Proceedings of the 1994 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI'94) • Las Vegas, NV • Oct. 2-5, 1994

An Active Sensing Method Using Estimated Errors for Multisensor Fusion Systems

Toshiharu Mukai and Masatoshi Ishikawa
Faculty of Engineering, University of Tokyo
Bunkyo-ku, Tokyo 113, Japan
e-mail: tosh@k2.t.u-tokyo.ac.jp

Abstract — An active sensing method for multisensor fusion systems with actuators is proposed. To realize active sensing with multiple sensors, i) where to position sensors, ii) how to associate data, and iii) how to fuse data should be determined. The authors propose a new method mainly concerning i). The method utilizes estimated errors of estimated values to determine optimal sensor locations where useful data are expected to be obtained and effectively associated. An algorithm to calculate nearly optimal sensor locations, instead of exact optimal locations, is also proposed to reduce calculation. As examples, the active sensing method is applied to multi-target tracking by a system with two hand-eye cameras, and visual and tactile fusion in a system with a camera and a tactile sensor. By using this method, the sensing strategy is optimized for the object of measurement.

I. Introduction

In recent years, sensor fusion, which is a technique in engineering to realize multisensor recognition systems more powerful than single-sensor systems, has attracted much attention [1, 2]. Research on sensor fusion has involved, for example, studies of an autonomous land vehicle (ALV), which has multiple sensors and geometrical maps and moves autonomously [3] and visual and tactile fusion using visual and tactile sensors usually to control a manipulator [4].

The ultimate purpose of many of these sensor fusion systems is for operation in the real world. To achieve this purpose, it is essential to be able to sense unknown or nearly unknown objects. For example, when an ALV encounters anything on its route, the ALV should recognize it and determine its position to decide how to react toward it. In such cases, how and what to sense is not clear in the design stage of the system. Therefore the system should formulate a sensing strategy gradually, using information obtained over time.

In addition, the goals of these systems are generally not only to acquire information but also to control objects or environments. For example, one of the goals of an ALV is to reach its destination, and one of the goals of a manipulator using visual and tactile fusion is to handle objects using sensed information.

It is natural to consider that such a sensor fusion system includes not only multiple sensors but also actuators to produce behavior. In this case, the actuators can be used as elements of sensing. For instance, when a robot hand holds a object, it may search for a good grip by moving its fingers. In such a case, we can say that there is close relationship between behavior of actuators and sensing.

Methods using actuators for sensing as mentioned above are known as active sensing [5–7]. In active sensing of unknown objects, it is necessary to determine the actuator's motion to realize optimal sensing by effectively using information obtained over time. This corresponds to formulating a sensing strategy. Similar concepts are known in psychology as the affordance concept of Gibson (an object exists with actions on it being included) [8] and the perceptive circulation of Neisser (perception is based on the repetition of the cycle: measurement → identification of the model → prediction → measurement) [9]. These are considered to be parts of basic human sensing behavior.

Active sensing in engineering has been studied to some extent. However, even the definition is not yet fixed and further studies are expected. In this paper, we propose a method to realize active sensing. Although there are many types of active sensing, here, we concentrate on the problem of multisensor systems positioning sensors at locations where necessary information is expected to be obtained.

In active sensing using multiple sensors, problems arise such as

i) where to position sensors,
ii) how to associate information from multiple sensors,
iii) how to fuse associated data.

0-7803-2072-7/94 $4.00 © 1994 IEEE

Here, the term "associate" is used to mean acquiring correct association of data from sensors and sensed objects. In this paper, we mainly deal with and propose a new method for i). Conventional methods of the nearest neighbor data association algorithm [7] and parallel Kalman filters [10] are used for ii) and iii).

In the following sections, first, the method to select optimal sensor locations using estimated errors is proposed. In this method, not only the accuracy of estimates but also the suitability for association is considered. In addition, a concrete algorithm for the proposed method is described.

Furthermore, as examples, we first study a multi-target tracking problem with two cameras mounted on the tips of manipulators. The movement of objects is assumed to be modeled. The cameras are moved to the locations where the best data are expected to be obtained. Next, we discuss the problem of acquiring position and orientation of a three-dimensional object using a fixed camera and a tactile sensor. In this case, applying information obtained by the camera to the three-dimensional model, optimal tactile sensor locations are determined. That is, rough position and orientation of the object is detected by the camera and, using this information, sites for contact by the tactile sensor are determined.

II. Selection of Sensor Locations

A. Selection of Sensor Locations to Minimize Estimated Errors

Here, we discuss the problem of how to select sensor positions to realize the best observation of time-variant objects. Tracking of moving objects with hand-eye cameras can be considered as an example. For simplicity, first, we concentrate on the case of one object and multiple sensors.

The state transition equation of the state vector $\boldsymbol{x}$ of the object and the observation equation (in the observation vector, outputs from multiple sensors are aligned) are

$$\boldsymbol{x}(t+1) = F(t)\boldsymbol{x}(t) + G(t)\boldsymbol{w}(t) \quad (1)$$

$$\boldsymbol{z}(t) = \boldsymbol{h}(\boldsymbol{x}(t), t) + \boldsymbol{v}(t), \quad (2)$$

where t is time, F and G are matrices, $\boldsymbol{h}(\cdot)$ is a nonlinear vector function, and $\boldsymbol{w}(t)$ and $\boldsymbol{v}(t)$ are zero mean noises with variances $Q(t)$ and $R(t)$, respectively. To estimate the state $\boldsymbol{x}$ from the observation $\boldsymbol{z}$, a Kalman filter is used. The merits of the filter are the following:

i) The recursive loop structure of the filter allows the estimates of the state to be adjusted incrementally with each new set of measurements.

ii) The prediction of the next state vector is incorporated as part of the filter algorithm.

iii) The variances of the estimate are evaluated as part of the algorithm.

iv) The estimator is optimal in a Bayesian (minimum variance) sense if all the state and observation noises are Gaussian, otherwise it is the optimal linear estimator.

The system is nonlinear, hence an extended Kalman filter is used. That is, the prediction of the state vector $\boldsymbol{x}(t+1)$ and the variance $P(t+1)$ at time t are

$$\hat{\boldsymbol{x}}(t+1|t) = F(t)\hat{\boldsymbol{x}}(t|t) \quad (3)$$

$$P(t+1|t) = F(t)P(t|t)F^T(t) + G(t)Q(t)G^T(t), \quad (4)$$

and after observation at time $t+1$, the estimates are updated by

$$\hat{\boldsymbol{x}}(t+1|t+1) = \hat{\boldsymbol{x}}(t+1|t) + W(t+1)[\boldsymbol{z}(t+1) - \boldsymbol{h}(\hat{\boldsymbol{x}}(t+1|t), t)] \quad (5)$$

$$P^{-1}(t+1|t+1) = P^{-1}(t+1|t) + \left(\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{x}}|_{\hat{\boldsymbol{x}}(t+1|t)}\right)^T R^{-1}(t+1)\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{x}}|_{\hat{\boldsymbol{x}}(t+1|t)} \quad (6)$$

with

$$W(t+1) = P(t+1|t+1) \cdot \left(\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{x}}|_{\hat{\boldsymbol{x}}(t+1|t)}\right)^T R^{-1}(t+1), \quad (7)$$

where $\frac{\partial \boldsymbol{h}}{\partial \boldsymbol{x}}|_{\hat{\boldsymbol{x}}(t+1|t)}$ is the Jacobian of $\boldsymbol{h}(\boldsymbol{x}, t)$ with $\boldsymbol{x}$ at $\hat{\boldsymbol{x}}(t+1|t)$, $\hat{\boldsymbol{x}}(t|t)$ is the estimate of $\boldsymbol{x}(t)$ at t using observations up to time t, and $\hat{\boldsymbol{x}}(t+1|t)$ is the estimate of $\boldsymbol{x}(t+1)$ using observations up to time t and the state transition equation (1). $P(t|t)$ corresponds to the variance $E[(\boldsymbol{x}(t) - \hat{\boldsymbol{x}}(t|t))(\boldsymbol{x}(t) - \hat{\boldsymbol{x}}(t|t))^T]$, where $E[\cdot]$ represents expectation.

Here, let us consider the movement of sensors to realize better estimation by the Kalman filter. Note that changing sensor locations means changing $\boldsymbol{h}(\cdot)$. Then, when we want to specify that $\boldsymbol{h}(\cdot)$ or P depends on sensor location, notation such as $\boldsymbol{h}(\cdot; \boldsymbol{x}_s)$ or $P(\cdot; \boldsymbol{x}_s)$ is used, where $\boldsymbol{x}_s = [\boldsymbol{x}_{s1}^T, \boldsymbol{x}_{s2}^T, \ldots, \boldsymbol{x}_{sM}^T]^T$: $\boldsymbol{x}_{sj}$ is the location of sensor j and M is the total number of sensors. In this paper, we define a good sensor location as one where the variance of

$\hat{\boldsymbol{x}}$ is small. That is, after prediction of $P(t+1|t+1;\boldsymbol{x}_s)$ at time t, we calculate

$$\mathcal{R}(\boldsymbol{x}_s) = \det P(t+1|t+1;\boldsymbol{x}_s), \tag{8}$$

and select sensor location $\boldsymbol{x}_s$ to minimize $\mathcal{R}(\boldsymbol{x}_s)$. As one can see in (6), $P(t+1|t+1;\boldsymbol{x}_s)$ does not invlove $\boldsymbol{z}(t+1)$. Hence, one can calculate $P(t+1|t+1;\boldsymbol{x}_s)$ at time t. Note that $\mathcal{R}(\boldsymbol{x}_s)$ is positive because P is positive definite and that $\mathcal{R}(\boldsymbol{x}_s)$ corresponds to the (scalar) variance of $\hat{\boldsymbol{x}}$, because it is the product of all eigenvalues of $P(t+1|t+1)$. Minimizing $\mathcal{R}(\boldsymbol{x}_s)$ is equivalent to selecting the location where we can obtain the best data as the result of fusing the prediction by the state transition equation and observations by sensors. Using the determinant of the variance of the estimate as the evaluation function is known as a D-optimum approximate design in optimum experimental designs [11] and the determinant of the variance is called the general variance.

It is difficult to calculate the defined $\mathcal{R}(\boldsymbol{x}_s)$ directly, so we discretize location $\boldsymbol{x}_s$ and select the optimal location from the finite candidates. Furthermore, selecting the location with small $\mathcal{R}(\boldsymbol{x}_s)$ is equivalent to selecting the location where

$$\mathcal{E}^p(\boldsymbol{x}_s) \equiv 1/\mathcal{R}(\boldsymbol{x}_s) = \det P^{-1}(t+1|t+1;\boldsymbol{x}_s) \tag{9}$$

is large. By using this evaluation function $\mathcal{E}^p$, since P^{-1} can be calculated from (6) directly, we need not calculate P from P^{-1}.

B. Selection of Location to Make Good Association Possible

Next, let us consider the case of multiple objects and sensors. When data of objects are obtained from multiple sensors, we must associate them to clarify which data originated from which object.

The state transition equation of object i is

$$\boldsymbol{x}_i(t+1) = F_i(t)\boldsymbol{x}_i(t) + G_i(t)\boldsymbol{w}_i(t) \quad (i = 1,\ldots,N), \tag{10}$$

corresponding to (1). The subscript i denotes the number of the object and N is the total number of objects. The measurement of this object through sensor j, corresponding to (2), is

$$\boldsymbol{z}_{j;i}(t) = \boldsymbol{h}_j(\boldsymbol{x}_i(t),t) + \boldsymbol{v}_{j;i}(t) \quad (j = 1,\ldots,M). \tag{11}$$

where $\boldsymbol{v}_{j;i}(t)$ is the zero mean noise independent of $\boldsymbol{w}_l$ $(l = 1,\ldots,N)$ and $\boldsymbol{v}_{m;i}$ $(m = 1,\ldots,M; m \neq j)$, and its variance is given by $R_{j;i}(t)$. $\boldsymbol{z}_{j;i}$ is assumed not to be observed when the object is out of range.

When outputs of sensor j are obtained, we must know which object is the origin of the outputs. Assume that L outputs from sensor j, $\tilde{\boldsymbol{z}}_{j;k}(t)$ $(k = 1,\ldots,L)$, are obtained at time t. Although various methods can be used for association, we use the nearest neighbor data association algorithm for simplicity. In this method, using

$$\hat{\boldsymbol{z}}_{j;i}(t+1|t) = \boldsymbol{h}_j(\hat{\boldsymbol{x}}_i(t+1|t)), \tag{12}$$

we associate $\tilde{\boldsymbol{z}}_{j;k}(t+1)$ which minimizes the Mahalanobis generalized distance

$$\begin{aligned}\mu_{j;ik} \equiv\ & (\hat{\boldsymbol{z}}_{j;i}(t+1|t) - \tilde{\boldsymbol{z}}_{j;k}(t+1))^T S_{j;i}(t+1|t)^{-1} \\ & \cdot(\hat{\boldsymbol{z}}_{j;i}(t+1|t) - \tilde{\boldsymbol{z}}_{j;k}(t+1))\end{aligned} \tag{13}$$

with $\boldsymbol{z}_{j;i}(t+1)$. $S_{j;i}(t+1|t)$ is the variance of $\hat{\boldsymbol{z}}_{j;i}(t+1|t)$ and is calculated as

$$S_{j;i}(t+1|t) = \frac{\partial \boldsymbol{h}_j}{\partial \boldsymbol{x}}|_{\hat{\boldsymbol{x}}_i(t+1|t)} P_i(t+1|t) \left(\frac{\partial \boldsymbol{h}_j}{\partial \boldsymbol{x}}|_{\hat{\boldsymbol{x}}_i(t+1|t)}\right)^T, \tag{14}$$

where P_i is the variance of the state vector $\boldsymbol{x}_i$. When $\mu_{j;ik}$ $(k = 1,\ldots,L)$ is larger than the appropriately defined threshold, no output is associated with $\boldsymbol{z}_{j;i}(t+1)$. The vector $\boldsymbol{z}_i(t+1)$ in which $\boldsymbol{z}_{j;i}(t+1)$ are aligned along j is the observation vector of object i, and in its variance $R_i(t+1)$, $R_{j;i}(t+1)$ are aligned diagonally. Using this $\boldsymbol{z}_i(t+1)$, the Kalman filter of the object i is formed as (3)$\sim$(7) with subscript i.

For accurate association using (13), sensor outputs from different objects must not be close. Equation (13) denotes the selection of the output which is nearest to the predicted output. Hence, in order to make this mechanism effective, the Mahalanobis generalized distance on sensor j between two arbitrary objects:

$$\begin{aligned}D_{j;i_\alpha,i_\beta} \equiv\ & (\hat{\boldsymbol{z}}_{j;i_\alpha}(t+1|t) - \hat{\boldsymbol{z}}_{j;i_\beta}(t+1|t))^T \\ & \cdot(S_{j;i_\alpha} + S_{j;i_\beta})^{-1} \\ & \cdot(\hat{\boldsymbol{z}}_{j;i_\alpha}(t+1|t) - \hat{\boldsymbol{z}}_{j;i_\beta}(t+1|t)) \\ & (i_\alpha, i_\beta = 1,\ldots,N; i_\alpha \neq i_\beta)\end{aligned} \tag{15}$$

must be large. A good sensor location is defined as one which makes this distance large. When an object i_α or i_β is out of range of sensor j, $D_{j;i_\alpha,i_\beta}$ is assumed not to be obtained. In this paper, considering that the selectivity decreases suddenly when the distance becomes small, we use

$$\mathcal{E}^a(\boldsymbol{x}_s) \equiv -\sum_{j=1}^{M} 1/D_j \quad \text{with } D_j \equiv \min_{i_\alpha,i_\beta} D_{j;i_\alpha,i_\beta} \tag{16}$$

as the evaluation function of the accuracy of association of the total system. When no $D_{j;i_\alpha,i_\beta}$ is obtained, a small positive value is used as D_j. The larger $\mathcal{E}^a(\boldsymbol{x}_s)$ is, the better the association is.

C. Definition of Evaluation Function

We want to define an evaluation function which yields accuracy in both the estimates:

$$\mathcal{E}_i^p(\boldsymbol{x}_s) \equiv 1/\mathcal{R}_i(\boldsymbol{x}_s) \equiv \det P_i^{-1}(t+1|t+1;\boldsymbol{x}_s) \quad (i=1,\ldots,N) \tag{17}$$

and the association. For this purpose, first, we clarify the procedure for computing $\mathcal{E}_i^p(\boldsymbol{x}_s)$ and $\mathcal{E}^a(\boldsymbol{x}_s)$ when there are several sensors and objects.

i) For the sensors at the candidate location $\boldsymbol{x}_s$, compute $\hat{\boldsymbol{z}}_{j;i}(t+1|t)\,(i=1,\ldots,N;\ j=1,\ldots,M)$,

ii) Using $\hat{\boldsymbol{z}}_{j;i}(t+1|t)$ above, evaluate $\mathcal{E}^a(\boldsymbol{x}_s)$,

iii) For each i $(i=1,\ldots,N)$, evaluate $\mathcal{E}_i^p(\boldsymbol{x}_s)$ using sensors which obtain $\hat{\boldsymbol{z}}_{j;i}(t+1|t)$ in their range.

The evaluation function representing the error of the total system is

$$\mathcal{R}_{\mathrm{All}}(\boldsymbol{x}_s) \equiv \sum_{i=1}^{N} \mathcal{R}_i(\boldsymbol{x}_s), \tag{18}$$

and minimizing this function corresponds to maximizing

$$\mathcal{E}_{\mathrm{All}}^p(\boldsymbol{x}_s) \equiv \frac{1}{\mathcal{R}_{\mathrm{All}}(\boldsymbol{x}_s)} = \frac{1}{\sum_{i=1}^{N} 1/\mathcal{E}_i^p(\boldsymbol{x}_s)}. \tag{19}$$

Using this $\mathcal{E}_{\mathrm{All}}^p$ and $\mathcal{E}^a(\boldsymbol{x}_s)$, we define the evaluation function which yields accuracy in both the estimates and the association as

$$\mathcal{I}(\boldsymbol{x}_s) \equiv \mathcal{E}_{\mathrm{All}}^p(\boldsymbol{x}_s) + \lambda \mathcal{E}^a(\boldsymbol{x}_s), \tag{20}$$

where λ is an appropriate positive value. The sensor locations which maximize $\mathcal{I}(\boldsymbol{x}_s)$ are defined as desired sensor locations. Locations maximizing $\mathcal{E}_{\mathrm{All}}^p$ and ones maximizing $\mathcal{E}^a$ may not agree. In this case, the location which maximizes $\mathcal{E}_{\mathrm{All}}^p$ among locations allowing correct association should be selected. The value of λ determines whether obtaining estimates with small errors or associating correctly is focused on, and to what extent they are weighted. However, because it is necessary to know the values of $\mathcal{E}_{\mathrm{All}}^p$ and $\mathcal{E}^a$ to set λ appropriately, we cannot establish a good algorithm to set λ before observations.

III. Algorithm

A. One Object

First, let us consider the case of only one object ($N=1$). The accuracy of association $\mathcal{E}^a$ need not be considered because there is only one object. Even in the case of only one object, calculation is tedious when there are many sensors or $\boldsymbol{x}_s$ is of a high dimension.

Now, let us consider $\boldsymbol{x}, \boldsymbol{h}_j(\cdot), \boldsymbol{z}_j$ and R_j corresponding to each sensor j. In this case, because there is only one object, the subscript i is omitted. When no correlation of observation noise is assumed, we can write

$$\begin{cases} \boldsymbol{x}_s = [\boldsymbol{x}_{s1}^T, \boldsymbol{x}_{s2}^T, \ldots, \boldsymbol{x}_{sM}^T]^T \\ \dfrac{\partial \boldsymbol{h}}{\partial \boldsymbol{x}} = [\dfrac{\partial \boldsymbol{h}_1}{\partial \boldsymbol{x}}^T, \dfrac{\partial \boldsymbol{h}_2}{\partial \boldsymbol{x}}^T, \cdots, \dfrac{\partial \boldsymbol{h}_M}{\partial \boldsymbol{x}}^T]^T \\ \boldsymbol{z} = [\boldsymbol{z}_1^T, \boldsymbol{z}_2^T, \ldots, \boldsymbol{z}_M^T]^T \\ R = \mathrm{diag}(R_1, R_2, \cdots, R_M). \end{cases} \tag{21}$$

When

$$\bar{P}_j^{-1}(t+1) \equiv \left(\frac{\partial \boldsymbol{h}_j}{\partial \boldsymbol{x}}|_{\hat{\boldsymbol{x}}(t+1|t)}\right)^T R_j^{-1}(t+1) \frac{\partial \boldsymbol{h}_j}{\partial \boldsymbol{x}}|_{\hat{\boldsymbol{x}}(t+1|t)}, \tag{22}$$

(6) can be rewritten as

$$P^{-1}(t+1|t+1) = P^{-1}(t+1|t) + \sum_{j=1}^{M} \bar{P}_j^{-1}(t+1). \tag{23}$$

As a first step, considering only sensor 1, we select sensor location $\boldsymbol{x}_{s1}$ using the evaluation function

$$\mathcal{E}^p(\boldsymbol{x}_{s1}) \equiv \det P^{-1}(t+1|t+1;\boldsymbol{x}_{s1}), \tag{24}$$

where

$$P^{-1}(t+1|t+1;\boldsymbol{x}_{s1}) = P^{-1}(t+1|t) + \bar{P}_1^{-1}(t+1;\boldsymbol{x}_{s1}). \tag{25}$$

Next, using the fixed $\boldsymbol{x}_{s1}$ and the evaluation function $\mathcal{E}^p(\boldsymbol{x}_{s1},\boldsymbol{x}_{s2})$ defined by using

$$P^{-1}(t+1|t+1;\boldsymbol{x}_{s1},\boldsymbol{x}_{s2}) = P^{-1}(t+1|t) + \bar{P}_1^{-1}(t+1;\boldsymbol{x}_{s1}) + \bar{P}_2^{-1}(t+1;\boldsymbol{x}_{s2}), \tag{26}$$

we select $\boldsymbol{x}_{s2}$. Iterating this procedure, we can acquire a fairly accurate location for $\boldsymbol{x}_s$ (greedy algorithm).

B. Multiple Objects

Next, we discuss the case of multiple objects. In this case, the accuracy of association $\mathcal{E}^a$ must be considered. In the algorithm, sensor locations are fixed one by one

similarly to the algorithm of only one object. Considering that there are multiple objects, we can write the following, similar to (23):

$$P_i^{-1}(t+1|t+1) = P_i^{-1}(t+1|t) + \sum_{j=1}^{M} \bar{P}_{i;j}^{-1}(t+1). \quad (27)$$

As a first step, considering only the sensor 1, we select desired sensor location $\boldsymbol{x}_{s1}$ using the evaluation function

$$\mathcal{I}(\boldsymbol{x}_{s1}) = \left(\sum_{i=1}^{N} \frac{1}{\det P_i^{-1}(t+1|t+1;\boldsymbol{x}_{s1})} \right)^{-1} - \lambda \frac{1}{D_1(\boldsymbol{x}_{s1})}, \quad (28)$$

where

$$P_i^{-1}(t+1|t+1;\boldsymbol{x}_{s1}) = P_i^{-1}(t+1|t) + \bar{P}_{i;1}^{-1}(t+1;\boldsymbol{x}_{s1}). \quad (29)$$

Next, using fixed $\boldsymbol{x}_{s1}$, we select $\boldsymbol{x}_{s2}$ from

$$\mathcal{I}(\boldsymbol{x}_{s1},\boldsymbol{x}_{s2}) = \left(\sum_{i=1}^{N} \frac{1}{\det P_i^{-1}(t+1|t+1;\boldsymbol{x}_{s1},\boldsymbol{x}_{s2})} \right)^{-1} - \lambda \left(\frac{1}{D_1(\boldsymbol{x}_{s1})} + \frac{1}{D_2(\boldsymbol{x}_{s2})} \right), \quad (30)$$

where

$$P_i^{-1}(t+1|t+1;\boldsymbol{x}_{s1},\boldsymbol{x}_{s2}) = P_i^{-1}(t+1|t) + \bar{P}_{i;1}^{-1}(t+1;\boldsymbol{x}_{s1}) + \bar{P}_{i;2}^{-1}(t+1;\boldsymbol{x}_{s2}). \quad (31)$$

Iterating this procedure, we can acquire a fairly accurate location for $\boldsymbol{x}_s$.

C. Method of Fusion

As the method to fuse information from spatially separated sensors, we can obtain the following equation transforming (5) for each object:

$$\begin{aligned} &\hat{\boldsymbol{x}}_i(t+1|t+1) \\ &= P_i(t+1|t+1)[P_i^{-1}(t+1|t)\hat{\boldsymbol{x}}_i(t+1|t) \\ &\quad + \sum_{j=1}^{N} \left(\frac{\partial \boldsymbol{h}_j}{\partial \boldsymbol{x}}|_{\hat{\boldsymbol{x}}_i(t+1|t)} \right)^T R_{j;i}^{-1}(t+1)\{\boldsymbol{z}_{j;i}(t+1) \\ &\quad - \boldsymbol{h}_j(\hat{\boldsymbol{x}}_i(t+1|t)) + \frac{\partial \boldsymbol{h}_j}{\partial \boldsymbol{x}}|_{\hat{\boldsymbol{x}}_i(t+1|t)} \hat{\boldsymbol{x}}_i(t+1|t)\}]. \end{aligned} \quad (32)$$

This is the same kind of equation as that for a parallel Kalman filter.

IV. Application 1: Target Tracking

A simulation of two cameras mounted on the tips of manipulators (three degrees of freedom) cooperatively tracking two objects on a two-dimensional plane surrounded by walls has been performed (Fig. 1). Since there are two objects, the evaluation function $\mathcal{I}$ defined by (20) is used.

$\boldsymbol{x}_i = [x_i, y_i, \dot{x}_i, \dot{y}_i]^T$ is used as the state vector of the object $i\,(i = 1, 2)$, where (x_i, y_i) is the position of the object normalized so that the length of a side of the wall is one. The state transition equation corresponding to (10) with

$$F_i = \begin{bmatrix} 1 & & \Delta t & \\ & 1 & & \Delta t \\ & & 0.98 & \\ & & & 0.98 \end{bmatrix} \quad (33)$$

is used, where Δt is the unit of discretization of time. When an object runs into a wall, it is assumed to be reflected. The observation equation corresponding to (11) is

$$\boldsymbol{h}_j(\boldsymbol{x}(t),t) + \boldsymbol{v}_{j;i}(t) \equiv -\frac{f\xi_{j;i}}{\eta_{j;i}} + [C_1 + C_2\{\frac{f\xi_{j;i}}{\eta_{j;i}}\}^2]\bar{v}, \quad (34)$$

where $\bar{v}$ is the zero mean noise of which variance is a unit matrix, and $(\xi_{j;i}, \eta_{j;i})$ are the coordinates of object i in camera j's coordinate system. f is the focal length of the cameras, and C_1 and C_2 are positive constants. In this equation, it is shown that the noise becomes large when the output is far from the center of the screen. When an object is behind the camera, the output is assumed not to be obtained. It is assumed that one pitch of each joint of the manipulators is 5 degrees and maximum movement in one time step is 4 pitches.

Fig. 2 shows the error of the estimated position $\hat{\boldsymbol{x}}(t|t)$:

$$d(t) \equiv \sum_{i=1}^{2} \{(x_i(t) - \hat{x}_i(t|t))^2 + (y_i(t) - \hat{y}_i(t|t))^2\}^{\frac{1}{2}}, \quad (35)$$

when the cameras are fixed or moving. This figure shows that, by using this active sensing method, the error $d(t)$

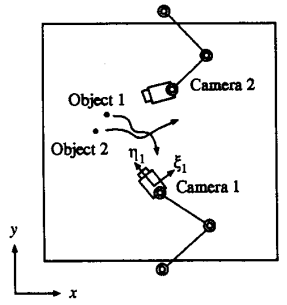

Fig. 1 Hand-eye camera tracking system

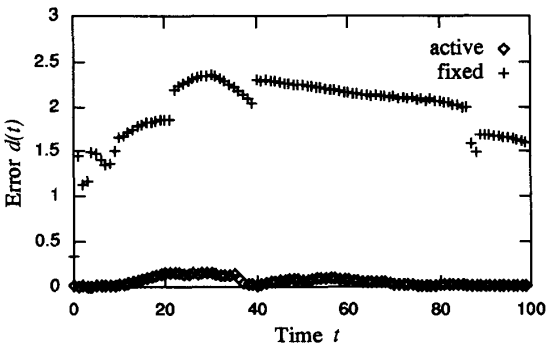

Fig. 2 Results of simulation

becomes small because the hand-eye cameras move to optimal locations according to the object positions. Furthermore, it was observed that when two objects are close, both cameras are located at the locations where both objects are within the ranges, whereas when the objects are far apart, each camera follows a different object. These results show that an appropriate sensing strategy is formulated according to the states of the objects.

V. Application 2: Visual and Tactile Fusion

In this section, we consider the problem of obtaining position and orientation of a three-dimensional object by visual and tactile fusion. The sensor system is assumed to have a three-dimensional geometrical model of the object. It consists of a fixed camera and a tactile sensor mounted on the tip of a manipulator. The camera is used to obtain rough position and orientation and, using this information, sites for contact with the tactile sensor are determined. The method described in Section II can be used to accomplish this by considering the geometrical model.

First let us define a rotation. We denote the rotation ϕ_x around x axis, ϕ_y around y axis and ϕ_z around z axis in turn by $\boldsymbol{\phi} = [\phi_x, \phi_y, \phi_z]^T$. The matrix corresponding to this rotation is

$$R(\boldsymbol{\phi}) = R_z(\phi_z)R_y(\phi_y)R_x(\phi_x) \tag{36}$$

$$R_z(\phi_z) = \begin{bmatrix} \cos\phi_z & -\sin\phi_z & 0 \\ \sin\phi_z & \cos\phi_z & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$R_y(\phi_y) = \begin{bmatrix} \cos\phi_y & 0 & \sin\phi_y \\ 0 & 1 & 0 \\ -\sin\phi_y & 0 & \cos\phi_y \end{bmatrix}$$

$$R_x(\phi_x) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\phi_x & -\sin\phi_x \\ 0 & \sin\phi_x & \cos\phi_x \end{bmatrix}.$$

The three-dimensional model of the object is defined using points and segments of lines in the model coordinate system in which the origin is at the representative point O of the object. The position of the object is denoted by X_O and $\boldsymbol{\phi}_O$, where $\boldsymbol{X}_O = [X_O, Y_O, Z_O]^T$ is the position of the representative point O and $\boldsymbol{\phi}_O = [\phi_{Ox}, \phi_{Oy}, \phi_{Oz}]^T$ is the rotation of the object around the point O in the world coordinate system. For simplicity, $\boldsymbol{X}_O$ and $\boldsymbol{\phi}_O$ are assumed to be constant with time. Coordinates $\boldsymbol{x}_W$ in the world coordinate system corresponding to coordinates $\boldsymbol{x}_M$ in the model coordinate system are

$$\boldsymbol{x}_W(\boldsymbol{X}_O, \boldsymbol{\phi}_O; \boldsymbol{x}_M) = R(\boldsymbol{\phi}_O)\boldsymbol{x}_M + \boldsymbol{X}_O. \tag{37}$$

First, we define the observation with the camera. For simplicity, it is assumed that the center of the camera is at the origin of the world coordinate system and the optical axis corresponds to the z axis. Using a pinhole camera model, we denote the observation by

$$\boldsymbol{x}_f = \boldsymbol{h}_c(\boldsymbol{x}_W) + \boldsymbol{v}_c = \begin{bmatrix} x_W f/z_W \\ y_W f/z_W \end{bmatrix} + \boldsymbol{v}_c, \tag{38}$$

where $\boldsymbol{x}_f$ are the coordinates of the observed point in the screen coordinate system, f is the focal length and $\boldsymbol{v}_c$ is the observation noise with variance R_c. Substituting (37) into (38), we obtain

$$\begin{aligned} \boldsymbol{x}_f &= \boldsymbol{h}_c(R(\boldsymbol{\phi}_O)\boldsymbol{x}_M + \boldsymbol{X}_O) + \boldsymbol{v}_c \\ &= \tilde{\boldsymbol{h}}_c(\boldsymbol{X}_O, \boldsymbol{\phi}_O; \boldsymbol{x}_M) + \boldsymbol{v}_c. \end{aligned} \tag{39}$$

In observation with the camera, we detect characteristic points from the image and select the model points corresponding to these characteristic points. Using these model points, the variables $\boldsymbol{X}_O$ and $\boldsymbol{\phi}_O$ can be estimated. Here, we suppose that the model point with coordinates $\boldsymbol{x}_M$ corresponding to the screen coordinates $\boldsymbol{x}_f$ can be determined easily and we use the Kalman filter to estimate $\boldsymbol{X}_O$ and $\boldsymbol{\phi}_O$. Equation (39) is the observation equation corresponding to (2), and $\boldsymbol{X}_O$ and $\boldsymbol{\phi}_O$ are the state vectors corresponding to $\boldsymbol{x}$ in (2).

Next, let us consider the tactile sensor mounted on the tip of the manipulator. The tactile sensor obtains coordinates of points or parameters of edges of the object in the world coordinate system. First, let us consider the observation of points. We define the observation with the

tactile sensor of the point with world coordinates $\boldsymbol{x}_W$ corresponding to model coordinates $\boldsymbol{x}_M$ as

$$\boldsymbol{x}_T = \boldsymbol{x}_W(\boldsymbol{X}_O, \boldsymbol{\phi}_O; \boldsymbol{x}_M) + \boldsymbol{v}_t, \tag{40}$$

where $\boldsymbol{v}_t$ is the observation noise with variance R_t. This is the observation equation of the Kalman filter, and $\boldsymbol{X}_O$ and $\boldsymbol{\phi}_O$ are the variables to be estimated.

Next, let us consider the situation where the tactile sensor is in contact with an edge of the object. In this case, it is assumed that we can obtain the parameters of the edge and we can easily determine the edge in the model corresponding to this edge. From this relation, the position and orientation, $\boldsymbol{X}_O$ and $\boldsymbol{\phi}_O$, are estimated.

An edge (which is not perpendicular to the z axis) is denoted by

$$\begin{cases} x_M = a_M z_M + p_M \\ y_M = b_M z_M + q_M. \end{cases} \tag{41}$$

Now, let us acquire the representation of this edge in the world coordinate system. Let $R(\boldsymbol{\phi}_O) = (r_{ij})$. Transforming a point on the edge represented by (41) by $\boldsymbol{X}_O$ and $\boldsymbol{\phi}_O$, we obtain

$$\begin{cases} x_W = r_{11}(a_M z_M + p_M) + r_{12}(b_M z_M + q_M) \\ \qquad\qquad + r_{13} z_M + X_O \\ y_W = r_{21}(a_M z_M + p_M) + r_{22}(b_M z_M + q_M) \\ \qquad\qquad + r_{23} z_M + Y_O \\ z_W = r_{31}(a_M z_M + p_M) + r_{32}(b_M z_M + q_M) \\ \qquad\qquad + r_{33} z_M + Z_O. \end{cases} \tag{42}$$

Substituting

$$z_M = \frac{z_W - r_{31}p_M - r_{32}q_M - Z_O}{r_{31}a_M + r_{32}b_M + r_{33}}, \tag{43}$$

obtained by transforming the third equation of (42), into the first and second equations, we obtain

$$\begin{cases} x_W = a_W z_W + p_W \\ y_W = b_W z_W + q_W, \end{cases} \tag{44}$$

where

$$\begin{cases} a_W = \dfrac{r_{11}a_M + r_{12}b_M + r_{13}}{r_{31}a_M + r_{32}b_M + r_{33}} \\ b_W = \dfrac{r_{21}a_M + r_{22}b_M + r_{23}}{r_{31}a_M + r_{32}b_M + r_{33}} \\ p_W = r_{11}p_M + r_{12}q_M + X_O \\ \quad - \dfrac{r_{31}p_M + r_{32}q_M + Z_O}{r_{31}a_M + r_{32}b_M + r_{33}}(r_{11}a_M + r_{12}b_M + r_{13}) \\ q_W = r_{21}p_M + r_{22}q_M + Y_O \\ \quad - \dfrac{r_{31}p_M + r_{32}q_M + Z_O}{r_{31}a_M + r_{32}b_M + r_{33}}(r_{21}a_M + r_{22}b_M + r_{23}). \end{cases} \tag{45}$$

This is the equation of the edge (41) transformed by $\boldsymbol{X}_O$ and $\boldsymbol{\phi}_O$. We denote this relation as

$$\boldsymbol{p}_W = \boldsymbol{\xi}(\boldsymbol{X}_O, \boldsymbol{\phi}_O; \boldsymbol{p}_M), \tag{46}$$

where $\boldsymbol{p}_W = [a_W, b_W, p_W, q_W]^T$ and $\boldsymbol{p}_M = [a_M, b_M, p_M, q_M]^T$.

We assume that the tactile sensor can obtain parameters of edges in the world coordinate system. The parameter is denoted by $\boldsymbol{p}_T = [a_T, b_T, p_T, q_T]^T$. Although we can also use the Kalman filter to obtain the parameter from distribution of points of the edge, we omit the procedure for simplicity. The observation is defined by

$$\boldsymbol{p}_T = \boldsymbol{p}_W + \boldsymbol{v}_T = \boldsymbol{\xi}(\boldsymbol{X}_O, \boldsymbol{\phi}_O; \boldsymbol{p}_M) + \boldsymbol{v}_T, \tag{47}$$

where $\boldsymbol{v}_T$ is the observation noise with variance R_T. This is the observation equation of an edge, and $\boldsymbol{X}_O$ and $\boldsymbol{\phi}_O$ are variables to be estimated.

In the explanation above, we showed that the Kalman filter with the model can be used for estimating the position and orientation of an object. Equations (39), (40) and (47) denote observation. Equations (40) and (47) change according to which point with coordinates $\boldsymbol{x}_M$ or which edge with parameter $\boldsymbol{p}_M$ is sensed by the tactile sensor. Applying the method of sensor location selection described in Section II to these equations, we can select optimal points or edges for contact with the tactile sensor after rough sensing with the camera.

Specifically, defining $\boldsymbol{Y}_O = [\boldsymbol{X}_O^T, \boldsymbol{\phi}_O^T]^T$ and its variance as $P_{\boldsymbol{Y}_O}$, we can obtain the following by applying the Kalman filter to (39):

$$\begin{aligned} \hat{\boldsymbol{Y}}_O(t+1|t+1) &= \hat{\boldsymbol{Y}}_O(t|t) + W_1(t+1) \\ &\quad \cdot[\boldsymbol{x}_f(t+1) - \tilde{\boldsymbol{h}}_c(\hat{\boldsymbol{Y}}_O(t|t); \boldsymbol{x}_M(t+1))] \end{aligned} \tag{48}$$

$$\begin{aligned} P^{-1}_{\boldsymbol{Y}_O}(t+1|t+1) &= P^{-1}_{\boldsymbol{Y}_O}(t|t) \\ &\quad + \left(\frac{\partial \tilde{\boldsymbol{h}}_c}{\partial \boldsymbol{Y}_O}\Big|_{\hat{\boldsymbol{Y}}_O(t|t); \boldsymbol{x}_M(t+1)}\right)^T \\ &\quad \cdot R_c^{-1}(t+1)\frac{\partial \tilde{\boldsymbol{h}}_c}{\partial \boldsymbol{Y}_O}\Big|_{\hat{\boldsymbol{Y}}_O(t|t); \boldsymbol{x}_M(t+1)}, \end{aligned} \tag{49}$$

where

$$\begin{aligned} W_1(t+1) &= P_{\boldsymbol{Y}_O}(t+1|t+1) \\ &\quad \cdot \left(\frac{\partial \tilde{\boldsymbol{h}}_c}{\partial \boldsymbol{Y}_O}\Big|_{\hat{\boldsymbol{Y}}_O(t|t); \boldsymbol{x}_M(t+1)}\right)^T R_c^{-1}(t+1). \end{aligned} \tag{50}$$

Note that, here, t is an index of characteristic points and does not correspond to time. $\boldsymbol{x}_M(t+1)$ in (48) are the model coordinates corresponding to the observed point $\boldsymbol{x}_f(t+1)$. The Jacobians in these equations are calculated as

$$\frac{\partial \tilde{\boldsymbol{h}}_c}{\partial \boldsymbol{X}_O} = \frac{\partial \boldsymbol{h}_c}{\partial \boldsymbol{x}_W}\frac{\partial \boldsymbol{x}_W}{\partial \boldsymbol{X}_O} = \frac{\partial \boldsymbol{h}_c}{\partial \boldsymbol{x}_W} \tag{51}$$

$$\frac{\partial \tilde{\boldsymbol{h}}_c}{\partial \phi_k} = \frac{\partial \boldsymbol{h}_c}{\partial \boldsymbol{x}_W}\frac{\partial \boldsymbol{x}_W}{\partial \phi_k} = \frac{\partial \boldsymbol{h}_c}{\partial \boldsymbol{x}_W}\frac{\partial R(\boldsymbol{\phi}_O)}{\partial \phi_k}\boldsymbol{x}_M, \quad (52)$$

where $k = x, y$ or z and

$$\frac{\partial \boldsymbol{h}_c}{\partial \boldsymbol{x}_W} = \begin{bmatrix} f/z_W & 0 & -x_W f/z_W^2 \\ 0 & f/z_W & -y_W f/z_W^2 \end{bmatrix}. \quad (53)$$

Using these equations with the camera output $\boldsymbol{x}_f$, we can obtain rough position and orientation of the object. Similar equations can be obtained by applying the Kalman filter algorithm to (40) and (47). The Kalman filter for the observation of $\boldsymbol{x}_T$ is

$$\hat{\boldsymbol{Y}}_O(t+1|t+1) = \hat{\boldsymbol{Y}}_O(t|t) + W_2(t+1) \cdot [\boldsymbol{x}_T(t+1) - \boldsymbol{x}_W(\hat{\boldsymbol{Y}}_O(t|t); \boldsymbol{x}_M(t+1))] \quad (54)$$

$$P^{-1}_{\boldsymbol{Y}_O}(t+1|t+1) = P^{-1}_{\boldsymbol{Y}_O}(t|t) + \left(\frac{\partial \boldsymbol{x}_W}{\partial \boldsymbol{Y}_O}|_{\hat{\boldsymbol{Y}}_{O(t|t)}; \boldsymbol{x}_M(t+1)}\right)^T \cdot R_t^{-1}(t+1)\frac{\partial \boldsymbol{x}_W}{\partial \boldsymbol{Y}_O}|_{\hat{\boldsymbol{Y}}_{O(t|t)}; \boldsymbol{x}_M(t+1)} \quad (55)$$

$$W_2(t+1) = P_{\boldsymbol{Y}_O}(t+1|t+1) \cdot \left(\frac{\partial \boldsymbol{x}_W}{\partial \boldsymbol{Y}_O}|_{\hat{\boldsymbol{Y}}_{O(t|t)}; \boldsymbol{x}_M(t+1)}\right)^T R_t^{-1}(t+1), \quad (56)$$

and the Kalman filter for observation of $\boldsymbol{p}_T$ is

$$\hat{\boldsymbol{Y}}_O(t+1|t+1) = \hat{\boldsymbol{Y}}_O(t|t) + W_3(t+1) \cdot [\boldsymbol{p}_T(t+1) - \boldsymbol{\xi}(\hat{\boldsymbol{Y}}_O(t|t); \boldsymbol{p}_M(t+1))] \quad (57)$$

$$P^{-1}_{\boldsymbol{Y}_O}(t+1|t+1) = P^{-1}_{\boldsymbol{Y}_O}(t|t) + \left(\frac{\partial \boldsymbol{\xi}}{\partial \boldsymbol{Y}_O}|_{\hat{\boldsymbol{Y}}_{O(t|t)}; \boldsymbol{p}_M(t+1)}\right)^T \cdot R_T^{-1}(t+1)\frac{\partial \boldsymbol{\xi}}{\partial \boldsymbol{Y}_O}|_{\hat{\boldsymbol{Y}}_{O(t|t)}; \boldsymbol{p}_M(t+1)} \quad (58)$$

$$W_3(t+1) = P_{\boldsymbol{Y}_O}(t+1|t+1) \cdot \left(\frac{\partial \boldsymbol{\xi}}{\partial \boldsymbol{Y}_O}|_{\hat{\boldsymbol{Y}}_{O(t|t)}; \boldsymbol{x}_M(t+1)}\right)^T R_T^{-1}(t+1). \quad (59)$$

Equations (55) and (58) depend on $\boldsymbol{x}_M(t+1)$ and $\boldsymbol{p}_M(t+1)$. Thus, to maximize

$$\mathcal{E}^p = \det P^{-1}_{\boldsymbol{Y}_O}(t+1|t+1), \quad (60)$$

we select $\boldsymbol{x}_M(t+1)$ and/or $\boldsymbol{p}_M(t+1)$. This yields the sites for contact with the tactile sensor. Using the outputs of the tactile sensor, we can estimate more precise $\boldsymbol{Y}_O$ using (54)~(59). This is the method by which to obtain the position $\boldsymbol{X}_O$ and orientation $\boldsymbol{\phi}_O$ efficiently.

VI. Conclusion

An active sensing method in sensor fusion was proposed.

First, the method to obtain sensor locations for obtaining estimates with small error was proposed. In this method, not only the accuracy of the estimates, but also suitability for association of sensor data is considered. Furthermore, an algorithm to calculate the proposed method was described.

Next, as examples, the method was applied to a multitarget tracking system with multiple sensors, and visual and tactile fusion using a camera and a tactile sensor.

By using this method, one can acquire sensor locations for the object of measurement. In other words, one can tailor the sensing strategy to one's purposes.

References

[1] M.Ishikawa, "Sensor fusion : The state of the art", *Journal of Robotics and Mechatronics*, vol. 2-4, pp.235–244, 1991.

[2] R.C. Luo and M.G. Kay, "Multisensor integration and fusion in intelligent system", *IEEE Trans. Syst. , Man, Cybern.*, vol. 19-5, pp. 901-931, 1989.

[3] S.A. Shafer, A. Stentz, C.E. Thorpe, "An architecture for sensor fusion in a mobile robot", *Proc. IEEE Int. Conf. on Robotics and Automation*, pp. 2002-2011, 1986.

[4] P.K. Allen, *Robotic Object Recognition Using Vision and Touch*, Kluwer, 1987.

[5] M. Shimojo and M. Ishikawa, " An active touch sensing method using a spatial filtering tactile sensor", *Proc. 1993 IEEE Int. Conf. on Robotics and Automation*, vol. 1, pp. 948-954, 1993.

[6] J. Aloimonos, I. Weiss, and A. Bandyopadhyay, "Active vision", *Int. J. of Computer Vision*, pp. 333-356, 1988.

[7] A. Blake and A. Yuille, *Active Vision*, MIT Press, 1992.

[8] J.J. Gibson, *The Ecological Approach to Visual Perception*, Houghton Mifflin, 1970.

[9] U. Neisser, *Cognition and Reality*, Freeman, 1976.

[10] H.R. Hashemipour, S. Roy, and A. J. Laub, "Decentralized structure for parallel Kalman filtering", *IEEE Trans. Automatic Control*, vol. 33-1, pp. 88-94, 1988.

[11] Y. Dodge, V.V. Fedrov and H. P. Wynn eds., *Optimal design and analysis of experiments*, Elsevier science publishers, 1988.