A Wearable System That Captures Human Activities and Social Interactions

Seon-Woo Lee[1] **Sadanori Ito**[2] **Yasuyuki Sumi**[2,3] **Kenji Mase**[2,4] **Kiyoshi Kogure**[5]
senu@hallym.ac.kr sito@atr.jp sumi@i.kyoto-u.ac.jp mase@itc.nagoya-u.ac.jp kogure@atr.jp
[1]Div. of Info. Eng. & Telecommunications, Hallym University, Chuncheon, KOREA
[2]ATR Media Information Science Laboratories, Kyoto, JAPAN
[3]Graduate School of Informatics, Kyoto University, Kyoto, JAPAN
[4]Information Technology Center, Nagoya University, Nagoya, JAPAN
[5]ATR Intelligent Robotics and Communication Laboratories, Kyoto, JAPAN

ABSTRACT

This paper describes a wearable system for detecting, recognizing and recording what a user sees, says, and does. The developed system is a core component of our project of building a machine-readable interaction corpus, with stationary sensing systems installed ubiquitously in an indoor environment. The system is composed of a mobile PC, a camera, a microphone, and two microcontroller-based sensing modules. One of the modules is for recognizing the motion of a user such as sitting, standing, walking, and ascending/descending a stairway. The other has a CMOS image sensor to detect and recognize the ID tags with infrared LED, which are attached to other people, artifacts, and some places. Using the camera and microphone, the mobile PC collects the video/audio data as well as the recognized information about current motion, location, people and/or artifacts.

Keywords

Context-sensing, experience capturing, wearable device, activity recognition, social interaction

I. INTRODUCTION

As Weiser has suggested, large number of computers are pervading our environment. To fully achieve this vision, we need a new human-computer interaction paradigm based on embodied interactions beyond existing current Human-Computer Interactions frameworks based on desktop metaphor and Graphical User Interfaces. Therefore, we aim to build a machine-readable dictionary of interaction protocols among humans, artifacts and environments.

Initially, we proposed to build an interaction corpus, a semi-structured set of a large amount of data collected by various sensing systems [1]. This corpus may serve as an infrastructure for researchers to analyze and model social protocols of human interactions with other humans and/or artifacts. Our approach is characterized by the integration of stationary sensing systems installed in indoor environments and wearable sensing systems worn by users, as shown in Figure 1. We need many sensors and integrate the sensed data to capture and understand our daily experiences.

In this paper, we describe more specifically the wearable sensing system for recognizing and collecting such experiences of user. The wearable system is composed of two sensing modules: activity and interaction sensing modules. The activity sensing module recognizes the motion activities mostly occurred in our daily life such sitting, standing and walking. The other one, interaction sensing module find the relative distance and angle between the wearer and other person or artifact with a infrared signal transmitter, called 'IrID tag'. The found relative distance can be used to analyze the human-human and human-artifact interactive activities.

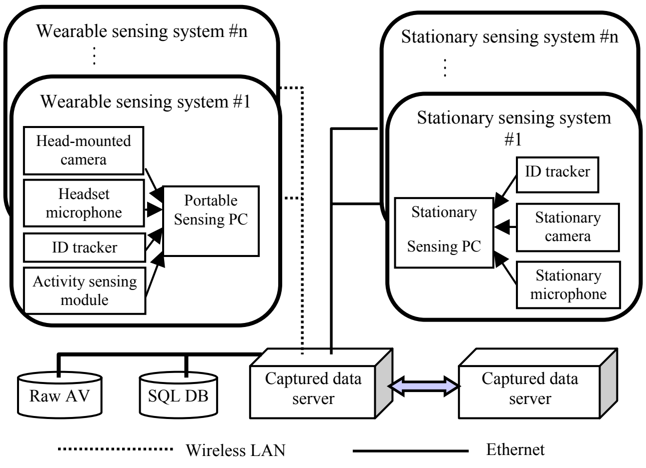


Figure 1. Block diagram of the proposed interaction corpus capturing system

2. SYSTEM DESCRIPTION

Figure 2 shows that the developed wearable system is composed of a mobile PC, a head-mounted display and camera, two microphones, a motion sensing module, and an infrared signal tracking module (IrID tracker). When the modules detect and recognize some motions or IrID Tags, they send the recognized information to the mobile PC using separated serial communication ports. To avoid using such an obsolete interface, we plan to make a centralized sensing system to connect each microcontrollor-based sensing module via an I^{2}C (Inter-Integrated Circuit) bus.

2.1 Motion-Sensing Module

Recognizing the user's activities such as sitting, standing, walking, or running is an essential part of capturing the user context [3~5]. To do that, some interesting ideas have been suggested. The developed motion-sensing module is designed for the wearable system from our PDA-based previous work [2] to recognize two poses and five motions of a user, including:

- Sit (*Si*) and Stand (*St*) poses.
- Stand to Sit (*StSi*) and Sit to Stand (*SiSt*) motions.
- Normal walking motion (*N*): means walking on level ground.
- Up(*U*) and Down(*D*) walking motion: mean walking up and down on stairways, respectively.

As shown in Figure 2(a), the motion-sensing module is implemented in two separated boxes. One box (denoted as the "leg box"), including a bi-axial accelerometer (ADXL 202EB from Analog Devices Inc.) and a gyroscope (Gyrostar ENV-05D from Murata Co.), is assumed to be in the user's right or left trouser pocket and measures the acceleration and angle of the user's thigh. The other one ("waist box"), including a RISC microcontroller (C8051F125, 16 MHz from Cygnal Co.), a 9 V battery, a power regulator, an RS-232 signal converter, and a connector.

The accelerometer in the leg module measures the forward and upward accelerations of the user's thigh, which are denoted by $a_x(t), a_z(t)$, respectively. The system then measures the angle (denoted by $\theta(t)$) of the user's thigh movement using a digital integrator of the angular velocity, $\dot{\theta}(t)$, obtained from the gyroscope, and the microcontroller reads the two acceleration and angular velocity signals every 40 msec. After reading, the microcontroller tries to detect the user's current motion with the following algorithm.

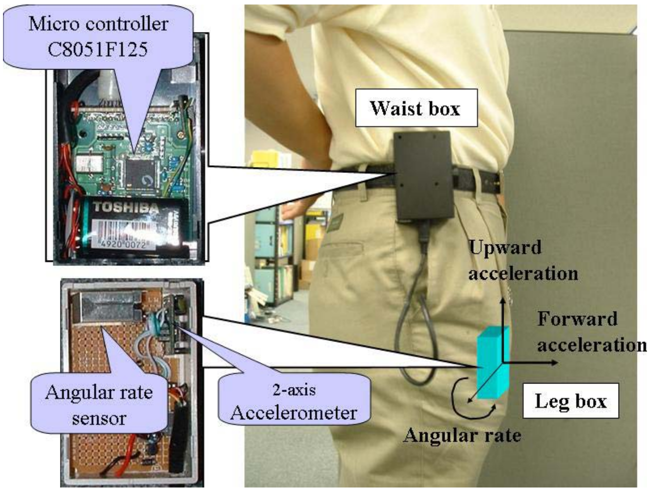
We define the following values as a basic feature vector:

$$\{\sigma_x(t), \sigma_z(t), \sigma_\theta(t), \Delta\theta_1(t), \Delta\theta_2(t), \Delta\theta_3(t)\} \quad (1)$$

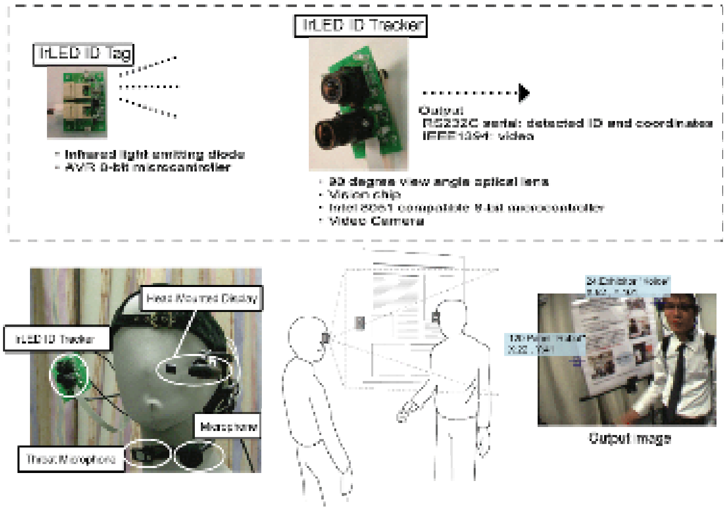
where $\sigma_x(t)$, $\sigma_z(t)$, and $\sigma_\theta(t)$ are the standard deviations over 25 samples of the forward acceleration, upward acceleration, and the thigh angle, $\theta(t)$, respectively, while $\Delta\theta_{\{1,2,3\}}(t)$ represent the past three angle differences when the angle direction changed. Each value of the angle difference can be obtained from the integration of angular velocity in a time interval between zero crossings of $\theta(t)$.

It is easy to recognize the *Si* and *St* poses by using the capability of the accelerometer to detect an absolute gravitational acceleration when the values and are less than a threshold value at every 1 (sec), as shown in Figure 3. On the contrary, the *SiSt* and *StSi* motions can be recognized by testing the following conditions:

- If $\sigma_\theta(t) > Th_{StSi}$, $\Delta\theta_1(t)$ >30°, and the upward acceleration $a_z(t)$ > Th_{StSi}, then current motion is *StSi.*
- If $\sigma_\theta(t) > Th_{SiSt}$, $\Delta\theta_1(t)$ <-40°, and the upward acceleration $a_z(t)$ < Th_{SiSti}, then current motion is *SiSt.*



(a) Motion sensing module



(b) ID tracking module

Figure 2. The suggested wearable sensing system

For the three walking behaviors, we could determine some common characteristics of the selected feature values, for many people such as:

- *N* walking: $\Delta\theta_1(t)$ is slightly positive, $\Delta\theta_2(t)$ is slightly negative, $\Delta\theta_3(t)$ is strongly positive.

- *U* walking: $\Delta\theta_1(t)$ is strongly negative, $\Delta\theta_2(t)$ is strongly positive, and both certainly have a longer duration.
- *D* walking: $\Delta\theta_1(t)$ is slightly positive, $\Delta\theta_2(t)$ is moderately negative, $\Delta\theta_3(t)$ is moderately positive.

Whenever the motion-sensing module detects a new zero crossing of $\theta(t)$, it tries to detect and discriminate the current motion by using a nearest neighborhood method with center values obtained a priori for the three walking behaviors. If the differences are less than some threshold values, then the module eventually recognizes new walking behavior. The center values should be obtained from a set of prior experiments on each user, because every user has different feature values. In order to assist with this task, we developed also an automatic calibration method that can provide statistical values for such features.

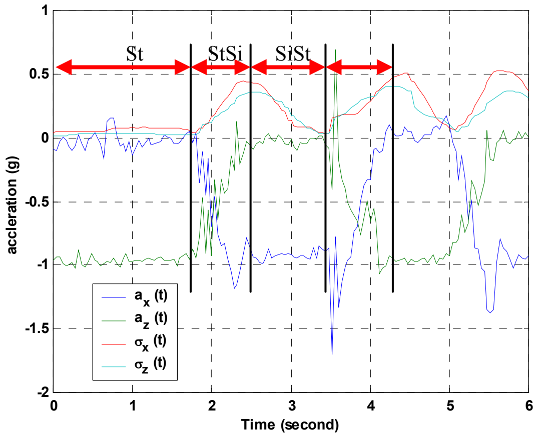


Figure 3. Typical sensor signals for *St, Si, StSi, SiSt*

To evaluate the performance of the proposed recognition method, the experiments were performed with a male subject (age 36, height 170cm, weight 67 kg). The subject walked in three times along a corridor (90 m) then down a stairway, and return to starting point. Table 1 shows the average results of the recognition method with the center values obtained from the developed automatic calibration method. We can get a similar performance to our PDA-based previous work [2], even by using an 8-bit microcontroller.

Unit (%)	*N*	*U*	*D*	Missing
N	92.2	1.2	2.1	4.5
U	0	91.7	0	8.3
D	1.3	0	92.1	6.6

Table 1. Recognition ratios

2.2 Interaction Sensing Module

Relative position between humans or between humans and artifacts is important for many applications used in an interactive context. Applications can infer activities or appropriate services that provide a user with such different types of relative position information as a person turning toward a display up to a certain distance or persons facing each other at a distance within which they can read each other's facial expression. In [4], the method of sensing and modeling human communication networks by using a wearable sensing device has been proposed. Such technologies often require the installation of large-scale infrastructures to calculate relative position. We, however, describe a decentralized accurate wearable relative positioning sensing module. The developed system consists of transmitters that use infrared light emitting diodes, called IrID Tag, and a localization device that uses a vision chip and an 8-bit microcontroller, called IrID Tracker, as shown in Figure 2(b).

The IrID Tag includes an 8-bit microcontroller (AT90S2323, Atmel Co.), an infrared LED (λ=850 nm, 2θ=30°, P_{max}=50 mA), as shown in Figure 2 upper-left corner. The IrID Tag emits its data in a total of 10 bits using Manchester encoding at the rate of 200 Hz. The 10-bit data is composed of a start bit (1 bit), unique ID (6 bits), a parity bit (1 bit), and a stop bit (2 bits). Using the coin battery, the IrID Tag runs 34 hours. As expected, there is a trade-off between the power consumption and the ID data emission rate.

The developed ID tracking module is composed of a RISC 8-bit microcontroller (C8051F125), the same as the motion-sensing module and a CMOS image sensor (M64283FP, Mitsubishi Co., black-white 128x128pixel) with an infrared-filtered 90-degree lens. The microcontroller reads the image sensor signal every 2.5 msec, i.e., a 400Hz frame rate, by using the 8 x 8 pixel area windowing. Each ID signal is detected within 100 msec (i.e., 10 Hz). When one or more IrID Tags are recognized, the ID tracking module sends to the mobile PC the ID of the detected tag and the coordinates within its view, x and y (0~127), via a serial communication port.

The spatial field-of-view is 4 m x4 m at a distance of 2 m when using the 90 degree lens, and the spatial resolving power is 31.25 mm at that distance.

These wearable devices- an IrIrID Tag and IrID Tracker- can detect relative positions in real-time as an "interaction scope" that is defined by distance and angle to the object in a physical space.

Visual acuity is the ability of the eye to read characters or facial expressions, which is a significant factor in defining the possible spatial range of interactive activities involving objective humans or artifacts [7]. General scope can be defined by resolution when an object is an artifact that has visual information resources, such as a poster or a display.

Scope for an objective person is determined not only by resolution but also social relation. To define scope for a person, we observed small group interaction activities in a workspace (Figure 4) and defined the shape of the scope that surrounded almost every observed interaction distance and angle (Figure 5). These scopes were generalized and assembled as a radiation pattern of the IrID Tag. Figure 4 shows the relative position of three people by using a high-accuracy multi-camera Vicon (R) motion tracker.

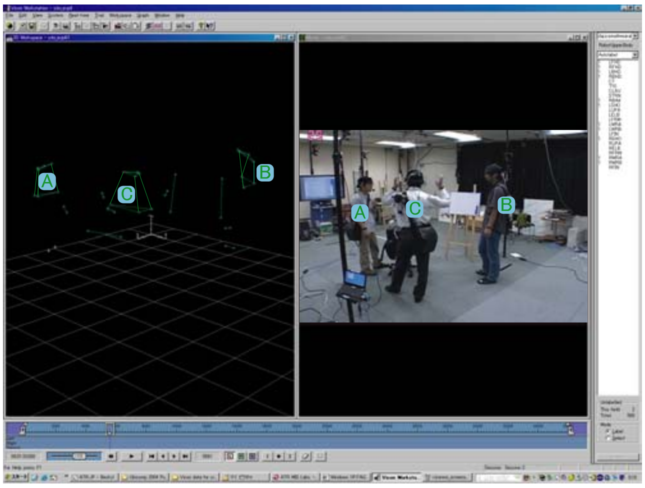


Figure 4. Observing interaction activities

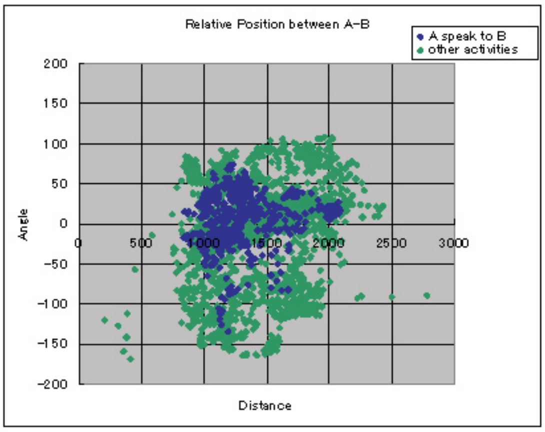


Figure 5. Relative position between A and B

The observer's view field information is important for determining which tag is observed. We chose the head direction as the view field direction because tracking the actual view field in everyday life would require unacceptable costs. The angle and shape of the view field is defined by the observed interaction activities in the workspace (Figure 5).

3. CONCULUSION

We have developed a wearable sensing system to detect and collect what a user experiences in an environment with multiple, pervasive self-identified devices. The designed wearable system can be used as a basic framework for our on-going researches into capturing human interactions.

The activity sensing system can detect two poses and five basic motions of one subject with the average 92% recognition ratio. The head-worn type interaction sensing module capture the visual data that the user sees with recognized multiple IrID Tags in the same view.

For the future works, we are evaluating the activity sensing method for multiple subjects in various environmental conditions like different shoes, shape of floor and stairways. We also plan to collect the massive interaction corpus by using the developed interaction sensing system in a events like Open Lab.

ACKNOWLEDGEMENTS

The work was partly supported by the National Institute of Information and Communications Technology, Japan. We would like to thank Megumu Tsuchikawa, Yasuhiro Katagiri, and members of ATR Media Information Laboratories, for their continuous support. We also would like to thank Norihiro Hagita at ATR Intelligent Robotics and Communications Laboratories for valuable comments.

REFERENCES

1. Sumi, Y., Matsuguchi, T., Ito, S., Fels, S., Mase, K.: Collaborative capturing of interactions by multiple sensors. *Adjunct Proceedings (poster presentation) Fifth International Conference on Ubiquitous Computing (UbiComp2003)*. (2003) 193–194
2. Lee, S.W., Mase, K.: Activity and location recognition using wearable sensors. *IEEE Pervasive Computing* **1** (2002) 24–32
3. Bao, Ling, Intille, S. S.: Activity Recognition from User-Annotated Acceleration Data. *Proc. 2nd Intl' Conf., PERVASIVE2004*. (2004) 1-17
4. Najafi, B., Paraschiv-Ionescu, K.A.A., Loew, F., Bula, C.J., Robert, P.: Ambulatory system for human motion analysis using a kinematic sensor: Monitoring of daily physical activity in the elderly. *IEEE Trans. on Biomedical Engineering* **50** (2003) 711–723
5. K. V. Laerhoven and O. Cakmakci, What shall we teach our pants? *Proc. of The Fourth Int'l Symposium on Wearable Computers*, (ISWC 2000), pp. 77-83.
6. Moore, D.J., Want, R., Harrison, B.L., Gujar, A., Fishkin, K.: Implementing phicons: Combining computer vision with infrared technology for interactive physical icons. *Proc. ACM UIST'99*, . (1999) 67–68
7. J. Gibson and A. Pick: Perception of another person's looking behavior, *American Journal of Psychology*, 76, 386-394, 1963.
8. Ito, S., Iwasawa, S., Kogure, K., Hagita, Sumi, Y., Mase, K.: InteractionScope: Non-fixed Wearable Positioning for Location-aware System, *Adjunct Proceedings (poster presentation) of the Sixth International Conference on Ubiquitous Computing (Ubicomp 2004)*, September 7-10, Nottingham, England. (2004)