People tracking by fusing different kinds of sensors, floor sensors and acceleration sensors

Tetsushi Ikeda, Hiroshi Ishiguro, and Takuichi Nishimura

***Abstract*—To realize accurate tracking of people in the environment, many studies have been proposed using vision sensors, floor sensors, and wearable devices. Floor sensors can reliably detect current positions, but it is difficult to estimate correct associations between observations and multiple people. To solve the problem, we propose to combine acceleration sensors that are attached to the human body. Since the signals from floor sensors and acceleration sensors synchronize when they observe same person walking, these signals are not independent. The synchrony between the signals is evaluated based on statistical test to find correct association. People tracking examples are shown to confirm the effectiveness of the proposed method. Significant improvement in correct association rate is achieved compared to the results only by floor sensors.**

I. Introduction

In order to realize intelligent environment that supports human activities, much work on sensor network has been done by integrating many kinds of sensors in the environment and wearable sensors. Accurate and reliable tracking of people and knowing their positions in the environment is one of fundamental problems in sensor networks. The sensors used in previous approaches are classified into three main groups.

(1) Vision sensors

Vision sensors are widely used to understand the scene in the environment, and much works have been done to recognize human behavior using vision sensors [1]. Vision sensors provide much information about people in the environment, not only their positions but shape, color, and gestures. One of the defects is that it is affected by changes in environmental condition such as light and obstacles.

(2) Floor sensors

By spreading touch sensor or pressure sensor network on the floor, the positions of people are accurately detected. Recently, floor sensors have received increasing attention and several studies have been proposed to recognize human behavior using floor sensors [2] [3] [4] [5]. Floor sensors are very reliable, but it does not provide any information to distinguish each person. This characteristic causes ambiguity of association between observations and people.

(3) Wearable devices

T. Ikeda and H. Ishiguro are with the Department of Adaptive Machine Systems, Graduate School of Engineering, Osaka University, Osaka, JAPAN (e-mail: ikeda@ed.ams.eng.osaka-u.ac.jp).

T. Nishimura is with the National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, JAPAN.

Positions of people are detected by using wearable device on the body. Many systems have been proposed using infrared [6], ultrasonic wave [7], RFID [8], and wireless LAN [9]. Since the ID information to distinguish each person is explicitly sent to the system, personal identification during tracking is perfect. However, we have to install many reader devices in the environment to obtain positions of people accurately. And it is desirable that people carry the device in natural manner.

In this paper, we propose to integrate floor sensors and acceleration sensors that are attached to the body. A typical problem to track people with floor sensors is shown in Fig. 1. When two people go across, it is difficult to distinguish the two interpretations only by floor sensor signals.

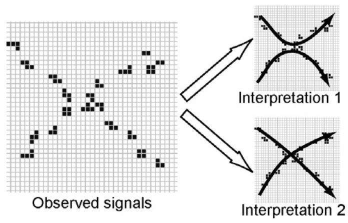


Fig. 1. An ambiguity of the association when walking people are observed by floor sensors.

Since the signal from an acceleration sensor reflects motion of each person and it contains ID information, the association problem is solved effectively. By sending acceleration signal via ratio waves, we can use the information anytime the association problem occurs. The number of the reader device that is required to communicate with acceleration sensors is one. We assume that our method is applied to track people who are carrying cellular phones that have acceleration sensors inside. So we are carrying acceleration sensors naturally and our method works under current information infrastructure.

The problem is how to integrate floor sensors and acceleration sensors in different representations. Many works has been done in the research area of the sensor fusion, and typical approach is to convert each sensory signal to a common representation before integration. For example, to integrate signals from microphones and video cameras, locations of a sound source is estimated in sound and video independently. Then the precise position is computed by integration of estimated locations. However, since acceleration sensors do not have location information, it is difficult to apply previous sensor fusion methods.

In this paper, we propose novel integration method to integrate floor sensors and acceleration sensors based on statistical method. We solve the association problem by evaluating independency between these sensory signals based on statistical test.

In section 2, the sensor fusion algorithm based on a statistical test is described. In section 3, the algorithm to integrate floor sensors and acceleration sensors are described in detail. In section 4, experimental results are shown. In section 5, we conclude the paper.

II. Integration of Different Kinds of Sensors based on the Statistical Test

In this section, we propose a integration method of following two types of sensors:

1. **Position sensor**: Sensors that observes positions of targets. The observations are not labeled with target IDs;
2. **Wearable sensor**: Sensors that is attached to each target. It observes the motion of the target.

A. Multiple target tracking using the position sensors

Since the observations from the position sensor contain only position information and do not contain ID information, ambiguities may arise in estimating trajectories of multiple targets. Tracking multiple targets using the position sensors requires solution to both data association and state estimation problems [10] [11]. The most successful algorithm is the multiple hypothesis tracker (MHT) [12]. MHT generates and maintains a set of hypotheses, where each hypothesis associates past observations with targets in a different way. Each hypothesis is evaluated by its posterior probability and the final result is the hypothesis with the highest probability. However, since MHT postpone decisions and examine all possible combination association as a new set of observation arrives, the number of hypotheses grows exponentially. The growth of the hypotheses is shown in Fig. 2, where each branch denoting a different assignment of an observation to a target.

Though several heuristics are proposed to cope with this problem, it is essentially difficult to select correct association

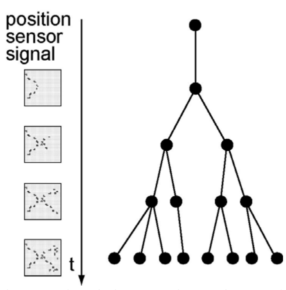


Fig. 2. Exponential growth of the number of association hypotheses in multiple hypotheses tracking. A path from the root node to a leaf node represents an association hypothesis.

only from position sensors. A promising approach is to combine different kinds of sensors.

B. Evaluating association hypotheses by integrating different kinds of sensors

In this paper, we propose to disambiguate the association by integrating wearable sensors. For example, the signals from floor sensors and acceleration sensors on the body change in correlated manner if they observe same person. By evaluating the correlation between these sensors, probable association hypotheses are selected and the number of hypotheses becomes tractable (Fig. 3).

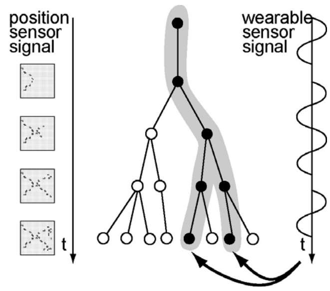


Fig. 3. Evaluation and selection of association hypotheses using correlation between position sensors and wearable sensors.

The problem is how to evaluate correlation between position sensors and wearable sensors. Since the signals are in the different representation, a method that computes correlation between different kinds of sensors is required.

C. Sensor integration based on evaluating synchrony between signals

Position sensors and wearable sensors display synchrony and the signals are not independent if they observe same information source. Recently, several studies of sensor integration have been done by extracting synchrony between the signals from different kinds of sensors based on statistical methods (Fig. 4).

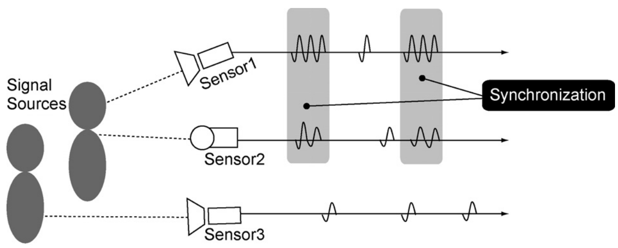


Fig. 4. Sensor integration based on computation of synchrony between sensors.

Hershey et al. [13] observed people speaking alternately with a camera and a microphone. They extracted synchrony between the audio signal and the brightness of the pixel around the speaker's mouse. They localized the speaker in the image by computing mutual information between the signals. This method has extended and has been applied to especially sound source localization problem [14] [15]. A limitation of the method is the assumption that the target does not move in the images. In coping with a moving target, object detection is applied [16] [17].

However, previous statistical sensor integration methods cannot be applied to the case that multiple signal source overlap in the array sensor signals like video cameras and floor sensors. In this paper, we propose to generate multiple hypotheses of the trajectories from the position sensor array based on MHT and select hypotheses by evaluating correlation between position sensors and wearable sensors.

Previous methods have not been applied to binary sensors like floor sensors. We propose to convert both signals into binary representation and apply a statistical test to evaluate correlation.

D. Evaluate synchrony based on the chi-square test of goodness-of-fit

We propose to select hypotheses of trajectories by testing if the acceleration sensor signals on a target is independent from the floor sensor signals that the hypothesis associates. The association problem between signals from different kinds of sensors is described as a hypothesis test [6]. Whether time series x(t) and y(t) observes same information source is decided by a hypothesis test:

$$H_0 : x(t), y(t) \sim p(x)p(y)$$
$$H_1 : x(t), y(t) \sim p(x, y) \quad (1),$$

where H_0 states that the observations are statistically independent and H_1 states dependent.

In the case of testing dependency between discrete signals, the chi-square test of goodness-of-fit is used. When both signals are in binary representation, a two-way contingency table is created (Table. I).

TABLE I. TWO-WAY CONTINGENCY TABLE.

	y = 1	y = 0	Total
x = 1	z_{11}	z_{10}	$z_{1.}$
x = 0	z_{01}	z_{00}	$z_{0.}$
Total	$z_{.1}$	$z_{.0}$	$z_{..}$

Where z_{ij} in the table is the number of the observation (x,y) = (i,j), and

$$z_{i.} = \sum_j z_{ij}, \quad z_{.j} = \sum_i z_{ij}, \quad z_{..} = \sum_{i,j} z_{ij}.$$

When the null hypothesis H_0 is presumed true, the χ^2 value is distributed as chi-square with 1 degree of freedom:

$$\chi^2 = \sum_{i,j} \frac{(z_{ij} - \hat{z}_{ij})^2}{\hat{z}_{ij}}$$

where $\hat{z}_{ij}$ is the theoretical frequency of observing (x,y)=(i,j) asserted by the null hypothesis H_0 and can be expressed as

$$\hat{z}_{ij} = \frac{z_{i.} z_{.j}}{z_{..}}.$$

The proposed algorithm is summarized in Table. II.

TABLE. II
MULTIPLE HYPOTHESES TRACKING ALGORITHM
BY INTEGRATING FLOOR SENSORS AND ACCELERATION SENSORS.

1. As new observations are obtained, generate hypotheses of target trajectories based on the observation of the position sensors.
2. For all generated hypotheses, convert the floor sensor signals associated to each target to binary signals. Convert signals from acceleration sensors to binary signals.
3. For all hypotheses, perform the chi-square test of goodness-of-fit to the binary signals computed in 2.
4. If the pair of signals is considered independent in 3, remove the hypothesis.
5. Go to 1.

III. ALGORITHM TO INTEGRATE A FLOOR SENSOR AND ACCELERATION SENSORS

In this section, the details of the algorithms to integrate floor sensors and acceleration sensors are described.

A. Multiple target tracking using the position sensors

An example of the observations of the floor sensor is shown in Fig. 5. In the figure, the observations in a period are overlapped.

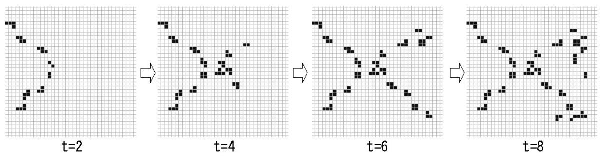


Fig. 5. Examples of floor sensor signals.

The problem of estimating trajectories of multiple people based on floor sensors are:

1. The exponential growth of the number of hypotheses. Keeping all hypotheses is not realistic.
2. Discrete changes of the observation. Since the floor sensors observes the feet of people, only discrete positions of the feet are observed. And the observation of both feet sometimes disappears.

We apply MHT to estimate possible trajectories of the targets. It is difficult to handle huge number of hypotheses, but in the next step it is better to evaluate longer observation to decide dependency between signals. We introduce following models and assumptions.

1) Modeling trajectories based on the Kalman Filter

We use the Kalman Filter to estimate a continuous trajectory from discrete observations. Each target is modeled by a sequence of state vectors. The state vector of the target i at t consist of positions and velocities in two dimensional coordinates:

$$\mathbf{X}_{i,t} = [x, y, \dot{x}, \dot{y}]'.$$

The system model is as follows:

$$\mathbf{X}_{i,t} = \mathbf{F}\mathbf{X}_{i,t-1} + \mathbf{w},$$

where

$$\mathbf{F} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$\mathbf{w}$ is process noise vector.

The observation vector $\mathbf{Z}_{i,t}$ is computed using following observation model:

$$\mathbf{Z}_{i,t} = \mathbf{H}\mathbf{X}_{i,t} + \mathbf{v}$$

where

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix},$$

$\mathbf{v}$ is measurement noise vector. Positive values are set to diagonal elements in $\mathbf{w}$ and $\mathbf{v}$.

A standard update rule is introduced. Only the prediction step of the Kalman Filter is executed when there are no observations that are associated to the target.

2) Association constraint

To reduce number of the association between the observation and the target, we introduce following constraint.

1. Associate all observations that are adjacent to each other to the same target, and associate the succeeding observation at same position to the same target.
2. Limit the observations that are associated to the target to positions whose Mahalanobis Distance between the observation and the estimated position by the Kalman Filter is less than a threshold (*f_md_threshold)* [11].

3) N-Scan Approximation

To limit the number of the hypotheses, we apply n-Scan Approximation [18] to resolve the assignment ambiguity at time t-N (N=*f_scanback_length*) using observation until current time t (Fig. 6). The posterior probability that is computed by the Kalman Filter is used to evaluate

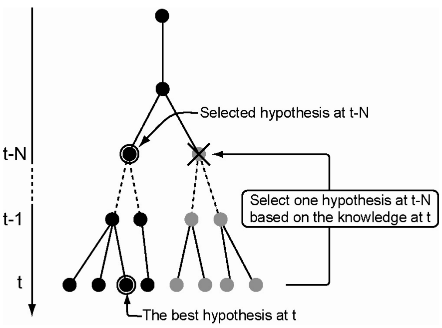


Fig. 6. N-scan approximation.

hypotheses.

B. Extracting the binary signal that represents the time of contact

1) Preprocess floor sensor signals

For each hypothesis, the binary signal that represents the time of contact is computed. This conversion is like a differential processing in one dimensional signal.

Suppose x_t is the center of the current observations and x_{t0} is the center of the previous observations. The binary signal f(t) is defined as follows:

$$f(t) = \begin{cases} 1 \ if \mid x_t - x_{t0} \mid > f\_dist\_threshold \\ 0 \text{ otherwise} \end{cases}.$$

We introduce the threshold since the center of the observation sometimes moves a little after the contact of a foot to floor sensors. The threshold should be larger than the length of the move, and we set the value to about half length of the sole. An example of the extracted binary signal is shown in Fig. 7.

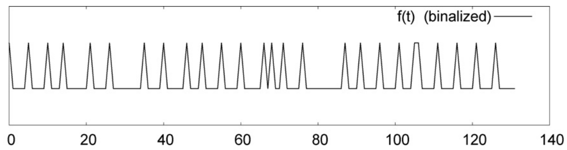


Fig. 7. An example of the extracted binary signal from floor sensor signals that are associated to a target.

2) Extract the binary signal from acceleration sensors

Signals from acceleration signal show almost binary property that is correlated to the walking motion. We convert acceleration signals to binary signal that represents the time of contact (Fig. 8).

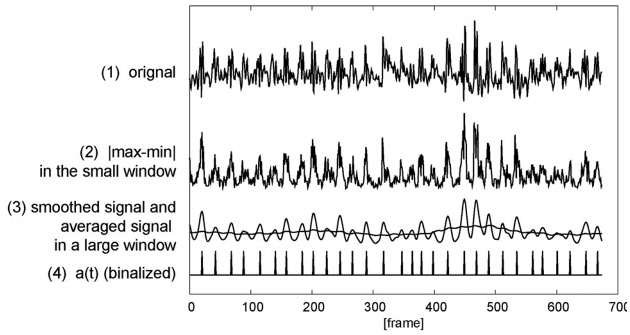


Fig. 8. An example of the extracted binary signal from an acceleration sensor signal.

1. An original observed signal.
2. Extract maximum difference of signals in *a_diff_wlen* frames. The large change of the value is detected.
3. Smooth the signal computed in 2 in a small window (Hanning window, window length = *a_smooth_wlen*). Compute local average signal in a large window (window length = *a_localavg_wlen*)
4. Extract peak of the signal computed in 2 if the value is larger than local average. Set binary signal a(t) as follows:

$$a(t) = \begin{cases} 1 & \text{if there is a peak at time t} \\ 0 & \text{otherwise} \end{cases}$$

C. Evaluating synchrony between sensory signals based on chi-square test of goodness-of-fit

The floor sensor signals and acceleration signals from a person is correlated according to the walking motion. The chi-square test of goodness-of-fit is performed to test if the signals are independent. The length of the signal for the test is set to *correlation_length*.

H0: (*a(t), f(t)*) is independent

H1: (*a(t) ,f(t)*) is not independent,

where *a(t), f(t)* is binary signal that are extracted from acceleration sensor and floor sensors, respectively.

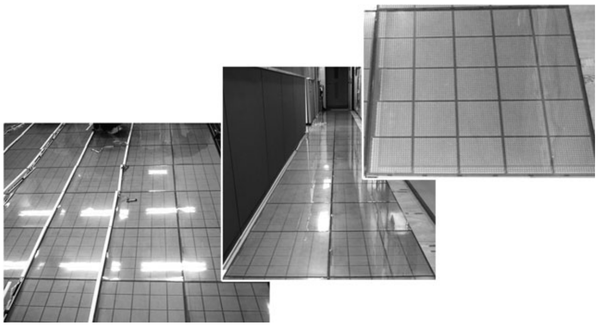
The purpose of the test is to select association hypotheses that (*a(t), f(t)*) is not independent. We set the level of significance α to a large value in the experiment.

IV. EXPERIMENTS

To confirm our approach, we applied the proposed method to track two people in the room.

A. Experimental setup

The floor sensor used in the experiments is VS-SS-SF55 (Vstone Co. [19]) shown in Fig. 9. A carpet is laid on the floor

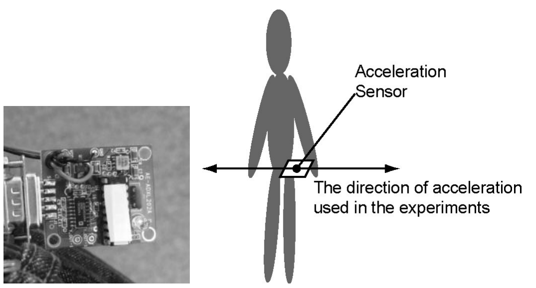


Size of unit sensor	100mm x 100mm
Number of sensors	1400
Output signal	binary
Frame rate	8 Hz

Fig. 9. The floor sensor used in the experiment. The squares in the figure are unit sensors. The frame rate depends on the number of sensors.

sensor in use.

The acceleration sensor used in the experiment is ADXL202 (Analog Devices, Inc.). The sensor is attached to the waist of the body (Fig. 10). The sensor measures acceleration of two axes and the signal of the axis shown in Fig. 10 is used in the experiments.



Output signal	16bit integer
Frame rate	37 Hz

Fig. 10. Acceleration sensor used in the experiments.

We assumed the number of people is known and all people in the environment have an acceleration sensor. The parameters used in the experiment are shown in Table. III.

TABLE III. PARAMETERS USED IN THE EXPERIMENTS.

Parameters in processing floor sensors	
Process noise variance (position)	$(0.04)^2$
Process noise variance (velocity)	$(0.02)^2$
Measurement noise variance	$(1.0)^2$
f_md_thrshold	1.0
f_scanback_length	48
f_dist_threshold	15

Parameters in processing acceleration sensors	
a_diff_wlen	3
a_smooth_wlen	15
a_localavg_wlen	101

Parameters in statistical test	
Level of significance (α)	0.2
correlation_length	128

B. Experimental data

We captured two types of trajectories shown in Fig. 11. The trajectories of two people cross in type 1 ("cross") and do not cross in type 2 ("pass"). Seven type1 data and six type2 data are used in the experiment. The scenes are recorded on a video tape and correct associations are obtained for evaluation.

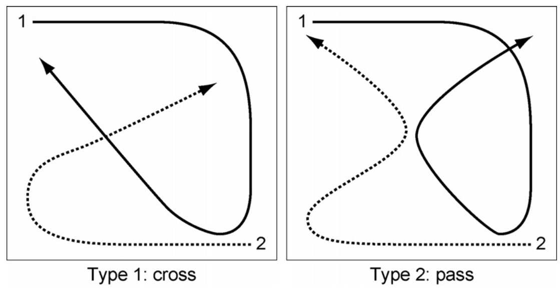


Fig. 11. Two types of trajectories in the experiment.

C. Results

We performed two people tracking experiments based on the proposed method and computed correct association rate between observation of floor sensors and target tracks. For comparison, the results by only floor sensors are computed. The results are shown in Table. IV.

In the type1 (cross) data, the correct association ratio is almost perfect for both methods. We consider the main reason is that the state vector used in the Kalman Filter model includes velocity terms. In the type2 (pass) data, the correct association ratio was increased significantly by integrating acceleration sensors.

Since two people encountered at about center of each captured sequence, the correct association rate is either almost perfect or about 50%. The rate depends on whether correct association was obtained after two people encountered.

TABLE. IV. EXPERIMENTAL RESULTS.
Figures in "floor sensor" column show percent of correct association of the best association only by touch sensor network. Figures in "integrated" column show the result with the proposed integration method.

Data	Type	Correct Association Rate[%]	
		Floor sensor	Integrated
1	cross	100.0	99.5
2		100.0	98.9
3		100.0	99.6
4		100.0	100.0
5		100.0	99.9
6		100.0	99.2
7		100.0	100.0
8	pass	43.5	99.9
9		47.2	99.8
10		45.1	99.5
11		54.8	99.7
12		99.7	99.2
13		56.7	99.4
Success Rate		8/13	13/13

D. Experiment (2)

We captured three people dancing shown in Fig. 12. We again compared tracking results with and without acceleration sensor signals using same parameters shown in Table. III. Without acceleration sensors, the number of hypotheses grows up very quickly and the memory has exhausted. By pruning hypotheses by integrating acceleration sensors, the result was almost correct (correct association

Fig. 12. Experimental data of dancing people.

ratio is 94.1%).

V. CONCLUSION

In this paper, we proposed a novel method to track multiple people in the environment by integrating floor sensors and acceleration sensors that are attached to the human body. Since many cellular phones have an acceleration sensor, the proposed framework is realistic approach. By using only floor sensors, it is difficult to estimate correct associations between observations and target tracks since floor sensors does not provide any ID information. We proposed to evaluate association hypotheses by extracting synchrony between two signals based on statistical test. By selecting association hypotheses based on the evaluation, the correct association hypothesis is estimated.

To confirm the effectiveness of the proposed method, we compared the results by only floor sensors and the results by integrating floor sensors and acceleration sensors. Significant improvement in correct association rate is achieved in the experiment.

In the future, we plan to apply the proposed method to track more than two people who show various behaviors like skipping and dancing.

REFERENCES

[1] W. Hu, T. Tan, L. Wang, and S. Maybank, "A survey on visual surveillance of object motion and behaviors," IEEE Trans. on Systems, Man and Cybernetics, Part C, Vol. 34, No. 3, pp. 334-352, 2004.

[2] R. J. Orr and G. D. Abowd, "The smart floor: A mechanism for natural user identification and tracking," Proc. Conference on Human Factors in Computing Systems (CHI 2000), pp. 303-306, 2000.

[3] H. Morishita, R. Fukui, and T. Sato, "High resolution pressure sensor distributed floor for future human-robot symbiosis environments," Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS'02), pp. 1246-1251, 2002.

[4] T. Murakita, T. Ikeda, and H. Ishiguro, "Human tracking using floor sensors based on the markov chain monte carlo method," Proc.17th Int. Conf. on Pattern Recognition (ICPR 2004), Vol.4, pp. 917-920, 2004.

[5] T. Mori, Y. Suemasu, H. Noguchi and T. Sato, "Multiple people tracking by integrating distributed floor pressure sensors and RFID system," Proc. IEEE Int. Conf. on Systems, Man and Cybernetics, Vol. 6, pp.5271-5278, 2004.

[6] R. Want, A. Hopper, V. Falc a″ o and J. Gibbons, "The active badge location system," ACM Trans. Inf. Syst., Vol. 10, No. 1, pp. 91-102, 1992.

[7] A. Harter, A. Hopper, P. Steggles, A. Ward and P. Webster, "The anatomy of a context-aware application," Proc. 5th Annual ACM/IEEE Int. Conf. on Mobile Computing and Networking (Mobicom '99), 1999.

[8] T. Amemiya, J. Yamashita, K. Hirota and M. Hirose, "Virtual leading blocks for the deaf-blind: A real-time way-finder by verbal-nonverbal hybrid interface and high-density RFID tag space," In Proc. of IEEE Virtual Reality Conference 2004, pp. 165-172, 2004.

[9] P. Bahl and V. N. Padmanabhan, "Radar: An in-building RF-based user location and tracking system," Proc. of IEEE INFOCOM 2000, Vol. 2, pp. 775-784, 2000.

[10] Y. Bar-Shalom and T. E. Fortmann. Tracking and Data Association, Academic Press, 1988.

[11] I. Cox. "A review of statistical data association techniques for motion correspondence," Int. J. of Computer Vision, 10(1):53--66, 1993.

[12] D.B. Reid, "An algorithm for tracking multiple targets," IEEE Trans. Automatic Control, 24(6):843--854, 1979.

[13] J. Hershey, H. Ishiguro, and J. R. Movellan, "Audio vision: Using audio-visual synchrony to locate sounds," Proc. Neural Information Processing Systems (NIPS'99), 1999.

[14] J. W. Fisher III, T. Darrell, W. T. Freeman, and P. Viola, "Learning joint statistical models for audio-visual fusion and segregation," Proc. Neural Information Processing Systems (NIPS'00), 2000.

[15] H. J. Nock, G. Iyengar, and C. Neti, "Assessing face and speech consistency for monologue detection in video," Proc. of ACM International Conf. on Multimedia 2002, pp. 303–306, 2002.

[16] M. Slaney and M. Covell, "Facesync: A linear operator for measuring synchronization of video facial images and audio tracks," Proc. Neural Information Processing Systems (NIPS'00), pp.814-820, 2000.

[17] T. Ikeda, H. Ishiguro, and M. Asada, "Attention to clapping - a direct method for detecting sound source from video and audio -," Proc. IEEE Conf. on Multisensor Fusion and Integration for Intelligent Systems (MFI2003), 2003, pp. 264–268.

[18] T. Kuiren, "Issues in the design of practical multitarget tracking algorithms," In Y. Bar-Shalom ed., Multitarget-Multisensor Tracking: Advanced Applications, pp. 43-87, Artech House Publishers, 1990.

[19] Vstone, Co., http://www.vstone.co.jp/e/etop.html.