Perceptual Functions for Context-Awareness of An Office Worker

Yusuke Manabe*, Hiromichi Saito*, Keisuke Akiyama*, Ryotaro Ikeda*, Satoshi Kanda* and Kenji Sugawara*
*Department of Information and Network Science
Chiba Institute of Technology
Narashino-shi, CHIBA 275-0016 JAPAN
{ymanabe, suga}@net.it-chiba.ac.jp

Abstract—**Symbiotic Computing is a kind of research project and concept to bridge the gap between Real Space (RS) and Digital Space (DS). The purpose is mainly to develop flexible information service or application as well as to establish the next generation information platform. In order to establish the framework, it is necessary to build 'Mutual Cognition' between human and system. Mutual Cognition broadly consists of two functions; 'RS Cognition' and 'DS Cognition'. This paper focuses on RS Cognition, which is perceptual function of software system for some events or human's activities. In this study, we develop two perceptual functions, which are sitting posture recognition and human's location recognition for an office worker, as a kind of RS perception task. As the results of experiments, we found that our developed functions are quite competent to recognize worker's activities.**

I. Introduction

Realization of symbiosis between human and computer system is a very important idea in the context of ubiquitous computing. However, under the current situation, people cannot receive satisfactory and suitable services they required because systems cannot perceive dynamic situation in Real Space (RS). In other words, a lot of current computer service has been realized based on information only in Digital Space (DS). We address this problem and propose a concept of Symbiotic Computing.

The purpose of Symbiotic Computing is mainly to develop flexible information service or application as well as to establish the next generation information platform based on multi-agent framework[1][2][3]. In addition, a specific characteristics of Symbiotic Computing is to aim at establishment of mutual cognition between RS and DS. Mutual cognition is a cognitive process defined by the relation between a personal feeling that "I know what you know about me" and a machine's activity as if "I know what you know about me."

Figure 1 shows a model of mutual cognition processing, which cinsists of 4 steps. The first step is to recognize human's activity. The next step is to infer the human's request from the activity and send it to DS. The third step is to give information or serivices provided by DS to RS according to situation. The last step is to check feeling of contentment of the human to evaluate the process of the mutual cognition. The realization of mutual cognition can provide suitable and secure services based on situation in RS to people.

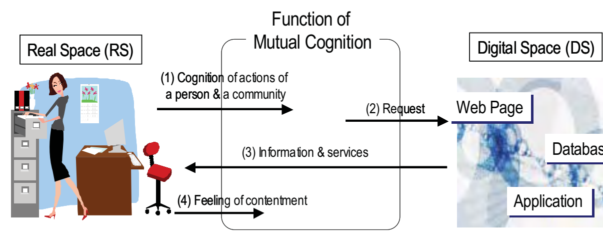

Fig. 1: Mutual Cognition Processing

In other words, mutual cognition model is one of the specific model of context-awareness computing[4]. According to referrence [4], the term context-aware was first used in 1994 by Schilit and Theimer[5]. Until now there are a lot of similar definition of context or context-awareness[6]. In order to clear our purpose, we follow the definition of context and context-aware in reference [6]. Especially, definition 2 is almost equivalent to the idea symbiotic computing is trying to realize.

Definition 1 : *Context is any information that can be used to characterise the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves.*

Definition 2 : *A system is context-aware if it uses context to provide relevant information and/or service to the user, where relevancy depends on the user's task.*

This paper especially focuses on the first step in the mutual cognition processing, that is, to recognize human's activities. This is called 'RS cognition', which is perceptual function of software system for some events or human's activities. Perceptual function should be implemented in a context-awareness system. As information in RS are usually obtained as multi-variable and multi-dimensional time series measured by various sensors such as camera, microphone, thermometer, acceleration sensor and so on, it is important to build intelligent algorithms to segment and cluster complicated time series data as well as to associate clusterd data with sementic

Proc. 9th IEEE Int. Conf. on Cognitive Informatics (ICCI'10)
F. Sun, Y. Wang, J. Lu, B. Zhang, W. Kinsner & L.A. Zadeh (Eds.)
978-1-4244-8040-1/10/$26.00 ©2010 IEEE

Fig. 2: Postulated Environment

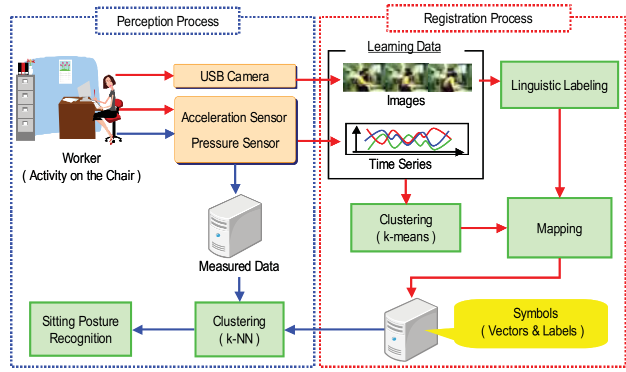

Fig. 3: Function 1: Sitting Posture Recognition

representations of some kind.

In this study, we postulate personal office room/space as environment and we develop two perceptual functions, which can recognize sitting posture on the chair and location point in the room for an office worker in the context of telework application [7][8]. We utilizes optical fiber pressure sensor mattress and acceleration sensor for sitting posture recognition and two USB cameras for location point recognition. Two experiments have been done to evaluate the developed functions. As the results, we find that the developed functions are quite competent to recognize human's activities.

This paper consists of five sections. Section II shows our postulate environment and proposed perceptual functions. Section III and IV show the two experiments and results for recognzing sitting posture on the chair and location point in the room. Finally, we discuss about the conclusions and future work in section V.

II. Proposed Functions

This section explains about the proposed perceptual functions; sitting posture recognition and location recognition.

A. Postulated Environment: Personal Office Room/Space

In this study, we address a personal office room/space as environment although it is impossible to adapt for all of human's activities and environments. Figure 2 illustrates a personal office room/space as environment. We postulate that the area of personal space is around from 20 m^2 to 30 m^2.An office chair is a best tool of context-awareness when a person works at the desk. Thus, pressure sensor mattress is put on the seat of a chair and acceleration sensor is attached to the chair back. Moreover, a worker may walk out and go to bookshelf for taking documents, books or something. Thus, two USB cameras are set up the high position and the angle between them is around 90 degree. Under the postulated environment we mentioned here, we develop the perceptual functions for recognizing a worker's activities.

B. Function 1: Sitting Posture Recognition

The first function is to recognize sitting posture on the chair with sensors. Figure 3 shows recognition process. This process is divided into two flows; registration process and perception process.

Registration process is to create symbols for perception process. Symbols denote a set of cluster vectors obtained from sensor data and linguistic labels (semantics) based on image data. The detail of this process is shown as follows.

1) Image data, pressure data and acceleration data are observed by respective devices, USB camera, pressure sensor mattress and acceleration sensor. USB camera is not used in perception process because it is used for confirming the actual sitting posture in registration process.
2) We used k-means algorithm to cluster each time series data set of pressure data and acceleration data respectively. Pressure data (54 dimensional vector) are clusterd into 10 clusters. Acceleration data (3 dimensional vector) are clustered into 5 clusters. However, raw clustering result may be noisy because of sensing error and human's activity. Thus, clustering result is smoothened with respect to time based on the follwing equation;
$$L(t) = Mode(l(t-5), \ldots, l(t), \ldots, l(t+5)), \quad (1)$$
where t denotes time, $L(t)$ denotes smoothened cluster label and $l(t)$ denotes raw cluster label.
3) The continuous segments, which are clustered as the same cluster over 120 seconds, are extracted as the stable intervals which human keeps sitting posture. Then the extracted intervals are elected as candidates of symbol.
4) Image sequence corresponding to the extracted intervals are checked by manual, linguistic labels are created to each interval based on actual sitting posture estimated from image.
5) Finally, some pairs of cluster vector for the stable interval and linguistic label ara registerd as symbols.

Meanwhile, in perception process, the created symbols are used as knowledge to recognize sitting posture based on

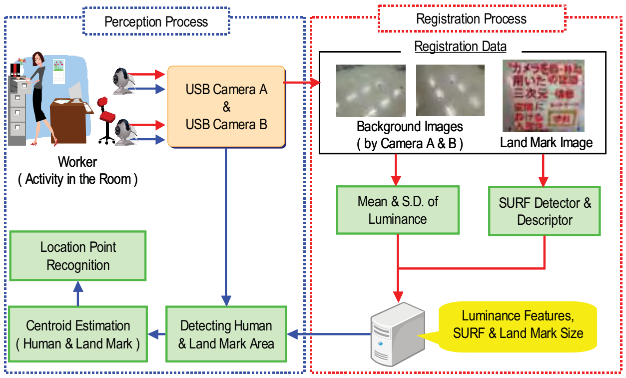

Fig. 4: Function 2: Human Location Recognition

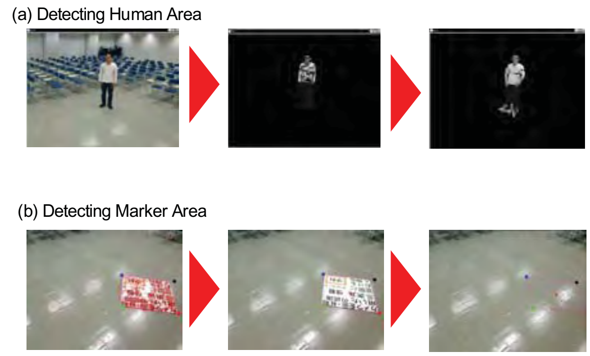

Fig. 5: Detecting Human and Landmark Area

observed data. The detail of perception process is shown as follows.

1) Pressure data and acceleration data are observed by respective devices and stored in database.
2) Next, respective sensor data are clustered by k-NN algorithm using the cluster vectors, which is created in registration process.
3) Clustering result is smoothened based on equation (1).
4) A specific linguistic label is assigned for the intervals being stabilized over 120 seconds.

C. Function 2: Location Point Recognition

The second function is to recognition human's location point in the small room/space by using two USB cameras. Figure 4 shows process flow. This process is also divided into two processes.

Registration process is shown as follows.

1) Background image sequences are captured by several two cameras (which are called camera A and B respectively). Needless to say, background images do not include human.
2) Next, image data of landmark object is prepared. Landmark object is located in the room/space in order to estimate human's precise location. Thus it is better that shape of landmark object is tetragonum with the top and the bottom as well as with the left and the right.
3) From the several background image sequences, the matrix of mean and standard deviation of luminance are calculated for 100 image frames. The pixel (i, j) is identified as background when it satisfies the following equation;
$$\bar{I} - \sigma \leq I(i, j) \leq \bar{I} + \sigma, \tag{2}$$
where I denotes luminance value, $\bar{I}$ denotes mean of luminance and σ denotes standard deviation of luminance.
4) From landmark image, Speeded Up Robust Feature (SURF)[9] is extracted. SURF is a robust image detector and descriptor, first proposed by Herbert Bay et al. in 2006. SURF is a scale- and rotation-invariant feature as well as it can detect the target object area speedily.
5) Finally, SURF and luminance matrix are registerd in order to detect landmarkand human.

In perception process, extracted features are used for detecting human area and landmarkarea respectively. The detail is shown as follows.

1) Human detection is rather detection of area excepting background. Figure 5(a) shows an example of human area detection based on our algorithms. This algorithm provides human's location point in broad terms.
2) Landmark detection is achieved by SURF detector and descriptor. Figure 5(b) shows an example of landmarkarea detection. If once landmarkis detected, it can be removed from the location.
3) From landmark area, length of four edges of landmark $l_1, l_2, l_3 and l_4$ are calculated. Then mapping ratio M_r is calculated as following equation;
$$M_r = \frac{4L_R}{l_1 + l_2 + l_3 + l_4}[cm/pixcel] \tag{3}$$
where L_R denotes actual length of an edge of landmarks.
4) In order to estimate precise location point of human, centroid point is calculated for each detection area.
5) Finally, precise human's location point is estimated by distance D between human's centroid and land mark's centroid and angle Θ between two vectors which are from land mark's centroid to top-right corner of landmark and human's centroid.

III. Expriment I: Evaluation of Function 1

Experiment I is to evaluate the function of sitting posture recognition.

A. Experimental Environment

Figure 6 shows experimental environment. A subject sits on the chair in front of desk and works. The contents of work are "using PC", "reading the book", "writing", "watching online videos" and so on.

Pressure sensor mattress is KINOTEX sensor and acceleration sensor is WiiRemote controller. Specification of each

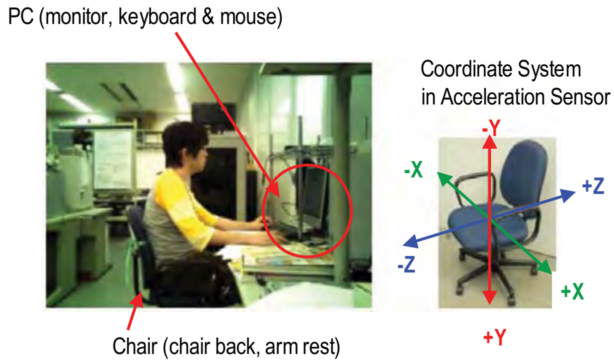

Fig. 6: Experimental Environment to Evaluate Function 1

TABLE I: Specification of each sensor device

(A) KINOTEX Pressure Sensor Mattress

Sensing Area Size (mm)	W200 x D270
Number of sensing points	54 (9 rows x 6 cols)
Pitch of sensing points (mm)	row=30, col=30
Measureable size (kPa)	about 0 - 10
Sampling rate (Hz)	about 40

(B) WiiRemote Controller

Size (mm)	H148 x W36.2 x D30.8
Measurement Axis	3-axis (X, Y, Z)
Wireless Communication	Bluetooth （Broadcom）

sensor device is shown as table I. WiiRemote controller is attached as the coordination system shown in Figure 6.

Development environments are as follows;

- PC : AMD Athlon Dual CoreProcessor 5000B 2.6GHz 4GB memory,
- OS : Windows XP Professional,
- Programming language : C#, Visual C++ (MFC)
- IDE : Visual Studio 2008 Express Edition (WiiRimote controller), Visual Studio 2005 Professional Edition (KINOTEX)
- Libraries : Wiimote Lib v1.7 [10]

B. Detail of Experiment

In this experiment, we measured pressure sensor data, acceleration data and image data in twice (which are called data set 1 and 2 respectively). Sampling rate is 1 Hz for all of data. Measurement time are 3680 seconds (data set 1) and 3755 seconds (data set 2) and we extracted only 3600 seconds from each data. We used data set 1 as learning data and used data set 2 as test data. Image data size is 320 x 240 and format is PNG (Portable Network Graphics).

Data clustering algorithms (k-means and k-NN) are achieved by *kmeans* function and *knn1* function in statistical package R[11]. Maximum iteration times of k-means is 1000000, the algorithm is Hartigan-Wong method.

C. Results

Figure 7 shows the result of clustering pressure sensor data by k-means algorithm for data set 1. Unfortunately, acceleration data could not be clustered successfully because of less of variant and lack of a part of data. Thus in this paper we mainly address the result of pressure sensor data and comlementaly address the result of acceleration data. From Figure 7, eight stable intervals are segmented by seven clusteres. Each image means that a subject mainly keeps the posture in each interval. From these images, we create four linguistic labels on sitting postures as follows;

1) *lean the chair* : This posture means that a subject leans the chair back.
2) *lean the desk* : This posture means that a subject leans the desk. For example, he puts his elbow or arm on the desk.
3) *proper sitting* : This posture means that a subject sits on the chair and mainly use keyboard while keeping proper sitting posture.
4) *nobody* : This posture means that nobody else sit on the chair.

Finally, we can create six symbols from data set 1 in Figure 8. Symbol #1, #2, #4, #5 and #6 are corresponding to cluster #1, #2, #7, #9 and #10 respectively. Symbol #7 is created by the mean of cluster #3 and #5 because their cluster vectors are very strong correlation (correlation coefficient is 0.95) as well as their images of sitting posture are very similar condition.

Figure 9 shows clusering and recognition result for data set 2. Cluster number in k-NN algorithm are corresponding to symbol number created from data set 1. From this figure, nine stable intervals (from 'A' to 'I') are segmented by six

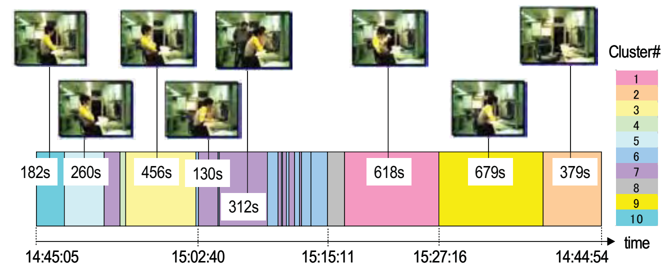

Fig. 7: Clustering Result by k-means (Data Set 1)

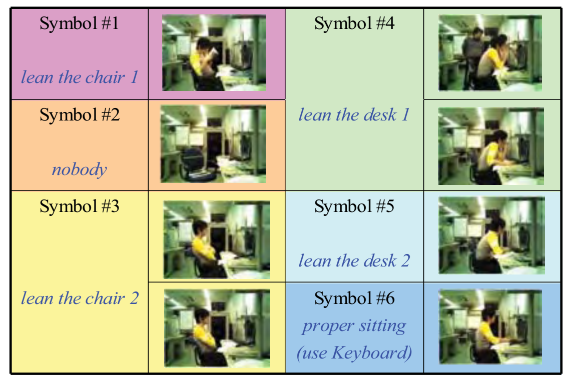

Fig. 8: Acquired Symbols (Data Set 1)

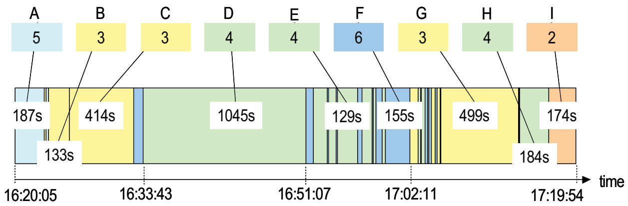

Fig. 9: Clustering Results by k-NN (Data Set 2)

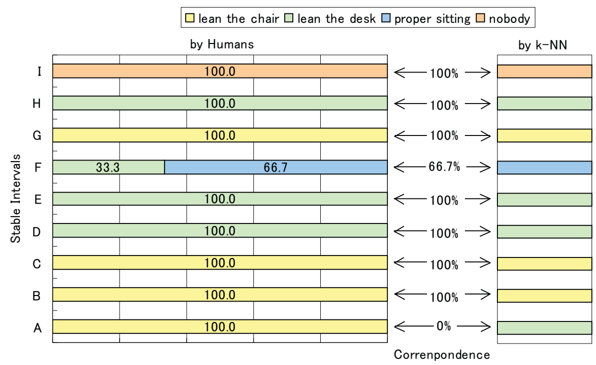

Fig. 10: Correspondence between Human's Clustering and k-NN Clustering

clusteres. For example, interval 'D' is recognized as *lean the desk* because it is labeled by symbol #4.

In order to evaluate the clustering and recognition accuracy, we checked movie files, which are created by combining image sequence in each interval. Figure 10 shows correspondence between human's clustering and k-NN clustering. The left bar graph denotes clustering result by humans and the right bar graph denotes clustering result by k-NN corresponding to Figure 9. As the result, we can found that the interval 'A' are not good recognition accuracy but the other intervals are good accuracy. The failure of the interval 'A' means that sitting posture and seat pressure pattern are not one-to-one correspondence, that is, a sitting posture can include some seat presure patterns.

Figure 11 shows variance of acceleration data of Z axis. Alphabes 'D', 'E', 'H' and 'I' correspond to the stable intervals in Figure 9. We can find that a worker is not using chair back when variance is small, because 'D', 'E' and 'H' intervals denote *leand the desk* and 'I' interval denotes *nobody*. Although acceleration data is three dimensional data (X, Y and Z), X axis data hardly change because the chair back with acceleration sensor does not axially change with respect to X axis. We checked that Y axis data changes but it does not include relations to sitting posture patterns.

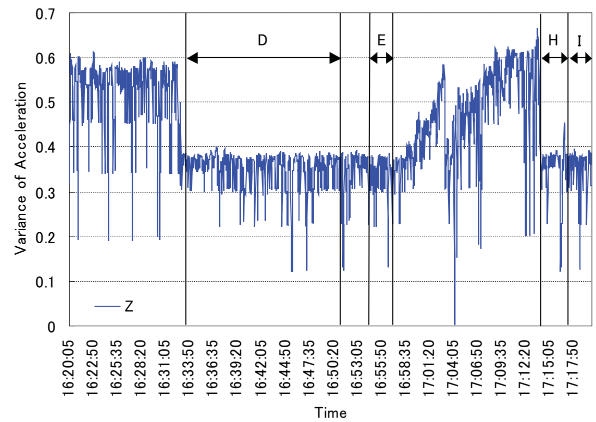

Fig. 11: Acceleration Data of Z axis (Data Set 2)

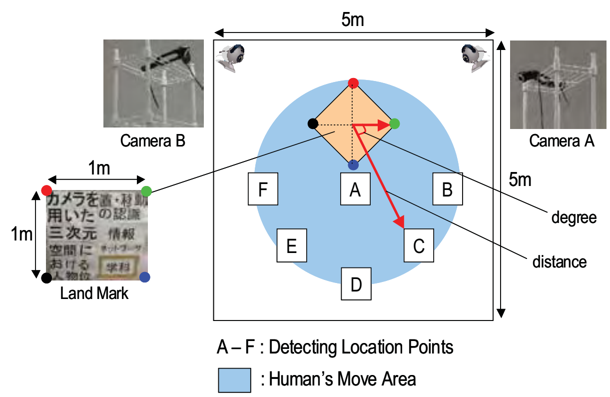

Fig. 12: Extperimental Environment to Evaluate Function 2

TABLE II: Specification of Logicool Webcam Pro 9000

Size (mm)	W89.9 x D117.9 x H39.8
Focus	10 cm – ∞ (Auto Focus)
Angle of View	76 degrees
Video Capture	2 megapixels (1600 x 1200) (MAX)
Frame Rate	30 frames / second (MAX)

IV. Expriment II: Evaluation of Function 2

Experiment II is to evaluate the function of human's location point recognition.

A. Experimental Environment

Figure 12 shows extperimental environment to evaluate function 2. USB cameras are Logicool Webcam Pro 9000, whose specification is shown as Table II. The area of space is around 25 m^2 (5 meters by 5 meters). Human's move area is within a circle, whose size is around 3 meters in diameter. A subject walks within the circle area. Recognition accuracy is evaluated for each detection location point; 'A' to 'F', distance D and angle Θ are defined in Figure 12. Landmark is square and its edge legth is 1 meter.

Development environments are as follows;

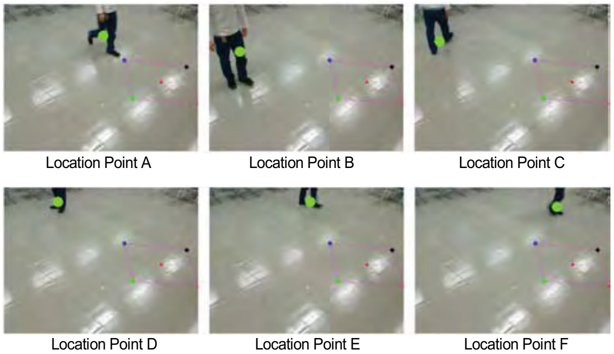

Fig. 13: Detecting Human's Centroid Position (Camera A)

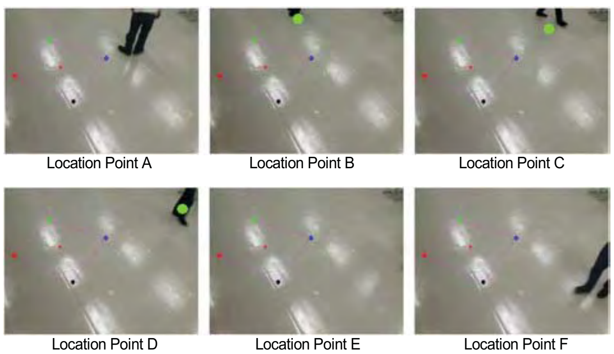

Fig. 14: Detecting Human's Centroid Position (Camera B)

- PC : AMD Athlon Dual CoreProcessor 5000B 2.6GHz 4GB memory,
- OS : Windows XP Professional,
- Programming language : C,
- IDE : Visual Studio 2008 Express Edition,
- Libraries : Open CV [12] 2.0.0 α, Video Input 0.1995.

B. Results

Figure 13 and 14 show the results of detecting human's centroid position for each recognition point by camera A and B respectively. Small green-painted circle denotes centroid position. We can find that each human's location point can be detected successfully in camera A. On the other hand, in case of camera B, we can find that some location points such as A, E and F are not detected successfully.

Next, Figure 15 shows distance error of actual value and estimated value between human's centroid point and land mark's centroid point for each location. The vertical axis denotes estimation error for distance. The average of error in camera A and camera B are 50.16cm, 83.75cm respectively. Figure 16 shows angle error of actual value and estimated value. The average of error in camera A and camera B are 21.75 degrees and 13.71 degrees respectively. From there graphs, we can find that in distance estimation camera A is better than camera B and in angle estimation camera B is better than camera A. In addition,

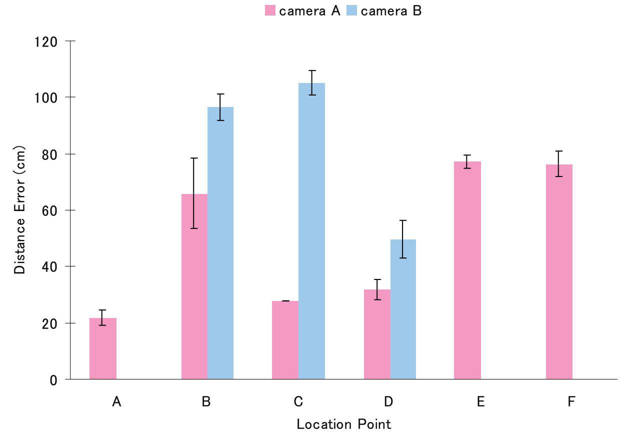

Fig. 15: Error between Actual and Estimated Distance for Each Location Point

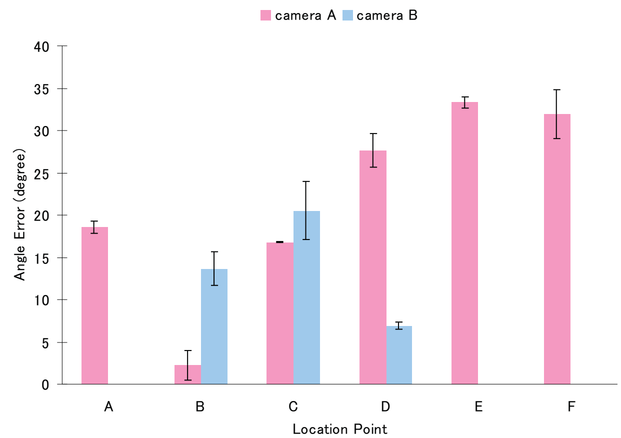

Fig. 16: Error between Actual and Estimated Angle for Each Location Point

The reason is mainly due to no depth and height correction. In Figure 13 and 14, the upper part of each image is far from the lower part due to perspective. Moreover, from figure 13 and 14 human's centroid points are up and down based on whether human's upper-body is captured or not. Actually, in case of good estimation points such as the position 'C' and 'D' in camera A, human's centroid point is near the detection point in the floor because human's upper-body is not captured. The position 'B' is fewer gap for depth direction although human's upper-body is captured. In case of bad estimation point such as the position 'F', it is located on the upper area of image although human's upper-body is not captured.

V. Conclusion

Symbiotic Computing is a post ubiquitous computing model to bring mutual cognition between Real Space and Digital Space. In this study, we focused on 'RS Cognition' which is perceptual function of software system and developed two perceptal functions; sitting posture recognition (function 1) and human's location point recognition (function 2) for an office worker. In function 1, we found that our proposed algorithm can create the useful symbols and recognize sitting postures based on pressure sensor time series data. In fuction 2, our proposed algorithm can detect the human's location. Thus our developed functions are quite competent to recognize human's activities.

In the next future, we are going to solve the following problems;

1) To create various symbols to recognize a lot of sitting posture pattern (function 1).
2) To solve depth correction problem in order to recognize human's location position in higher performance (function 2).
3) To apply our algorithm to the actual environment where includes chair and desk as well as drawer, chest, book shelf and so on (function 1 and 2).
4) To integrate our proposed algorithms and high-order cognition function such as reasoning or knowledge processing based on multi-agent framework (function 1 and 2).

References

[1] Symbiotic Computing Project, (URL) http://symbiotic.agent-town.com/ .

[2] T. Suganuma, K. Sugawara, T. Kinoshita, F. Hattori, N. Shiratori, Concept of Symbiotic Computing and its Agent-Based Application to a Ubiquitous Care-Support Service, IJCINI Vol. 3, No. 1, pp. 34–56, 2009.

[3] K. Sugawara, S. Fujita, T. Kinoshita, N. Shiratori, A design of cognitive agents for recognizing real space —towards symbiotic computing—, Proc. of 7th IEEE Internatinal Conference on Cognitive Informatics, pp. 277–285, 2008.

[4] S. Poslad, Ubiquitous Computing: Smart Devices, Environments and Interactions, WILEY, 2009.

[5] B. N. Schilit and M. M. Theimer, Disseminating Active Map Information to Mobile Hosts, IEEE Network, Vol. 8, No. 5, p.p. 22–32, 1994.

[6] A. K. Dey, G. D. Abowd, Towards a Better Understanding of Context and Context-Awareness, In the Workshop on The What, Who, Where, When, and How of Context-Awareness, as part of the 2000 Conference on Human Factors in Computing Systems (CHI 2000), 2000.

[7] K. Sugawara, S. Fujita and H. Hara, A Concept of Symbiotic Computing and Its Application to telework, Proc. of 6th IEEE Internatinal Conference on Cognitive Informatics, pp. 302–311, 2007.

[8] S. Konno, Y. Manabe, S. Fujita, K. Sugawara, T. Kinoshita, N. Shiratori, A Framework for Perceptual Functions of Symbiotic Computing, Web Intelligence/IAT Workshops, pp. 501–504, 2008.

[9] H. Bay, A. Ess, T. Tuytelaars, L. V. Gool, SURF: Speeded Up Robust Features, Computer Vision and Image Understanding (CVIU), Vol. 110, No. 3, pp. 346–359, 2008.

[10] Wiimote Lib, (URL) http://www.codeplex.com/WiimoteLib .

[11] The R Project for Statistical Computing, (URL) http://www.r-project.org/ .

[12] Open Computer Vision Library, (URL) http://sourceforge.net/projects/opencvlibrary/ .