Proceedings of
IEEE International Conference on
Multisensor Fusion and Integration for Intelligent Systems
Seoul, Korea, August 20 - 22, 2008

Personal Identification in Dynamic Images using UHF band RFID System for Service Provision

Ichiro Fukuda, Soichiro Morishita, and Hajime Asama

Abstract— **Today, many surveillance cameras are installed in public spaces. This paper presents a proposal of using these cameras for service provision. To do so, it is important to identify "Who" the target of service is and "Where" the target is. For this study, we use a UHF-band RFID system to acquire "Who" information and a surveillance camera to acquire "Where" information. We define the RFID tag's "Reading Probability" to associate "Who" and "Where" information. Additionally, to match up "Who" and "Where" information, we propose and define "Score". Using "Score", we simultaneously acquire information "Who" and "Where".**

I. INTRODUCTION

To examine service provision, many studies of intelligent environments have been conducted: Robotic Room[1], Intelligent Space[2], Smart Room[3], Aware Home[4], and Intelligent Room[5]. These methods require many sensors, sensor networks, robotic arms and so on to provide services such as turning on lights and transporting objects depending on the person's behavior. These methods are suitable when their envisioned environment is a private space, as used in Robotic Room, such as a hospital room or one's own home. However, they require additional space. Therefore, it is difficult to apply them to existing environments, especially for public spaces such as stations. Additionally, such devices are not easy to maintain and are too expensive to adopt because they require many sensors and ubiquitous devices. For those reasons, such systems have not been able to provide value-added services easily.

Obtaining dynamic images from fixed cameras such as surveillance cameras is efficient for sensing wide spaces from environment for service provision. Sensing and analyzing object persons without contact is made possible by obtaining dynamic images and using computer image processing. Therefore, dynamic image recognition methods have gained much attention; various related studies have progressed.

Furthermore, in recent years, numerous surveillance cameras has been increasing worldwide based on social recognition of their usefulness for crime prevention. In Britain, a closed-circuit Television (CCTV) network has been rapidly spreading with the aim of counter-terrorism and crime-fighting since the latter half of the 1990s. A report issued in 2006 states that there might now be 4.2 million CCTV cameras in Britain: one for every 14 people; a person's image can be taken by over 300 cameras in one day[6]. Using such existing devices is advantageous to construct a service provision system from the perspective of cost and feasibility.

We have been studying "service media" for adaptive service provision to walking persons and make it three phases:
(1) Extraction of a moving object, (2) Movement track acquisition, and (3) Presumption of a person's intention. This study addresses (1) and (2). Furthermore, it is important to identify "Who" is the target of service and "Where" the target is for realizing (3): presumption of person's intention. For that reason, we examine association of "Who" and "Where" information.

From the description presented above , it is a purpose of this study to construct an environment to acquire information for adaptive service provision for a walking person. For this discussion, information for adaptive service provision is

1. Information of a person's position or movement track ("Where" information)
2. Information of a person's attributes ("Who" information)

and information of a match-up of "Who" and "Where".

In subsequent chapters, we construct an environment to acquire "Who" and "Where" information simultaneously.

II. PROPOSED METHOD

A. System Outline

Moving object detection using background difference has been studied actively for a person's position identification[7]. For the present study, this method is applied also for moving object detection. It is possible, using this method, to obtain a moving objects position and a moving track ("Where" information). However, it is extremely difficult, using image processing alone, to obtain a person's attributes (e.g., sex, age, native language) and identify people. An RFID tag makes simplifies personal identification. However, commonly used RFID imposes extraordinary operation tasks such as holding RFID tag over an antenna (or scanner) on people. Using a UHF-band RFID system, personal identification is possible merely by wearing an RFID tag around one's

Manuscript received March 27, 2008.
I. Fukuda, S. Morishita, and H. Asama are with RACE, The University of Tokyo, Kashiwanoha 5-1-5, Kashiwa-shi, Chiba 277-8568, JAPAN (e-mail: {zone, mori, asama}@race.u-tokyo.ac.jp).

neck (Fig. 1) as a pendant. We therefore adopt a UHF-band RFID system.

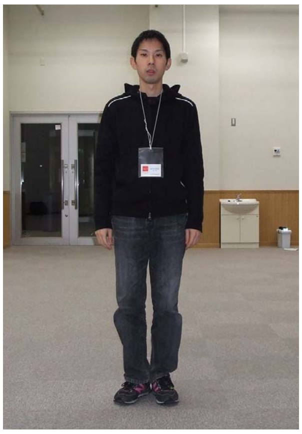

Fig. 1 Image of a person wearing an RFID tag pendant

A moving object's position ("Where") is obtained from image processing; personal attributes ("Who") are obtained by UHF band RFID. These two pieces of information are matched up mutually. Furthermore, we can then produce an environment to obtain "Where" and "Who" information simultaneously.

B. Analysis of UHF band RFID Attribute

A UHF-band RFID can read tags at the range of 3–8 [m], which is a reasonably long range. We carried out preliminary experiments on UHF-band RFID attribute using a UHF-band RFID system (RF-RW002; Mitsubishi Electric Corp.).

A graph of tag-antenna distance and response power when the tag and antenna are in a face-to-face position is portrayed in Fig. 2. When an antenna is installed at Point A or B in the same room, at these two points we obtained response power. This result shows that the response power is ***not*** constant according to the change of environment.

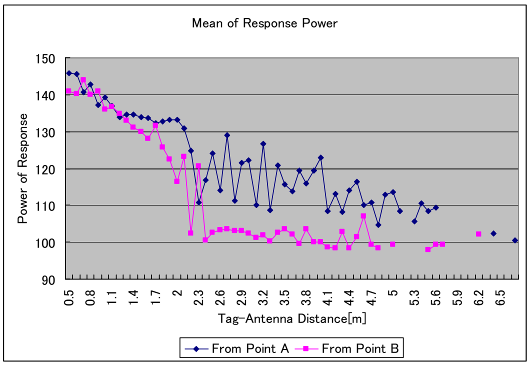

Fig. 2 Distance-Power of Response relation of UHF-band RFID System

Fig. 3 presents a graph of the angle of the tag to the antenna and response power. We confirmed that the response power is greatest when the angle of the tag and antenna are of a face-to-face position (ca. 0 [°]).

The values of Y-axis in Fig. 2 and Fig. 3 have no unit because RFID system trims the power of response values when outputs them and we use trimmed output values.

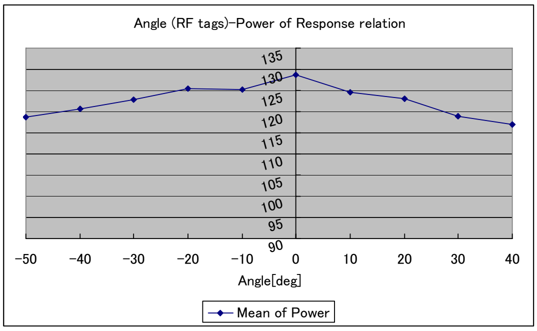

Fig. 3 Angle (RF tag)-Power of response relation

C. Probability Model using Tag Reading information

From results described in the previous section, it seems difficult to identify a moving objects' position using power of response alone. Therefore, we consider a probability model using the tag reading frequency and the angle of the tag and antenna instead of the distance and power of response relation.

The tag reading frequency is calculated by dividing the image area into small rectangular areas.

The tag reading data stream and moving object's location

and angle matrix are described as

$$\boldsymbol{r}_i = (r_{i1}, r_{i2}, \ldots, r_{iT-1}, r_{iT}), \quad r_{it} = \begin{cases} 0 \\ 1 \end{cases} \quad (t = 1, 2, \ldots, T)$$

$$\boldsymbol{Q}_j = \begin{pmatrix} x_{j1} \ldots x_{jT} \\ y_{j1} \ldots y_{jT} \\ \theta_{j1} \ldots \theta_{jT} \end{pmatrix} \quad , \quad (1)$$

where i, j respectively signify a unique IDs of tags and moving objects; T is the data size.

Then we define the tag reading probability as

$$p(x, y, \theta) = \frac{n_{xy\theta}}{N_{xy\theta}}, \quad (2)$$

where x, y is the grid number—the image pixel barycenter of a moving object in the area, and θ is the angle of the moving object to RFID antenna, as calculated from its trajectory (Fig. 4). Furthermore, $n_{xy\theta}$, $N_{xy\theta}$ are determined based on the number of reading data. In addition, $n_{xy\theta}$ is the number of data when $r = 1$; $N_{xy\theta}$ is the number of all data in grid of (x, y) and angle θ .

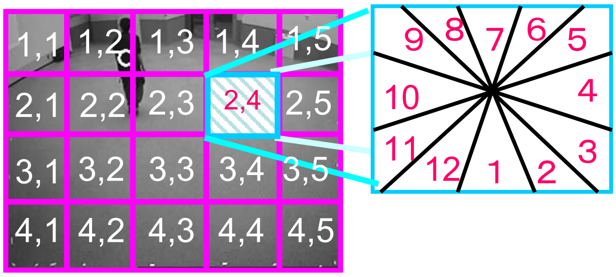

Fig. 4 Rule of dividing the grid and angle grid

First, a person walks around in the intended environment and we obtain tag reading data. Then we calculate the tag reading probability and produce a database.

If we obtain a moving track $\boldsymbol{Q}_j$, the "Score" is defined as follows.

$$S_{ij} = \sum_{t=1}^{T} \{ r_{it} p(x_{jt}, y_{jt}, \theta_{jt}) + (1 - r_{it})(1 - p(x_{jt}, y_{jt}, \theta_{jt})) \}, \quad (3)$$

Score S_{ij} is a value for determining whether the moving track and tag reading data fit. The score is considered a high value when track and reading data have good fit. We adopt it for matching up "Who" and "Where" information.

III. EXPERIMENT

We use 320 · 240 [pixel] images and produce 5 × 4 grids to make a tag reading probability database. The database is made from about 9000 frame images. The data acquisition rate and frame rate of the movie are both 15 [Hz].

A. Experiment using Multi Track data

Five pattern tracks (each track has 200 frames of data) are chosen randomly. We calculate Score using it and its corresponding tag reading data.

The five tracks are called Track 1 – Track 5. Track 1 – Track 5 are shown in Fig. 5, as determined by image processing. In addition, Fig. 6 – Fig. 10 show graphs of the Score transition. Each Score is calculated using a person's location data stream of Track 1 – Track 5, i.e., when $\boldsymbol{Q}_1$ – $\boldsymbol{Q}_5$ (where each is a corresponding data matrix of Track 1 – Track 5) and $\boldsymbol{r}_1$ – $\boldsymbol{r}_5$ (where each is corresponding tag reading data stream) are defined; then the graph of Fig. 6 shows S_{11} – S_{51} , Fig. 7 shows S_{12} – S_{52} , Fig. 8 shows S_{13} – S_{53} , Fig. 9 shows S_{14} – S_{54} , Fig. 10 shows S_{15} – S_{55} . Each graph shows the transition of Score.

Furthermore, Fig. 7 – Fig. 10 show, for each result, the largest Score value is S_{11} [S_{22} , S_{33} , S_{44} , S_{55}] at the end of the graph (value at 200th frame). However, it is difficult to determine the correct pair of track data and tag reading data when comparing S_{33} and S_{43} (or S_{34} and S_{44}) because each score transition is very similar. It is also difficult to do it in the case of S_{55} because S_{55} is ***not*** the largest value at most parts of the Score transition in Fig. 10.

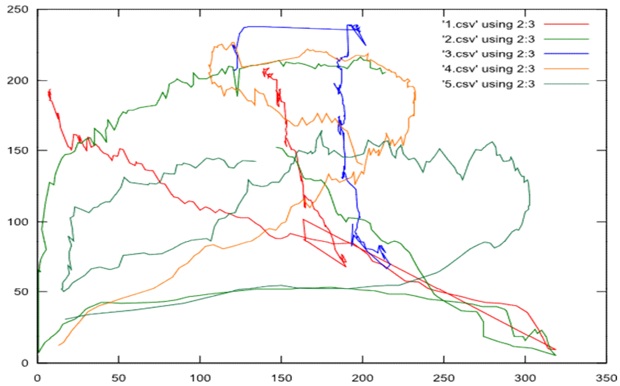

Fig. 5 Image of Tracks using Experiment A

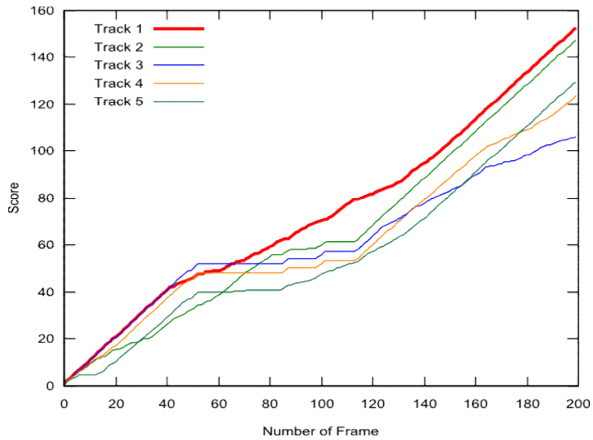

Fig. 6 Graph of Score transition S_{11} - S_{51}

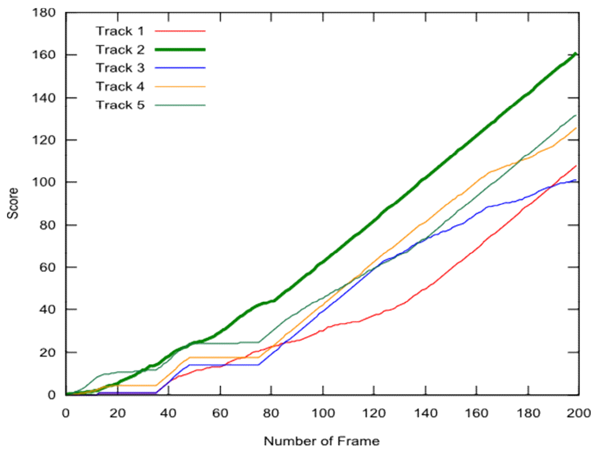

Fig. 7 Graph of Score transition S_{12} - S_{52}

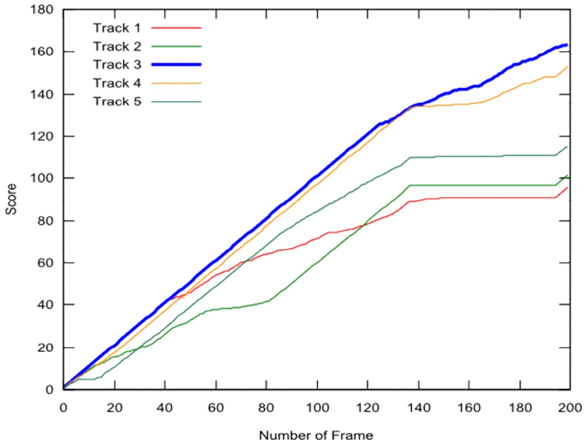

Fig. 8 Graph of Score transition S_{13} - S_{53}

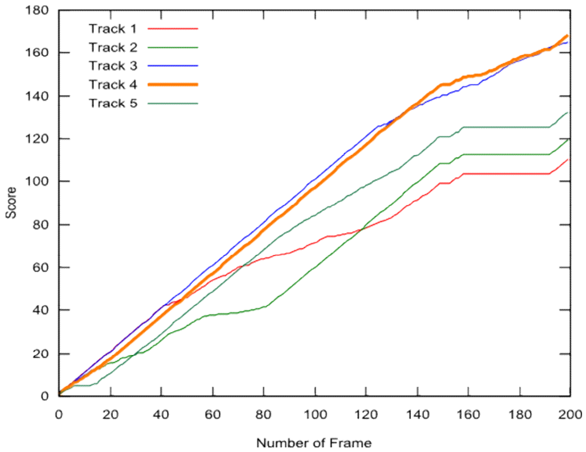

Fig. 9 Graph of Score transition S_{14} - S_{54}

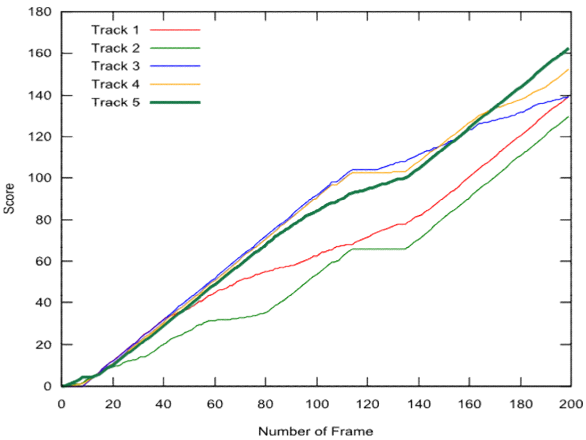

Fig. 10 Graph of Score transition S_{15} - S_{55}

B. Experiment using Multiple Persons' data

In this experiment, we adopt a multiple moving object detection method using the EM algorithm. Using it, we obtain tracks of multiple persons. In this case, we adopt it for two people shown simultaneously in an obtained image. It is an offline process to obtain a trajectory of moving objects and Score calculation. The original image obtained from a fixed camera and an example image of extracting moving object process are portrayed in Fig. 11.

Extracted tracks of two people are defined as Track A and Track B. Fig. 12 shows Track A and Track B. Figs. 13 and 14 show graphs of the Score transition. Fig. 13 is a graph using a tag reading data stream of Track A; Fig. 14 is using one of Track B, i.e. Fig. 13 is Score transition of S_{AA} , S_{BA} and Fig. 14 is one of S_{AB} , S_{BB} . Fig. 13 shows that S_{AA} is larger than S_{BA} after about the 120th frame and Fig. 14 shows that S_{BB} is larger than S_{AB} after about the 20th frame. Therefore, it is safe to say that track data and tag reading data are correctly matched.

C. Adoption of a Short length Score

Matching up of the "Who" and "Where" information method using in the previous two sections cannot be used with real-time matching. Therefore, we adopt a short length Score, which is calculated by adding the Score per 5–15 [frames](i.e. ca. 1[s]). Three pattern interval frames are used: 5 [frames], 10 [frames], and 15 [frames], and match-up accuracy rates are calculated. Data streams of track and tag reading are the same as those of Experiment B.

Results using data streams of Track A are presented in Fig. 15 and Fig. 16. Each is short length. The Score transitions are of 5 [frames] and 15 [frames] interval frames. Accuracy rates are shown in TABLE I. They are 66.0%, 72.5%, and 76.5%, respectively, for 5 [frames], 10 [frames], and 15 [frames]. In this case, the correct interval is when $S_{AA} > S_{BA}$.

On the other hand, accuracy rates of using data streams of Track B are shown in TABLE II. The accuracy rates are 63.1%, 66.7%, and 67.6% in order. In this case, the correct interval is when $S_{BB} > S_{AB}$.

Furthermore, result of paired t-test, the mean of difference (S_{AA}, S_{BA} and S_{BB}, S_{AB}) is statistically significant on the significant level one percent using data of 5[frames] short length Score(TABLE III).

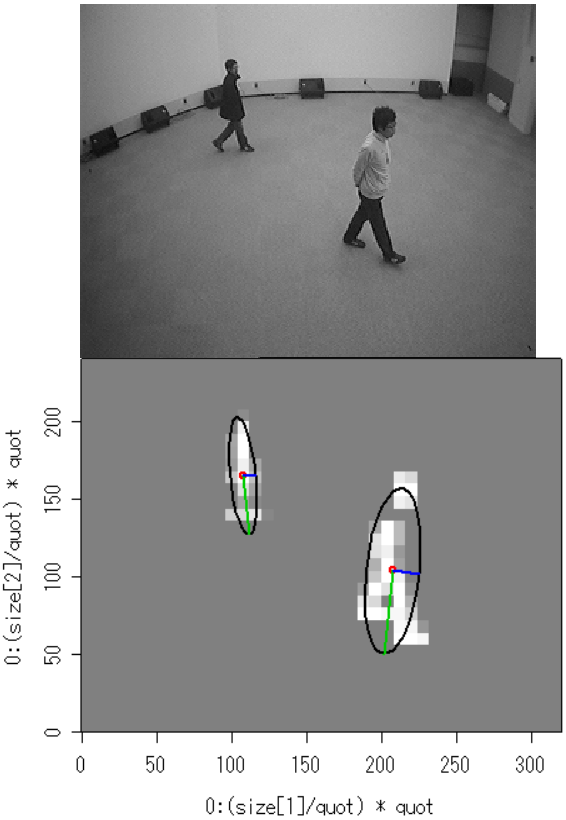

Fig. 11 Original image (upper) and image of extracting moving object process (lower)

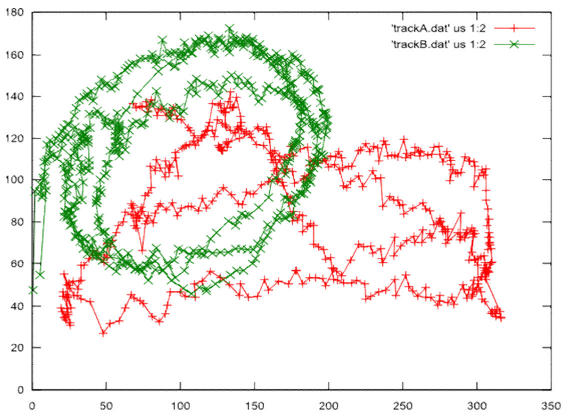

Fig. 12 Plot image of Track A and Track B used in Experiment B

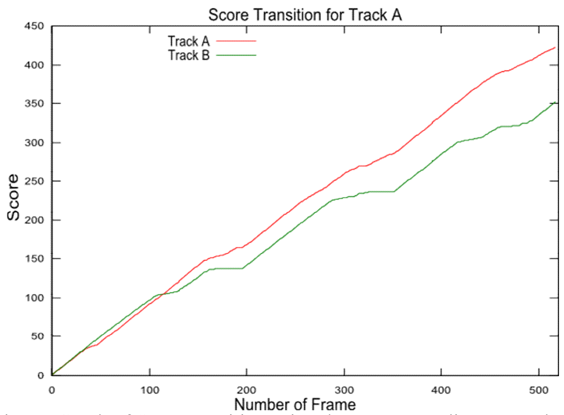

Fig. 13 Graph of Score transition using data corresponding to Track A

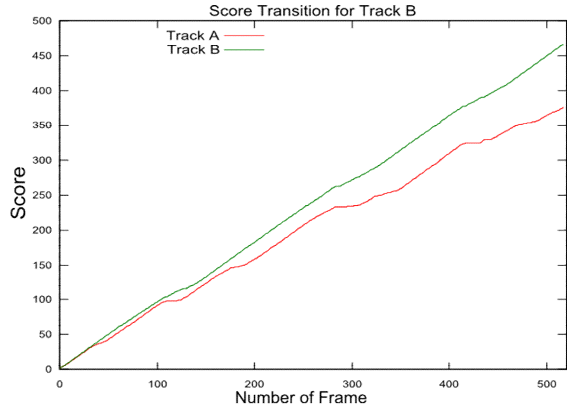

Fig. 14 Graph of Score transition using data corresponding to Track B

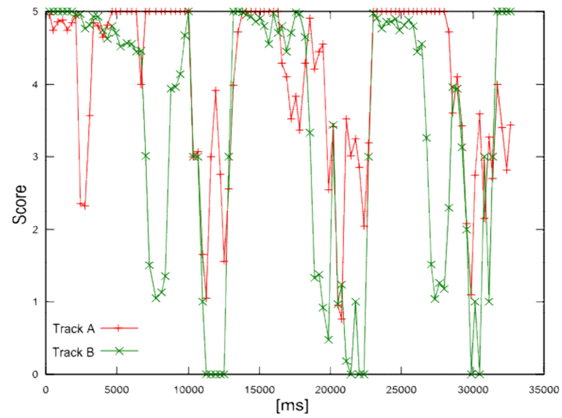

Fig. 15 Short length Score transition using data of Track A for 5 [frames]

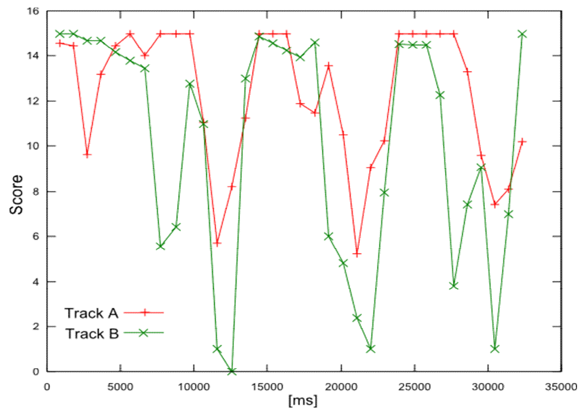

Fig. 16 Short length Score transition using data of Track A for 15 [frames]

TABLE I
Accuracy of Score comparison using short length Score
(Track A data)

	5[frame]	10[frame]	15[frame]
All num of interval	103	51	34
Num of correct interval	68	37	26
Accuracy Rate(%)	66.0	72.5	76.5

TABLE II
Accuracy of Score comparison using short length Score
(Track B data)

	5[frame]	10[frame]	15[frame]
All num of interval	103	51	34
Num of correct interval	65	34	23
Accuracy Rate(%)	63.1	66.7	67.6

TABLEIII
Result of t-test using short length Score(5[frames] data)

	t value	df	P value
S_{AA}, S_{BA}	4.5247992	102	1.63912E-05
S_{BB}, S_{AB}	6.2581308	102	9.27621E-09

The accuracy rate transition graph when the interval frame length is changed is shown in Fig. 17 for data of Track A. A higher accuracy rate when the length for a longer interval. From the explanation given above, the accuracy rate of track and tag reading information match-up and length of interval show a tradeoff relation.

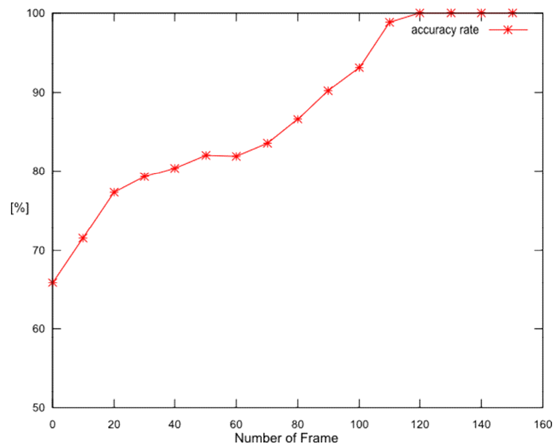

Fig. 17 Graph of accuracy rate transition by changing the interval length

IV. CONCLUSION

This paper presents discussion of "Who" and "Where" information, which is important for service provision. Obtaining these two pieces of information simultaneously is necessary to achieve adaptive service provision.

For adaptive service provision, we proposed this method to obtain "Who" and "Where" information simultaneously. "Who" information is obtained using the UHF-band RFID system; "Where" information is obtained by image processing to use background differences.

We carried out preliminary experiments using UHF-band RFID attributes. We proposed and adopted tag-reading probability.

We proposed "Score" to match up "Who" and "Where" information. We used experiments to analyze Score's attributes and whether we can match up these two pieces of information using Score.

We proposed the short length Score for real-time processing of match-up information. Results of experiments show a link between the interval frame length and accuracy. Using our proposed method, it was possible to obtain "Who" and "Where" information simultaneously.

REFERENCES

[1] T. Sato, "Robotic room: human behavior measurement, behavior accumulation and personal/behavioral adaptation by intelligent environment," Advanced Intelligent Mechatronics 2003 (AIM 2003), *Proc. 2003 IEEE/ASME International Conference on*, Vol. 1, pp. 515–520, Jul. 2003.

[2] K. Morioka, Joo-Ho Lee, and H. Hashimoto, "Human centered robotics in intelligent space," Robotics and Automation, *Proc. ICRA '02. IEEE International Conference on*, Vol. 2, pp. 2010–2015, 2002.

[3] A. Pentland, "Smart rooms, smart clothes," *Scientific American*, Vol. 274, No. 4, pp. 68–76, Apr. 1996.

[4] C. D. Kidd, R. Orr, G. D. Abowd, C. G. Atkeson, I. A. Essa, B. MacIntyre, E. D. Mynatt, T. Starner, and W. Newstetter, "The aware home: A living laboratory for ubiquitous computing research," *CoBuild '99: Proc. of the Second International Workshop on Cooperative Buildings, Integrating Information, Organization, and Architecture,* pp. 191–198, London, UK, 1999. Springer Verlag.

[5] R.A. Brooks, "The intelligent room project," *Proc. Second International Conference on Cognitive Technology, 1997.'Humanizing the Information Age'*, pp. 271–278, Aug. 1997.

[6] D. Murakami-Wood et al., "A report on the surveillance society," *Report for The Information Commissioner*, Sept. 2006.

[7] M. Yamada, K. Ebihara, and J. Ohya, "A new robust real-time method for extracting human silhouettes from color images," In *FG '98: Proc. of the 3rd International Conference on Face & Gesture Recognition*, p. 528, Washington, DC, USA, 1998. IEEE Computer Society.