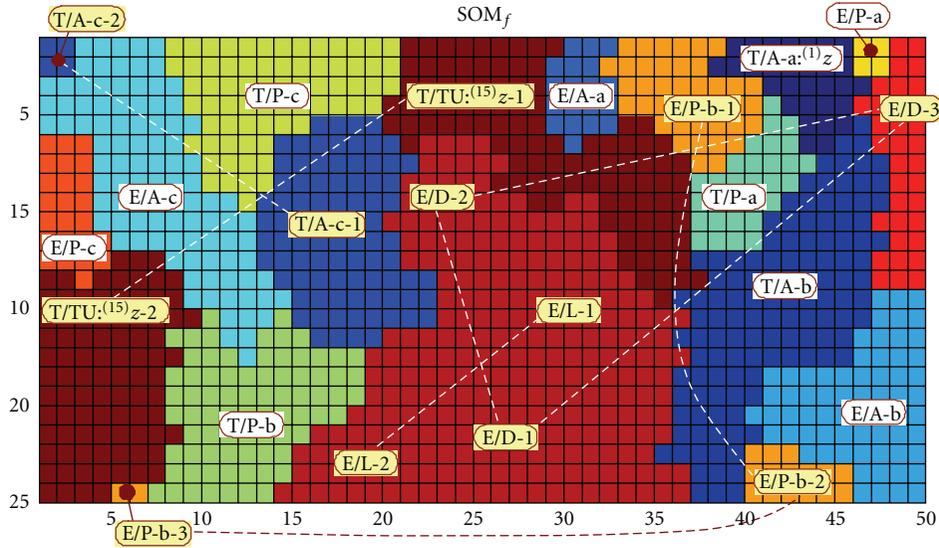


FIGURE 10: Clusters on the SOM_f map.

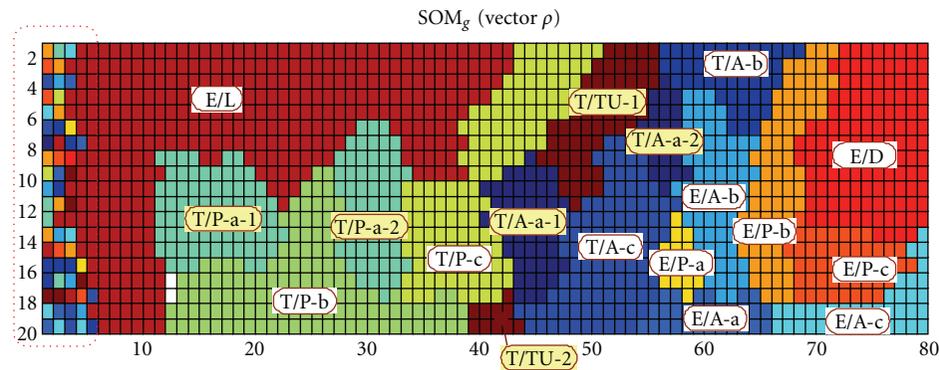


FIGURE 11: Clusters on the SOM_g map (case of vector ρ).

sufficiently. Due to the inadequate clustering, change of probabilistic distribution using SOM_g was discontinuous, and it is inferred that computation of approximation for the belief did not work well.

5.2. *Improvement of the SOM Mapping.* One technique to obtain adequate SOM clustering is to prepare adequate input data having sufficient but nonredundant information. From this viewpoint, investigating the input data that was used for the SOM training, it was found that the operator did not manipulate different types of commands defined by (18) simultaneously; hence, ρ was redefined as a scalar value by

$$\rho := 1 + {}^t\kappa_c + ({}^e\kappa_c \wedge 1)({}^e\kappa_c + 8) + ({}^e\kappa_h \wedge 1)({}^e\kappa_h + 16). \quad (21)$$

Figure 12 shows new SOM_g map obtained using the scalar ρ newly defined by (21). Unlike a former SOM generated by vector ρ , clear clusters without cluttered area are confirmed all over the map. To confirm this improvement, changes of time-concordance ratio R_{tim} of the result obtained using new SOM_g and of the former result by old SOM_g are compared, and the result was visualized as shown in Figure 13. The x ,

y , and z axes are the number of trials (the total is nine), intention modes (the total is fifteen), and the values of concordance ratios, respectively. Comparing the right 3D bar chart with the left one, most bars drawn in the right chart are higher than the others drawn in the left. Similar tendency of improvement was confirmed in case of the strength-concordance ratios R_{str} as shown in Figure 14. These graphs prove that accuracy of the intention-estimation was improved by reduction of dimension of the input data for the SOM learning.

6. Analyses of the Estimator's Behavior

The SOM-Bayes intention estimator using the improved SOM mapping are investigated by comparing with the human-discerned intentions \tilde{z} . Figure 15 shows the transitions of \tilde{z} and improved \hat{z} in the same manner as Figure 9. When compared with the former result in case of the vector- ρ (that are shown by red lines in Figure 9), the estimator using the scalar- ρ (red lines in Figure 15) improved in eight

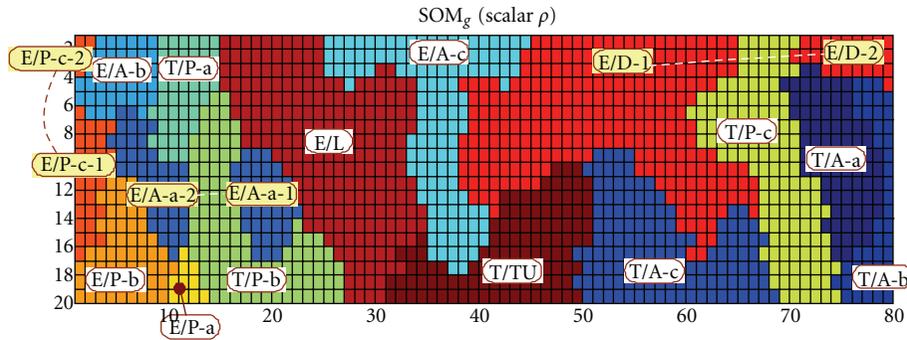


FIGURE 12: Clusters on the new SOM_g map (case of scalar ρ).

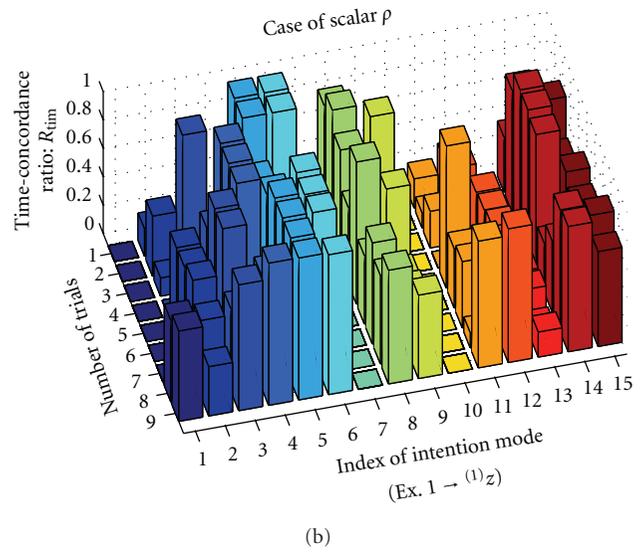
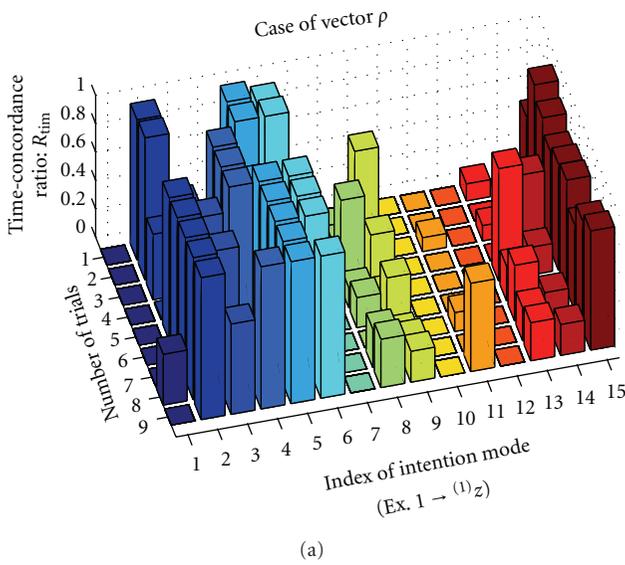


FIGURE 13: Comparison of time-concordance ratios R_{tim} in two cases: results using vector ρ (a) and scalar ρ (b).

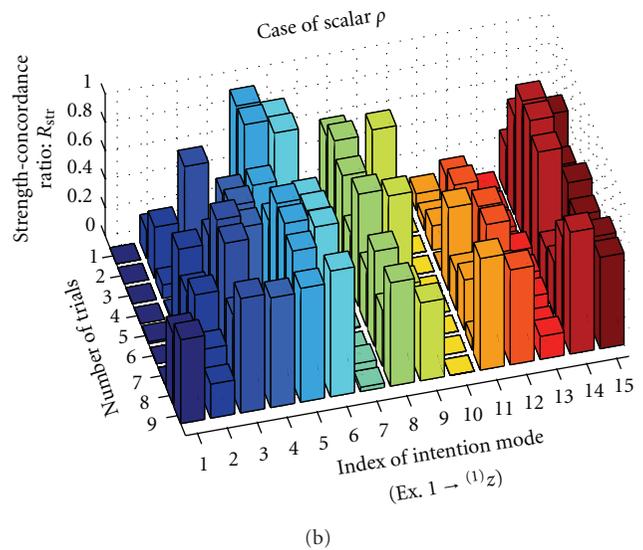
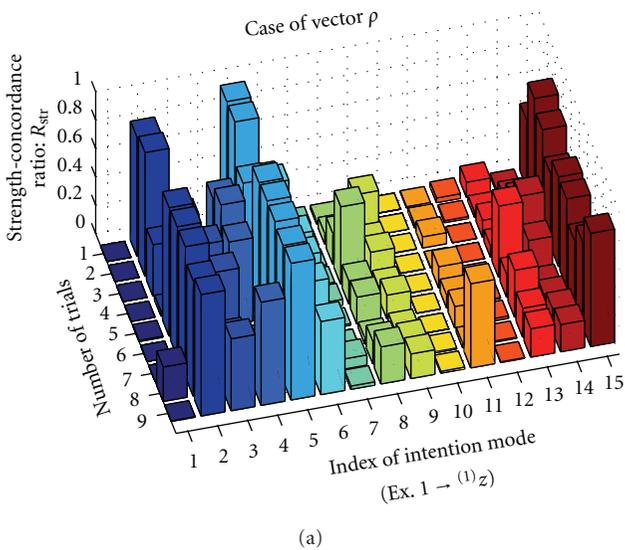


FIGURE 14: Comparison of strength-concordance ratios R_{str} in two cases: results using vector ρ (a) and scalar ρ (b).

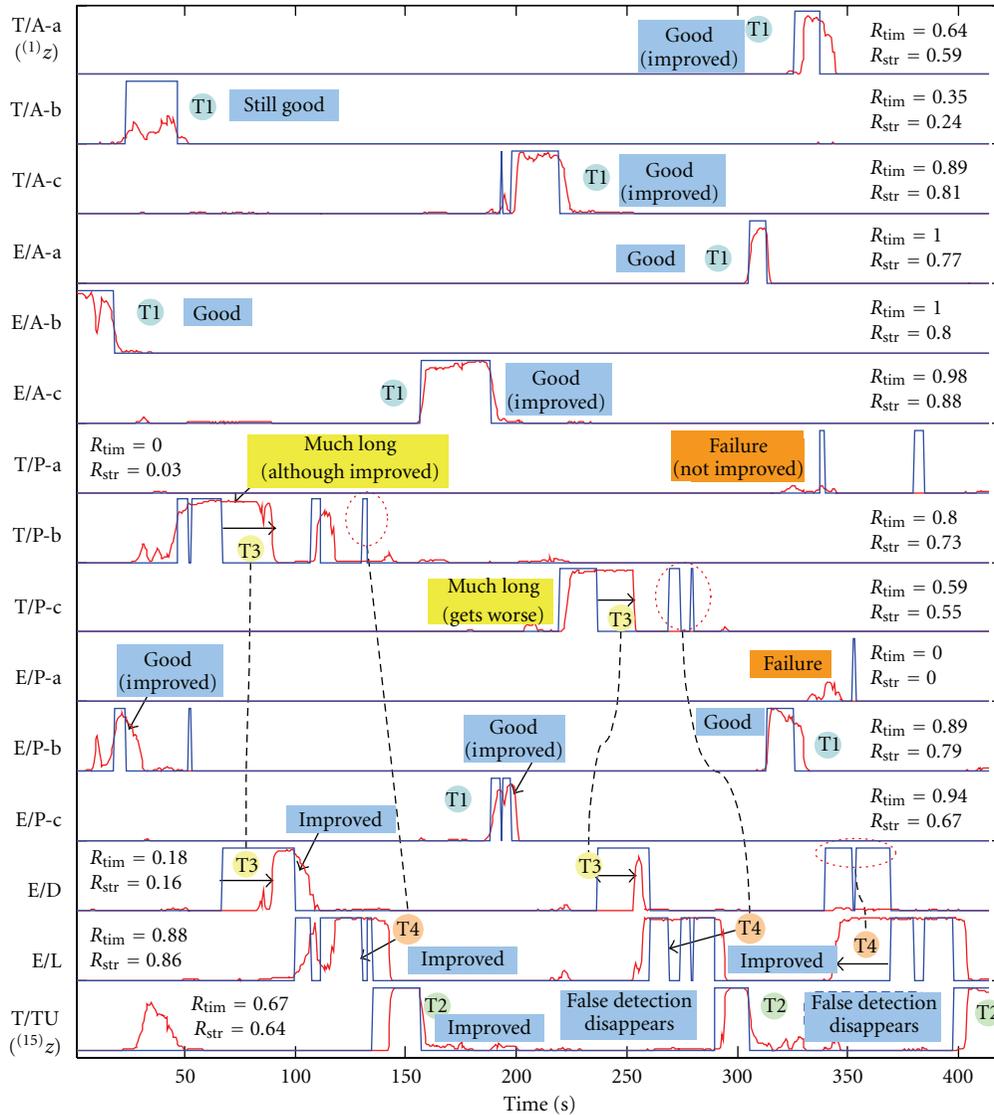


FIGURE 15: Transitions of intentions: red lines were identified by the SOM-Bayes estimator using scalar ρ . Blue lines were discerned by human analyst.

modes such as T/A-a, T/A-c, E/A-c, E/P-b, E/P-c, E/D, E/L, and T/TU. Detailed analysis is mentioned below.

6.1. *Tendency Analysis.* When the overlap, timing, strength, and types of the intention are qualitatively investigated by comparing \hat{z} against \tilde{z} in Figure 15, the following tendencies were found.

- (T1) Concerning *approaching*, the operational intentions in both cases of the truck (T/A-*) and of the excavator (E/A-*) were identified adequately. And *excavator positioning* (E/P-a,b) was also identified well.
- (T2) Concerning *truck’s transport and unloading* (T/TU), the identification of this mode was improved; however, the starting timings of the identified intentions were delayed.

(T3) The periods identified as truck’s positioning (T/P-b, T/P-c) were larger than that of the corresponding human intentions.

(T4) Although detection of the *excavator loading* (E/L) was improved, the periods were longer than human intentions.

The numberings of tendencies, (T1)–(T4), are written in Figure 15 to show relation between above-mentioned tendencies and waveforms. The reason of success mentioned in the tendency (T1) appears to come from large change in variables of the machine’s status. The reason of the delay indicated by the tendency (T2) is that the analyst regarded the end of the E/L action (that occurred before the T/T action) as a start of T/T.

The tendency (T3) for T/P was found by investigating which status changed at the same timing of E/D in \hat{z} (their

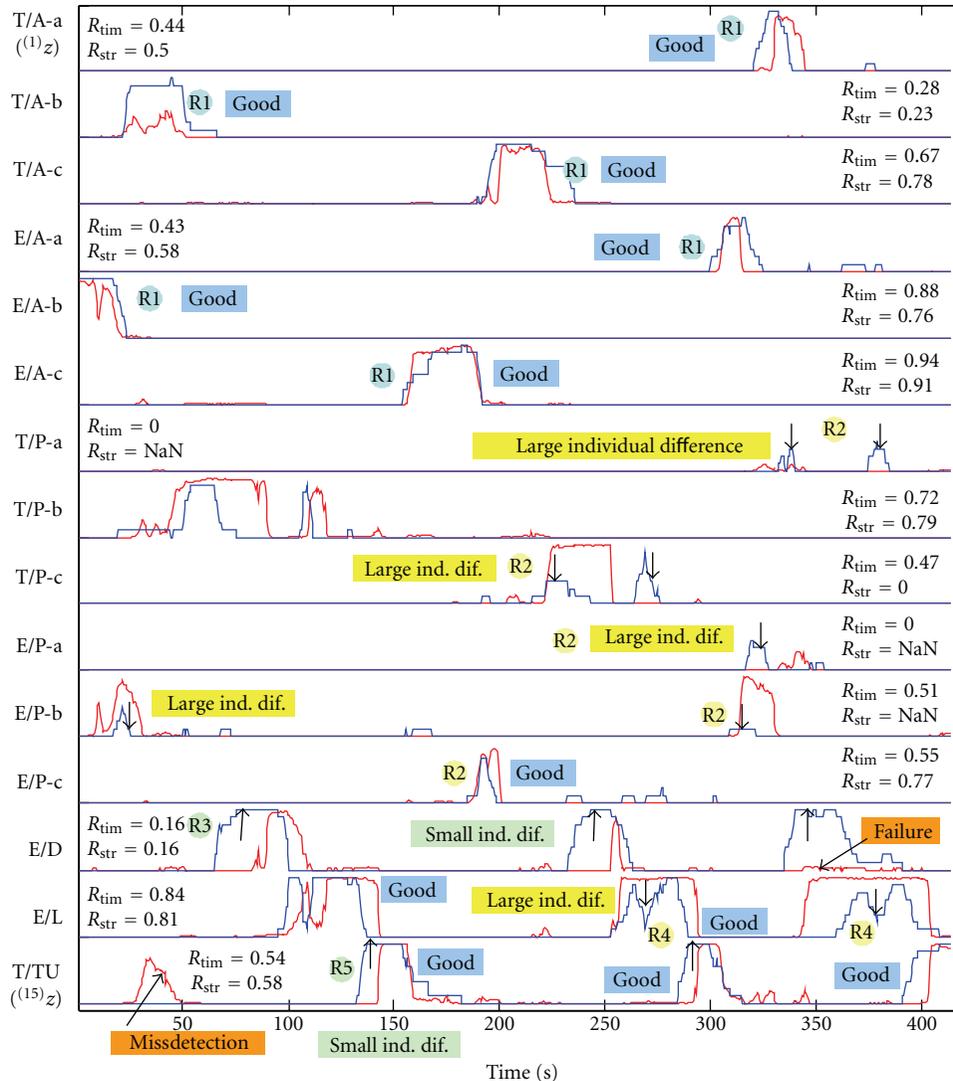


FIGURE 16: Transitions of intentions: red lines were identified by the SOM-Bayes estimator using scalar ρ . Blue lines were averages of results discerned by eight human analysts.

timings were about 85 [s], 255 [s]). And checking the raw logging data we found that the timing detected by the estimator was synchronized with change of the bucket's vertical manipulation while analyst's \tilde{z} began to change at the time the bucket were moved in a front-back direction before the vertical manipulation. Since the analyst can check the operator's body movement (as shown in Figure 8), it appeared that the intentions were discerned by unconsciously predicting the operator's hand reaching action based on the monitoring image. In other words, the analyst appears to guess other person's action earlier than actual action by considering the sequence of events and its causality. This might be the difference between the human discrimination and the estimator's identification.

Concerning the tendency (T4), the estimator's identification might be interpreted as follows: the truck's positioning at the phase of "loading of payload" was included in the E/L operation. Such interpretation of "T/P is a part of E/L"

can be acceptable for us. If we look at this from another point of view, this fact highlights human ambiguousness of criterion to classify intentions. This question will be analyzed in Section 6.3.

6.2. Qualitative Analysis of the SOM Clusters. To confirm findings about the SOM training, relation between characteristics of clusters and the accuracies of identified intention was investigated by checking Figure 10. In terms of the relative position among clusters, the cluster corresponding to continuous operations such as "E/D \rightarrow E/L" were formed close to each other. This does not contradict an intuitive feeling that intentions in continuous sequential operations resemble each other. To several modes of the excavator such as E/D, E/L, and E/P, multiple clusters were formed. This phenomenon appears to come from the difference of the working place since the number of multiple clusters are the same as the number of sites. (The transitions of their

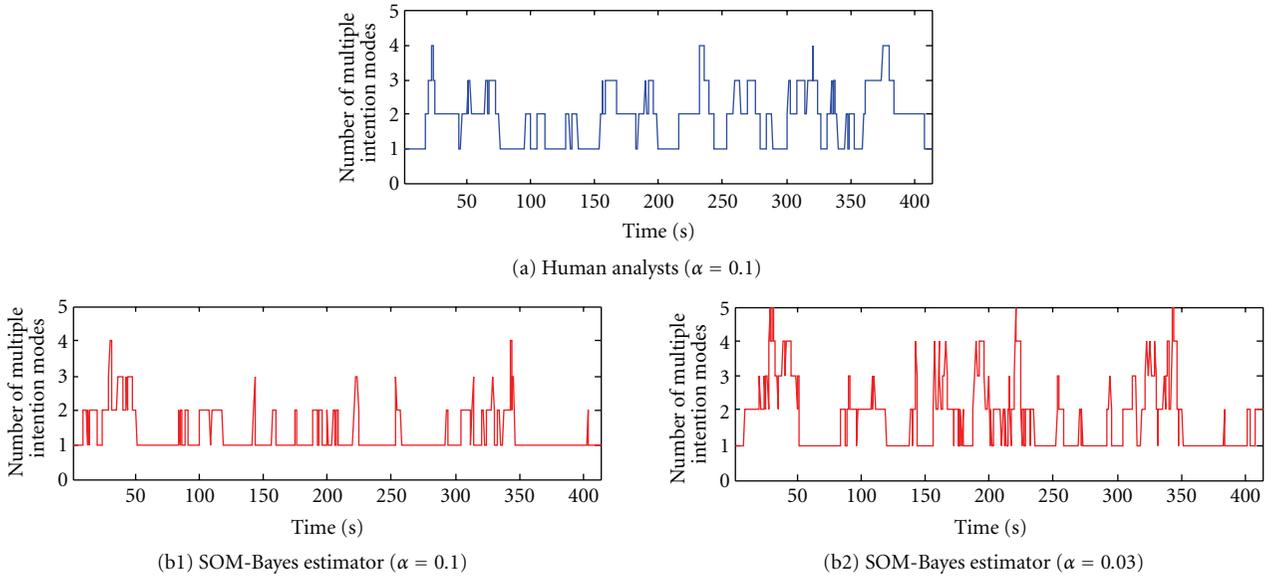


FIGURE 17: Transition of multiplicity of different intentions.

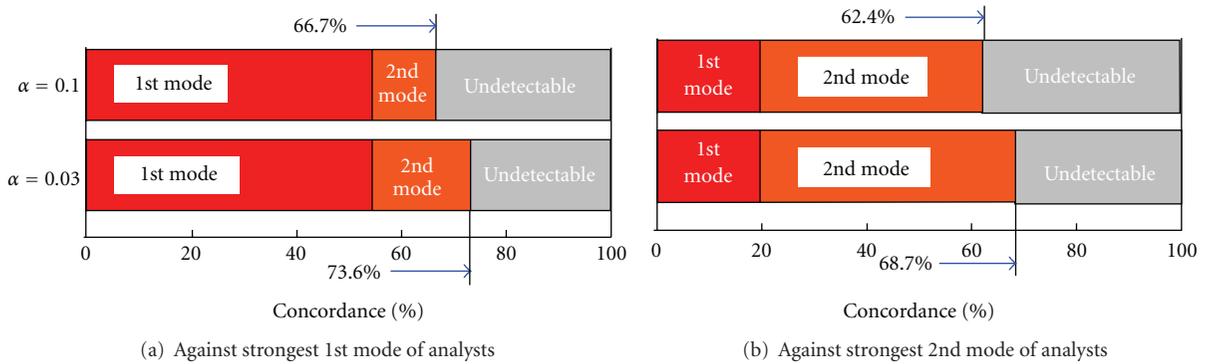


FIGURE 18: Concordance of identified intentions to human-discriminated ones when multiple different intentions are considered.

intention modes shown in Figure 15 change widely three times).

About size of clusters, the ED and EL for excavator occupy wide area in the SOM plane map. One could interpret this to mean that many nodes are assigned to conditions of most complex and significant operation in this digging task. In contrast, intention modes having small region in the map were not identified sufficiently. For instance, the cluster size of the identification failed E/P-a—the area is located around the coordinates of (46, 2) in Figure 10—is as small as four nodes which is equal to 0.3 percent of total area of the whole map. Since small cluster means small likelihood of the probability computation in the presented algorithm, it appears that this intention mode could not be identified sufficiently because of the small probability due to the small size of the corresponding cluster. Conversely, it would appear that the size of the total map should be adjusted so as to assign a sufficient number of modes to small occupied clusters.

6.3. Analysis of Individual Differences in Human Discrimination. In the previous section, performance of the intention

estimator was investigated by comparing with another intention, \tilde{z} , that was discriminated by human analyst. The \tilde{z} is, however, not absolutely unique since it was discriminated subjectively. As mentioned in Section 6.1, there is a possibility that discrimination of the operator’s intention differs in individuals. Therefore, other new eight participants were commissioned to discriminate intentions of the remote operation task, and individual differences among them were investigated. Intention modes discriminated by the eight participants are averaged, and the bog-standard intention is denoted by \bar{z} below. Figure 16 shows renewal graph of which old blue lines of \tilde{z} in Figure 15 were replaced by \bar{z} . The more close \bar{z} is to 1, the less differences discriminations by the analysts. Investigation of the tendencies in the characteristics of humans’ \bar{z} yields the following results.

- (R1) Concerning *approach* (T/A-a · · · E/A-c): maximum of components in \bar{z} are close to 1. Since their waveforms of \bar{z} look like triangle, individual differences in the timing of the starting and ending are confirmed. On the other hand, the estimation \hat{z} works well since \hat{z} matches \bar{z} relatively well.

- (R2) Concerning *positioning* (T/P-a · · · E/P-c): large individual differences in human discrimination are confirmed since each component in \bar{z} is small. Half of the corresponding components in \hat{z} were detected in the same timing of the \bar{z} case.
- (R3) Concerning *digging* (E/D): differences between individuals are small since the corresponding components in \bar{z} are close to 1 and form a rectangular shape. Identified intention \hat{z} is, however, insufficient because waveforms of \hat{z} does not match \bar{z} .
- (R4) Concerning *loading* (E/L): large individual differences are confirmed since waveforms of the components in \bar{z} split into triangle shapes. The corresponding components in \hat{z} cover the others in \bar{z} over long periods.
- (R5) Concerning *transport* (T/TU): the individual difference of the analysts is small since this component in \bar{z} is large and the shape of waveform is rectangular. Although the timing of \hat{z} is delayed against that of \bar{z} , the estimator works well.

From the result (R2) against (R1), it was found that even human analysts could not discriminate several intentions and the estimator could not also work well in such a case. The result (R3) indicates that the human discrimination differs from a way of the machine identification beyond the individual differences. This reason appears to come from human discrimination approach which utilizes also observation of operators' body motion. Or maybe human analysts utilizes task context for the discrimination by predicting the operator's intention. Therefore, a task scenario involving various judgement conditions might be required to enhance the presented intention estimator. It appears that same reason causes issues mentioned at the (R4) and (R5).

In order to evaluate the above-mentioned results quantitatively, the following indices are computed and were summarized in Table 2: the maximum of the intention level average found by eight participants, $\max(\bar{z})$, and the modified concordance ratios that are computed by (22) and (23) using new intentions identified by the improved estimator

$${}^{(k)}R_{\text{tim}2} := \frac{\sum_{\text{all } t'} {}^{(k)}\bar{z}}{\sum_{\text{all } t} {}^{(k)}\bar{z}} \quad t' = \{t \mid {}^{(k)}\hat{z}_t > 0.3\}, \quad (22)$$

$${}^{(k)}R_{\text{str}2} := \frac{\sum_{\text{all } t'} {}^{(k)}\hat{z}}{\sum_{\text{all } t} {}^{(k)}\bar{z}} \quad t' = \{t \mid {}^{(k)}\bar{z}_t > 0.7\}. \quad (23)$$

$R_{\text{tim}2}$ defined by (22) is essentially the same as the former ratio defined by (19). $R_{\text{str}2}$ defined by (23) shows the degree of the strength-concordance of the estimator's result \hat{z} against the intention periods which were agreed by 70 percent members of eight analysts. $\max(\bar{z})$ for ${}^{(4)}z, {}^{(7)}z, \dots, {}^{(12)}z$ described in Table 2 are as small as 0.5–0.88 while others are 1; hence, the individual differences to classify their modes were large. $R_{\text{str}2}$ of some modes could not be computed, that was $R_{\text{str}2} = \text{NaN}$ (not a number), because the \bar{z} -value of the mode could not

be determined due to large individual differences in the eight analysts. In other words, concerning such intention modes, it was impossible to compare the human discrimination with the machine estimation. Checking the other intention modes that have small individual differences (i.e., for ${}^{(i)}z$ satisfying $\max({}^{(i)}\bar{z}) = 1$), ${}^{(2)}z$ (T/A-b) indicates small concordance ratio as $R_{\text{tim}2} = 0.28$ and $R_{\text{str}2} = 0.23$. The upper second waveform in Figure 16 for this mode shows that the timing of \hat{z} 's activation coincided with \bar{z} but the value of \hat{z} was small. Other concordance ratio for ${}^{(13)}z$ of E/D indicates also small value as $R_{\text{tim}2} = 0.16$ and $R_{\text{str}2} = 0.16$, and this reason comes from the timing-shift given by the analysts' predictive discrimination. Except these particular cases, however, the other concordance ratios for ${}^{(1)}z, {}^{(3)}z, {}^{(5)}z, {}^{(6)}z, {}^{(14)}z$, and ${}^{(15)}z$, which were discriminable for the most human analysts, are large as $R_{\text{tim}2} = 0.44\text{--}0.94$ and $R_{\text{str}2} = 0.50\text{--}0.91$; hence, it can be said that estimation accuracies of each intention mode are comparatively good.

6.4. Multiplicity of Different Intentions. In former analysis relating to Figures 15 and 16, time transitions of most dominant intention-mode were mainly treated. The proposed algorithm of the intention estimator directly computes the ratio of all intention modes; hence, it is possible to investigate behavior of the multiple different intention. Transition of the number of the multiple difference intention is shown in Figure 17. The graph was obtained by drawing the change of the number of modes whose intention level becomes larger than a threshold α . The graph (a) is for comparison, and shows the pseudomultiplicity of modes (\bar{z}) discriminated by eight analysts in case of $\alpha = 0.1$. Although this graph does not show strictly multiplicity of one person's intentions and expresses an individual difference of eight person, it can be thought that the maximum of the multiplicity of standard person are four. Graphs (b1) and (b2) are obtained similarly using identified modes \hat{z} by the estimator, and are cases for $\alpha = 0.1$ and $\alpha = 0.03$, respectively. From these graphs, the estimator identified four or five candidates as intention modes when the threshold is set as $\alpha = 0.1$ or $\alpha = 0.03$, respectively. That is, both machine estimator and human analyst show similar possibility of multiple intentions.

Based on the results, concordance ratio between the machine estimator and human analysts were investigated by considering the multiplicity. Concordance percentage of first and second intention modes by the estimator against the human-discriminated intentions are shown in the graphs (a) and (b) in Figure 18, respectively. The upper and lower bar charts on each graph show the percentages of the concordance for $\alpha = 0.1$ and $\alpha = 0.03$, respectively. In any case shown in the figure, ratio of more than third mode is zero, that is, the estimator identified the same type of modes to human-discriminated ones by only the first or second dominant intention modes. It can be said that the proposed algorithm works comparatively well as the human does if the first and second candidates of the identified modes are considered because the sum of concordance ratio of the estimator's identification against human-discrimination indicates 62–73 percent.

TABLE 2: Maximum of intention level ($\max(\bar{z})$) and concordance ratios ($R_{\text{tim}2}, R_{\text{str}2}$).

Label	⁽¹⁾ z	⁽²⁾ z	⁽³⁾ z	⁽⁴⁾ z	⁽⁵⁾ z	⁽⁶⁾ z	⁽⁷⁾ z	⁽⁸⁾ z	⁽⁹⁾ z	⁽¹⁰⁾ z	⁽¹¹⁾ z	⁽¹²⁾ z	⁽¹³⁾ z	⁽¹⁴⁾ z	⁽¹⁵⁾ z
	T/A-a	T/A-b	T/A-c	E/A-a	E/A-b	E/A-c	T/P-a	T/P-b	T/P-c	E/P-a	E/P-b	E/P-c	E/D	E/L	T/TU
$\max(\bar{z})$	1	1	1	0.88	1	1	0.5	0.88	0.88	0.5	0.5	0.75	1	1	1
$R_{\text{tim}2}$	0.44	0.28	0.67	0.43	0.88	0.94	0	0.72	0.47	0	0.51	0.55	0.16	0.84	0.54
$R_{\text{str}2}$	0.50	0.23	0.78	0.58	0.76	0.91	NaN	0.79	0	NaN	NaN	0.77	0.16	0.81	0.58

TABLE 3: List of constants.

Symbols	Designation	Values in verification
α	Threshold to discern intention modes	0.1 or 0.03
$f \xi_i$	Reference vectors of the $SOM_f (i = 1, \dots, L_f)$	(Determined by the learning)
$g \xi_i$	Reference vectors of the $SOM_g (i = 1, \dots, L_g)$	(Determined by the learning)
σ_i	Standard deviation of i th column sequence in $\{s\}$	(Depend on data)
n_s	Size of vector s	28
n_z	Number of intention types	15
$(u_{\text{dim}}, v_{\text{dim}})$	Sizes of the SOM lattice	
L_f	Number of all nodes of SOM_f	1250 (50×25)
L_g	Number of all nodes of SOM_g	1600 (80×20)
$2L_r + 1$	Number of nodes of one side of a square-like area $A_{g_i^{[m]}}$	11
M	Number of all particles	1000
T	Total time of task (integer)	(Depend on trial)

7. Conclusion and Discussion

A method to estimate the operational intention in a human manipulation of a machine was introduced in the present paper. This method utilizes a clustering technique by SOM to compute a state transition probability of intentions. This method enables to omit troublesome process to specify types of information which should be used to build the estimator. By embedding the state transition property expressed by the SOM mapping into a particle filtering algorithm, the operator's intention is identified through a Bayes estimation. Applying the proposed method to the remote operation task, the estimator's behavior was analyzed, the pros and cons of the method were investigated, and ways for the improvement were discussed. Moreover, through investigation of normative data of intentions discriminated by human analysts, issues in the verification were also treated.

Concerning the design of the SOM-Bayes intention estimator algorithm, the following findings were obtained.

(F1) Reducing the redundancy in the input vector for the SOM training is effective to improve the estimator.

(F2) It was confirmed that an estimation accuracy of the intention mode is good (not good) when the corresponding cluster on the SOM plane map is large (small). Therefore, the size of the total map should be adjusted so as to assign a sufficient number of modes to small occupied clusters.

By the countermeasure mentioned in (F1), the accuracy was increased in eight intentional modes among a total of 15 modes, and the effectiveness of this findings was confirmed. Although experimental proof concerning (F2) was not performed in this paper, the proposed algorithm is a so-called frequentism method consisting of Bayes estimation; hence, the present authors predict that the approach of (F2) will be effective. As another approach for the improvement, modification of the input data can be considered in order to enhance a fixation of the related network in the SOM structure for the significant but rare events at the stage of the SOM training. A finding about individual difference in human discrimination is as follows:

(F3) there are differences in discrimination of intentions among individuals. Difficult circumstances to human analysts were also difficult to the present estimation algorithm.

For normal intention modes whose periods can be found by 70 percent members of all analysts, the concordance ratios between the identified intention by the proposed algorithm and the human-discriminated ones were as high as 0.44–0.94 if exceptional modes were removed in case of this remote operation task. Additionally, when the multiplicity of the different intentions is considered, the estimator could identify the same type of intention modes to human-discriminated ones by using the first and second dominant intention modes as high as 62–73%. The estimation accuracy is sufficiently high considering that this method does not utilize predictive estimation from the user's motion. Therefore, it can be concluded that the proposed algorithm works comparatively well.

Considering the findings based on the experiment, one method to improve the estimation is a combination with a prediction function using measurement of the operator's hand/body motion. Use of some task scenario appears to be another effective method. The former approach is, however, needs additionally the measurement device and the image-processing; hence, a tradeoff between complexity of the system and enhancement of the intention-estimation is required. The present authors would like to study such

TABLE 4: List of variables.

Symbols	Designation	Types/range/definition
η	Bayes normalization parameter	Real number
κ_c	Operation mode for crawlers	$\{0, 1, \dots, 8\}$
${}^e\kappa_a$	Operation mode for the arm of the excavator	$\{0, 1, \dots, 2\}$
${}^e\kappa_b$	Operation mode for the bucket of the excavator	$\{0, 1, \dots, 2\}$
${}^e\kappa_r$	Operation mode for the superstructure rotation of the excavator	$\{0, 1, \dots, 2\}$
ρ	Commands to the machine from an operator	Scalar integer
$\hat{\rho}^{(i')}$	Value corresponding to an operational command which is registered in i' th reference vector of ${}^g\xi$	Real number
Δ	Random value having standard deviation σ_i	Real number
$\text{bel}(z)$	Belief of the intention z (probabilistic distribution of z)	Vector with real numbers
$\overline{\text{bel}}(z_t)$	Prediction at time t (a predicted belief)	Vector with real numbers
$f c_t^{[m]}$	Node index of m th particle's BMN on SOM_f	$\{1, \dots, L_f\}$
$g c_t^{[m]}$	Node index of SOM_g whose reference vector is most close to a combination of the measured command and the predicted status	$\{1, \dots, L_g\}$
i	Index for a particle	Integer
l	Number of nodes whose $\hat{\rho}$ is close to ρ_t	Integer
m'	Number registered in the r th element of the array W (a particle number)	Integer
n_W	Length of the roulette array W (this is variable on each time counter)	Integer
q_M	Machine status (M -status)	Vector with real numbers
q_E	Environmental status (E -status)	Vector with real numbers
q_T	Task status (T -status)	Vector with real numbers
r	Random number	$\{1, \dots, n_W\}$
s	Information recognized by a human operator	Vector with real numbers
Δs_t	Random sample point around s_t	Vector with real numbers
t	Time counter	Integer
(u_c, v_c)	Coordinate value corresponding to the node $g c_t^{[m]}$ on the SOM_g plane map	Integers
$w_t^{[m]}$	Measurement probability corresponding to m th predicted particle at time t	Real number
z	Operational intention of a human	Vector in $[0, 1]^{n_z}$
\bar{z}	Average of operator's intention level discerned by eight human analysts	Vector in $[0, 1]^{n_z}$
\hat{z}	Intentions identified by the SOM-Bayes estimator (Average of the identified intentions given by M particles)	by (15)
\tilde{z}	Intention discerned by a human analyst	Vector in $[0, 1]^{n_z}$
${}^{(i)}z$	i th element of vectors z (i th intention mode)	Real number
$B z_t^{[m]}$	m th particle of the set Z_t ($m = 1, \dots, M$)	Vector with real numbers
$B \bar{z}_t^{[m]}$	Prediction at time t by m th particle	by (10)
${}^{(k)}R_{\text{tim}}$	Concordance ratio with respect to time on k th intention mode (time-concordance ratio)	$[0, 1]$
${}^{(k)}R_{\text{str}}$	Concordance ratio with respect to strength on k th intention mode (strength-concordance ratio)	$[0, 1]$
${}^{(k)}R_{\text{tim}2}$	Time-concordance ratio against average intention of eight analysts' discrimination on k th intention mode	$[0, 1]$
${}^{(k)}R_{\text{str}2}$	Strength-concordance ratio against average intention of eight analysts' discrimination on k th intention mode.	$[0, 1]$
$A_{g c_t^{[m]}}$	Square-like area on the SOM_g plane map around the node of $g c_t^{[m]}$	—

TABLE 4: Continued.

Symbols	Designation	Types/range/definition
W	Roulette array	Size is variable
Z_t	Set of particles at time t	Set of vector

TABLE 5: List of functions.

Symbol	Meanings	Definitions
$H()$	Conceptual human controller model. The input and output are assumed as s and ρ , respectively	in Figure 2
$f()$	State transition function describing change of the intention	(1)
$g()$	Output function of operational command	(2)
$p(z_t s_t, z_{t-1})$	Conditional probabilistic distribution of a transition of intention from z_{t-1} to z_t	
$p(\rho_t z_t)$	Conditional probabilistic distribution of judgment that outputs ρ_t if the intentions z_t happens to be true	
SOM_f	SOM map for the function f (SOM mapping function which is trained using an input vector consisting of z and s)	(5)
SOM_g	SOM map for the function g (SOM mapping function which is trained using an input vector consisting of z and ρ)	(6)

a combination approach by developing the proposed SOM-Bayes intention estimator in future work since we have been studying the hand-reaching action [29].

Appendix

Constants, variables, and functions used in this paper are summarized in the following lists. $\{x\}$ describes a time series sequence data of the vector signal x . Or it is used to express a set consisting of elements denoted in the braces. See Tables 3, 4, and 5.

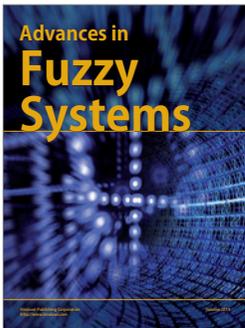
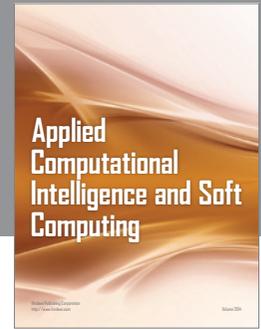
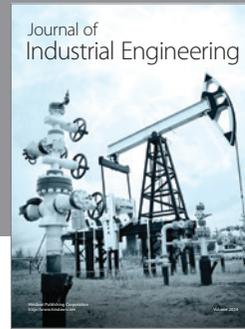
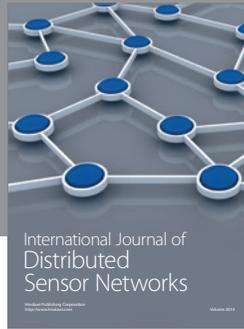
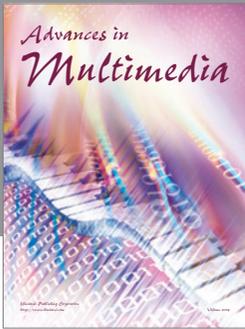
Acknowledgments

The present study were supported by a Grant-in-Aid for Scientific Research (A) of the Japanese Ministry of Education, Culture, Sports, Science, and Technology. The experiment was supported by many participants who embraced the authors' requests kindly. The present author appreciates their cooperation.

References

- [1] E. Horvitz, J. Breese, D. Heckerman, D. Hovel, and K. Rommelse, "The Lumière project: bayesian user modeling for inferring the goals and needs of software users," in *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*, pp. 256–265, Madison, Wis, USA, 1998.
- [2] I. Zukerman, D. W. Albrecht, and A. E. Nicholson, "Predicting users' requests on the WWW," in *Proceedings of the 7th International Conference on User Modeling*, pp. 275–284, Banff, Canada, 1999.
- [3] J. Gutman, "A means-end chain model based on consumer categorization process," *Journal of Marketing*, vol. 46, pp. 60–72, 1982.
- [4] K. Kasaoka and Y. Sankai, "Predictive control estimating operator's intention for stepping-up motion by exo-skeleton type power assist system HAL," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, (IROS '01)*, pp. 1578–1583, Maui, Hawaii, USA, November 2001.
- [5] K. Sakita, K. Ogawara, S. Murakami, K. Kawamura, and K. Ikeuchi, "Flexible cooperation between human and robot by interpreting human intention from gaze information," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, (IROS '04)*, pp. 846–851, Sendai, Japan, October 2004.
- [6] C. Breazeal, *Designing Sociable Robots*, MIT Press, 2002.
- [7] T. Shibata, "An overview of human interactive robots for psychological enrichment," *Proceedings of the IEEE*, vol. 92, no. 11, pp. 1749–1758, 2004.
- [8] K.-S. Park and D.-S. Kwon, "A cognitive modeling for mental model of an intelligent robot," *International Journal of Assistive Robotics and Mechatronics*, vol. 7, no. 3, pp. 16–24, 2006.
- [9] K. Furuta, "Control of pendulum: from super mechano-system to human adaptive mechatronics," in *Proceedings of the 42nd IEEE Conference on Decision and Control*, pp. 1498–1507, Maui, Hawaii, USA, December 2003.
- [10] F. Harashima and S. Suzuki, "Intelligent mechatronics and robotics," in *Proceedings of the IEEE International Conference on Emerging Technologies and Factory Automation, (ETFA '08)*, Hamburg, Germany, 2008.
- [11] W. D. Gray, B. E. John, and M. E. Atwood, "Project Ernestine: validating a GOMS analysis for predicting and explaining real-world task performance," *Human-Computer Interaction*, vol. 8, no. 3, pp. 237–309, 1993.
- [12] B. Price, "Frank and Lillian Gilbreth and the motion study controversy, 1907–1930," in *A Mental Revolution: Scientific Management since Taylor*, D. Nelson, Ed., The Ohio State University Press, 1990.
- [13] T. Suzuki, S. Sekizawa, S. Inagaki et al., "Modeling and recognition of human driving behavior based on stochastic switched ARX model," in *Proceedings of the 44th IEEE Conference on Decision and Control, and the European Control Conference, (CDC-ECC '05)*, pp. 5095–5100, Seville, Spain, December 2005.
- [14] H. Kawashima and T. Matsuyama, "Multiphase learning for an interval-based hybrid dynamical system," *IEICE Transactions*

- on *Fundamentals of Electronics, Communications and Computer Sciences*, vol. E88-A, no. 11, pp. 3022–3034, 2005.
- [15] T. Inamura, I. Toshima, H. Tanie, and Y. Nakamura, “Embodied symbol emergence based on mimesis theory,” *International Journal of Robotics Research*, vol. 23, no. 4-5, pp. 363–377, 2004.
- [16] Y. Inagaki, H. Sugie, H. Aisu, and T. Unemi, “A study of a method for intention inference from human’s behavior,” in *Proceedings of the IEEE International Workshop on Robot and Human Communication (Ro-Man ’93)*, pp. 142–145, 1993.
- [17] T. Choudhury, G. Borriello, and D. Haehnel, “The mobile sensing platform: an embedded system for activity recognition,” in *Proceedings of the IEEE Pervasive Magazine—Special Issue on Activity-Based Computing*, pp. 32–41, 2008.
- [18] J. A. Ward, P. Lukowicz, and H. W. Gellersen, “Performance metrics for activity recognition,” *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 1, article 6, 23 pages, 2011.
- [19] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.
- [20] S. Chizuwa and M. Kameyama, “Bayesian-network-based intention estimation for a user support system of an information appliance,” in *Proceedings of the 5th International Conference on Information Technology and Applications (ICITA ’08)*, pp. 71–76, Queensland, Australia, 2008.
- [21] T. Kohonen, “The self-organizing map,” *Proceedings of the IEEE*, vol. 78, no. 9, pp. 1464–1480, 1990.
- [22] S. Suzuki and F. Harashima, “Segmentation and analysis of console operation using self-organizing map with cluster growing method,” in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS ’09)*, pp. 4875–4880, St. Louis, Mo, USA, October 2009.
- [23] S. Suzuki and F. Harashima, “Bayesian intention estimator using self-organizing map and its experimental verification,” in *Proceedings of the 19th IEEE International Symposium on Robot and Human Interactive Communication (Ro-Man ’10)*, pp. 290–295, Viareggio, Italy, 2010.
- [24] S. Suzuki and F. Harashima, “Estimation of operational intentions utilizing self-organizing map with bayes filtering,” in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS ’10)*, pp. 2249–2255, Taipei, Taiwan, 2010.
- [25] S. Suzuki and F. Harashima, “Skill analysis focused on the hand discrete movement for machine manipulation,” in *Proceedings of the 13th IEEE Int. Conf. of Emerging Technologies and Factory Automation (ETFA ’08)*, pp. 156–163, Hamburg, Germany, 2008.
- [26] B. J. Baars, *A Cognitive Theory of Consciousness*, Cambridge University Press, New York, NY, USA, 1988.
- [27] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*, MIT Press, Cambridge, Mass, USA, 2005.
- [28] G. Winkler, *Image Analysis, Random Fields, and Dynamic Monte Carlo Methods*, Springer, Berlin, Germany, 1995.
- [29] S. Suzuki and F. Harashima, “Skill evaluation from observation of discrete hand movements during console operation,” *Journal of Robotics*, vol. 2010, Article ID 967379, 13 pages, 2010.
- [30] S. Suzuki, “Characteristics analysis of a task switching during console operation based on the self-organizing map,” *International Journal of Advanced Mechatronic Systems*, vol. 2, no. 5-6, pp. 306–317, 2010.
- [31] Helsinki University of Technology, Neural Networks Research Centre Web-site, <http://www.cis.hut.fi/research/som-research/nncr-programs.shtml>.
- [32] T. Kohonen, *Self-Organizing Maps*, Springer, Berlin, Germany, 1995.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

