Evaluation Study on Sensor Placement and Gesture Selection for Mobile Devices

Kazuya Murao
Graduate School of Engineering, Kobe University
1-1 Rokkodai-cho, Nada, Kobe, Hyogo 657-8501, Japan
murao@eedept.kobe-u.ac.jp

Tsutomu Terada
Graduate School of Engineering, Kobe University
1-1 Rokkodai-cho, Nada, Kobe, Hyogo 657-8501, Japan
tsutomu@eedept.kobe-u.ac.jp

Ai Yano
Fujitsu Laboratories, Ltd.
4-1-1 Kamikodanaka, Nakahara, Kawasaki, Kanagawa 211-8588, Japan
yano.ai@jp.fujitsu.com

Ryuichi Matsukura
Fujitsu Limited
4-1-1 Kamikodanaka, Nakahara, Kawasaki, Kanagawa 211-8588, Japan
r.matsukura@jp.fujitsu.com

ABSTRACT

Mobile phones and video game controllers using gesture recognition technologies enable easy and intuitive operations, such as scrolling a browser and drawing objects. However, usually only one of each kind of sensor is installed in a device, and the effect of multiple homogeneous sensors on recognition accuracy has not been investigated. Moreover, the effect of the differences in the motion of a gesture has not been examined. We have investigated the use of a test mobile device with nine accelerometers and nine gyroscopes. We have captured the data for 27 kinds of gestures for a mobile tablet. We experimentally investigated the effects on recognition accuracy of changing the number and positions of the sensors and of the number and kinds of gestures. The results showed that the use of multiple homogeneous sensors has zero or negligible effect on recognition accuracy, but that using an accelerometer along with a gyroscope improves recognition accuracy. They also showed that some gestures were not consistent among test subjects and interdependent, so selecting specific gestures to use can improve recognition accuracy.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous; H.5.1 [**Information interfaces and presentation**]: Multimedia information system—*Evaluation/methodology*; I.5.4 [**Pattern Recognition**]: Application—*Waveform analysis*

General Terms

Experimentation, Measurement

Keywords

Gesture recognition, Sensor, Mobile device, Accelerometer

1. INTRODUCTION

Many kinds of mobile devices containing small sensors have been released, and their applications and supporting services have attracted a great deal of attention. In particular, an accelerometer is installed in most current mobile devices, such as iPhone and Android-powered devices and controllers for the Wii and PS3 video game consoles, which enables easy and intuitive operations. Usually only one accelerometer is installed in commercial devices for energy saving and small footprint purposes. A high degree of accuracy is required since failure in gesture recognition degrades interface usability. The effect on accuracy of changing the number of sensors has not been investigated. Using multiple sensors increases the processing time for recognition, which would lead to less interface usability. In addition, whether recognition accuracy is affected by the number and kinds of gestures has not been studied.

We conducted an experiment capturing 27 kinds of gestures with 9 accelerometers and 9 gyroscopes and investigated the effect on recognition accuracy of changing the number of sensors and their positions and changing the number and kinds of gestures.

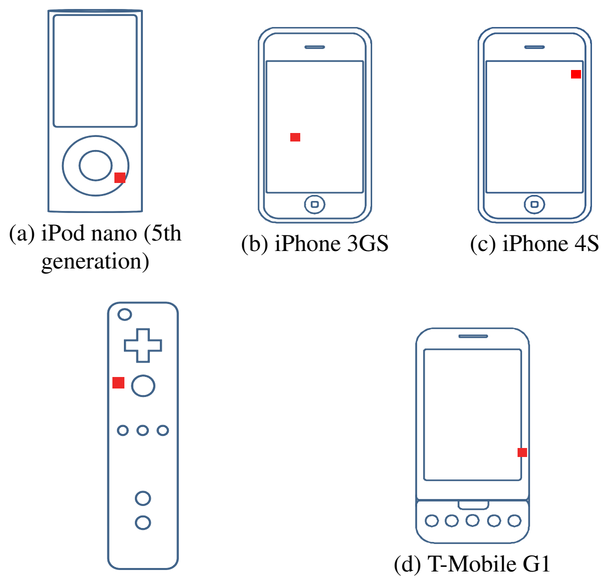
This paper is organized as follows. Section 2 describes the factors that affect recognition accuracy. Section 3 describes the experiment conducted to evaluate the effect of the number and position of sensors and of the number and kinds of gestures. The results are presented and discussed in Section 4. An application using gesture recognition technology is described in Section 5. The key points are summarized and future work is mentioned in Section 6.

2. FACTORS IN GESTURE RECOGNITION ACCURACY

The important factors affecting gesture recognition accuracy include the number of sensors, position of sensors, number of gestures, and kinds of gestures. While the form of the device is an important factor in terms of affordance, it is ignored here since it is not realistic to change the form of the device for recognition accuracy without taking its design into consideration.

Table 1: Sensors commonly installed in commercial devices.

Device	Number and kind	Manufacturer
Apple iPod nano (5th generation)	One 3-axis accelerometer	STMicroelectronics
Apple iPhone 3GS	One 3-axis accelerometer	STMicroelectronics
Apple iPhone 4S	One 3-axis accelerometer	STMicroelectronics
HTC Dream (T-Mobile G1)	One 3-axis accelerometer One 3-axis geomagnetic sensor	Asahi Kasei Microsystems
Nintendo Wii remote	One 3-axis accelerometer	STMicroelectronics
Nintendo Nunchuk	One 3-axis accelerometer	Analog Devices, Inc.
Nintendo Wii motion plus	One 2-axis gyro One 1-axis gyro	INVENSENSE EPSON TOYOCOM



(a) iPod nano (5th generation) (b) iPhone 3GS (c) iPhone 4S

(c) Wii remote (d) T-Mobile G1

Figure 1: Positions of sensors installed in commercial devices.

2.1 Number and positions of sensors

Table 1 summarizes the number and kinds of sensors commonly installed on commercial devices. Apple's iPod and iPhone and Nintendo's Wii remote and Nunchuk use one 3-axis accelerometer. Nintendo's Wii motion plus, a device for enhancing the Wii remote, contains a 2-axis gyro and a 1-axis gyro, which together detect 3-axis angular velocity. HTC Dream (also marketed as T-Mobile G1 in the US and parts of Europe), Android-powered phone, has a 3-axis accelerometer and a 3-axis geomagnetic sensor. One device usually has only one sensor of each kind, and, as far as we know, no device with multiple homogeneous sensors has been released.

The positions of the sensors in the iPod, iPhone 3Gs, iPhone 4S, Wii remote, and HTC Dream are shown in Figure 1 with ■ marks. The accelerometer and geomagnetic sensor are installed at the same position in the HTC Dream. As shown in the figure, there is no consistent positioning of the sensors. Sensor positions are not coordinated, especially in the iPod, iPhone 3GS, and iPhone 4S. This indicates that the positions probably depend on hardware limitations.

Murao et al. [?] measured the recognition accuracies of daily movements for all sensor combinations of five 3-axis accelerometers attached to a subject's wrists, ankles, and hip. The results show that recognition accuracy depends on the number of sensors and their combination. Therefore, the effect of the number and position of sensors should be taken into consideration in addition to energy consumption, implementation space, and wiring when developing new devices.

2.2 Number and kinds of gestures

The number and kinds of gestures also affect recognition accuracy. Many systems using gesture recognition technology with accelerometers have been proposed. The method proposed by Graeme et al. [?] annotates video-recorded activities by gesture recognition using one accelerometer mounted on the wrist since annotating video is difficult only by analyzing video. It uses hidden Markov models (HMMs) [?] for the recognition, resulting in only one mistake in 30 trials for three kung-fu martial art movements: cut, elbow, and punch.

The system proposed by Liu et al. [?] recognizes eight gestures including drawing a line or a circle (recommended by Nokia laboratories), with one 3-axis accelerometer. They captured more than 4,000 samples for 8 test subjects over a long period, using dynamic time-warping (DTW) [?] as a recognition algorithm. Accuracy of 98.6% was achieved by successively renewing the training data.

The system proposed by Junker et al. [?] recognizes ten daily short actions, such as pushing a button and drinking, and achieves approximately 80% precision and recall. The innovative feature of this system is that it partitions the stream of sensor data into several segments that represent atomic human movement by using the sliding-window and bottom-up (SWAB) algorithm [?].

The Georgia tech gesture toolkit [?] has been proposed as a tool to support gesture recognition. The toolkit enables ordinary users to use the HMM toolkit [?] with ease, which facilitates using HMM but requires specialized knowledge of speech recognition. One of its applications, gesture recognition with a 3-axis accelerometer on the wrist and on the elbow, achieved 93.3% accuracy for ten kinds of gestures such as grinding and sawing.

In these studies, the number of gestures to be recognized was not many, and the effect of the number and kinds of gestures was not reported. Though increasing the number of gestures to be recognized would enhance the interface, recognition accuracy would drop. Our investigation clarified the effect of the number and kinds of gestures, enabling application engineers to better select gestures used in applications.

3. EXPERIMENT

This section describes the gesture recognition experiment we conducted to investigate the effect on recognition accuracy of changing the number of sensors and their positions and of changing the number and kinds of gestures.

3.1 Setup and procedure

Data on 27 kinds of gestures (Table 2) that could be performed using a mobile tablet such as the iPad were captured 10 times for each gesture for 8 subjects (7 male and 1 female) 21–26 years old. A total of 2,160 samples were collected using 9 sensors placed on a test board, as shown in Figure 2. The board was W117×H155×D16 [mm] and weighed 200 [g]. The sensors were WAA-006 sensors, made by Wireless-T Inc.[1] Each sensor comprised a wireless 3-axis accelerometer and a 3-axis gyroscope. All the subjects were right-handed and performed the gestures while holding the lower right corner of the board with their right hand. The sampling frequency was 50 [Hz]. To reduce errors due to individual interpretation, the instructions were not given verbally. Instead, one of the authors demonstrated the actual movement for each gesture.

3.2 Preprocessing

Since our objective was to investigate the effect of sensors and gestures, the subjects were instructed to stand still before and after each gesture to mark the starting and ending points of each gesture. The waveform for each gesture was then extracted by checking for displacements in the captured data. If $x(t) - \overline{x}(t)$, where $x(t)$ is a captured data point and $\overline{x}(t)$ is the moving average, exceeded a threshold ϵ, the system detected a starting point. Otherwise, the system judged that the subject was not gesturing. We set ϵ to 200 [mG] since the fluctuation in the data while the subjects were stationary was up to 100 [mG]. Since the current value of $x(t)$ might have temporarily entered the region of $x(t) \pm \epsilon$ even while the subject was gesturing, the gesture was judged to have ended only after $x(t)$ had been within the region for more than 0.25 [s]. This interval was based on the results of pilot studies.

The data for one gesture consisted of 54 sequences (9 sensors × 2 kinds × 3 axes). The collected data were manually labeled.

3.3 Recognition

Time-series data is widely used in various fields such as science, medicine, economics, and engineering. Calculation of the similarity between time-series data is required in order to perform data-mining. A simple approach to measuring similarity is to use the Euclidean distance, but the results are susceptible to temporal distortion, and the number of samples in two data sequences must be equal.

A better approach is to use dynamic time-warping (DTW). The DTW algorithm can be used to calculate the temporal non-linear elastic distance between two sequences, which may vary in time, and the number of samples in the sequences need not be equal. For example, it can be used to find the similarity between two data sequences for *draw a circle in the air* when the rotation speeds differed.

The HMM is also famous for a method to treat time-series data and some studies use it for gesture recognition. HMM outputs likelihood of unknown gestures for predefined gestures, while DTW outputs spatial distance between gestures. Since distance is better index than likelihood to measure the difference between time-series, we adopted DTW in this paper.

The algorithm works as follows. For the sake of simplicity, we assume there is data for one axis. Testing data $X = (x_1, \cdots, x_m)$ and training data $Y = (y_1, \cdots, y_n)$, with lengths m and n, respectively, are compared, and an $m \times n$ matrix $\boldsymbol{d}$ is defined: $\boldsymbol{d}(x_i, y_j) = |x_i - y_j|$. Next, warping path $W = (w_1, \cdots, w_k)$, which is the path of the pairs of indices of X and Y, is found. W meets three conditions:

[1]Wireless Technologies, Inc. http://www.wireless-t.jp

Table 2: List of gestures

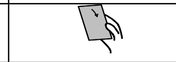
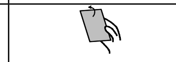
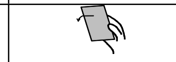
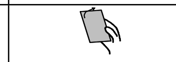
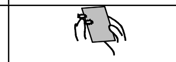
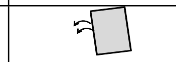
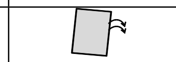
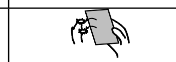
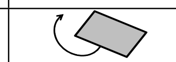
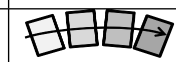
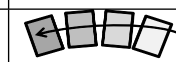
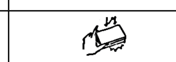
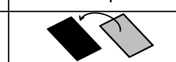
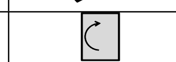
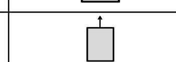
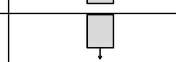
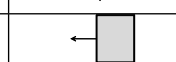
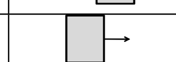
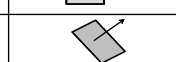
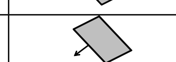
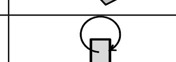
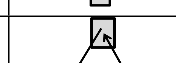
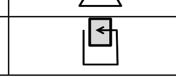
ID	Description	Illustration
1	Tilt to the near side	
2	Tilt to the far side	
3	Tilt to the left side	
4	Tilt to the right side	
5	Tap upper side twice	
6	Tap left side twice	
7	Swing twice to the left side quickly	
8	Swing twice to the right side quickly	
9	Shuffle cards	
10	Tap lateral edge as though sifting	
11	Scoop	
12	Lay cards	
13	Gather cards	
14	Rap table with the longer lateral edge	
15	Rap table with the surface of the board	
16	Knock the board twice	
17	Turn the board over	
18	Rotate clockwise on the table	
19	Shift up	
20	Shift down	
21	Shift left	
22	Shift right	
23	Shift diagonally up	
24	Shift diagonally down	
25	Draw a circle	
26	Draw a triangle	
27	Draw a square	

Figure 2: Experimental board with nine sensors.

- Boundary
  $w_1 = (1,1), w_k = (m,n)$
- Seriality
  $w_k = (a,b), w_{k-1} = (a',b') \Rightarrow a - a' \leq 1 \wedge b - b' \leq 1$
- Monotony
  $w_k = (a,b), w_{k-1} = (a',b') \Rightarrow a - a' \geq 0 \wedge b - b' \geq 0$

To find the path with the lowest cost that meets these conditions, the following steps are applied.

1. Initialization

   $DTW(0,0) = 0$
   $DTW(i,0) = \infty$ for $i = 1, \cdots, m$
   $DTW(0,j) = \infty$ for $j = 1, \cdots, n$

2. Do for $i = 1$ to m
   Do for $j = 1$ to n

$$DTW(i,j) = d(x_i, y_j) + min \begin{cases} DTW(i-1, j-1) \\ DTW(i-1, j \quad\;\;) \\ DTW(i \quad\;\;, j-1) \end{cases}$$

3. Output

   Return $DTW(m,n)/n$

The obtained cost $DTW(m,n)$ is the distance between X and Y. The returned $DTW(m,n)$ is divided by n since the DTW distance increases with the length of the training data.

The system recognizes gestures on the basis of the training data. The distances for all the training data are calculated, and the training data with the shortest distance is identified. A gesture labeled with the training data is then output. The DTW algorithm can be used for multiple axes of accelerometers and gyroscopes. The DTW calculation is carried out for each axis, and the sum of the distances for all axes is used as the distance of the gesture.

In our evaluation, recognition accuracy was analyzed in two ways: intra-subject and inter-subject. Intra-subject means the average accuracy over the eight subjects after four-fold cross-validation was conducted for each subject. Inter-subject means the use of one subject's data for testing and the other subjects' data for training, with the average accuracy being measured over all subjects.

4. RESULTS AND DISCUSSION

4.1 Number and positions of sensors

4.1.1 Number

The recognition accuracy on average over the gestures, the maximum accuracy, and the minimum accuracy for each number of sensors are shown in Table 3. The rows show the results for accelerometer only, gyroscope only, and both. Tables 3(a) and 3(b) are for intra-subject analysis and inter-subject analysis, respectively. For example, for two sensors, the number of combination was ${}_9C_2 = 36$. "Ave." shows the average accuracy over the 36 combinations, and "Max." and "Min." show the maximum and minimum accuracies among the 36 combinations, respectively.

These results show that the accuracies remained almost the same when the number of sensors was increased. Comparison of the intra-subject and inter-subject results shows that the intra-subject ones were much better. This is because the training data included only data for the subject making the gestures. Comparison of the results for the two kinds of sensors shows that the gyro was slightly inferior to the accelerometer and that using both an accelerometer and gyro improved accuracy. This is related to the kinds of gestures, which is discussed in Section 4.2.

The calculation cost increases with the number of sensors. The operation time for recognizing one gesture was 30 [ms] with one sensor and more than 500 [ms] with nine accelerometers and nine gyroscopes. The number of samples for training data was one for each gesture (total of 27 samples). The device used was an Android-powered phone (HTC One X; Android 4.0 OS; NVIDIA Tegra 3 (AP33/XMM6260) Quad-core 1.5 GHz) CPU. Even though the operation time could be longer if other applications are running in the background or the CPU is poor, the delay with one sensor would not be significant, but the delay with multiple sensors would likely irritate the user and lead to less usability.

4.1.2 Positions

To evaluate the positions of the sensors, the accuracy of one sensor for each position was measured. The results are shown in Figure 3. The horizontal axis shows the sensor positions corresponding to the numbers in Figure 2, and the vertical axis shows recognition accuracy. The results for the intra-subject analysis show 0.98–0.99 accuracy with negligible differences, between the number and kind of sensor. The results for the inter-subject analysis show that the accuracy was lower and that the differences between the positions were negligible. These results are supported by the data in Table 3: the range of accuracies among the same number of sensors was quite small.

There was, however, an approximately 0.01 difference in accuracy between sensors 3 and 4 for intra-subject analysis (Fig. 3(a)) even though these sensors were placed on the same board and in the same direction. To investigate this, we plotted the accelerometer waveforms for two samples of the "Gather cards" gesture, for which the difference was the greatest.The acceleration values for the x-axis, y-axis, and z-axis for two "Gather cards" gestures performed by the same subject are plotted in Figures 4(a), (b), and (c). respectively. The "Gather cards" gesture is sliding the board across a table as though gathering up cards, and the direction of movement corresponds to the Z-axis. Figure 4(c) shows the acceleration values for the Z-axis for sensors 3 and 4 for two samples of the gesture, resulting in similar waveforms. Since the gesture does not move in the X-axis and Y-axis direction, both values should be small. However, as shown in Figures 4(a) and (b), there was inconsistent noise in the data for sensor 3 while the values for sen-

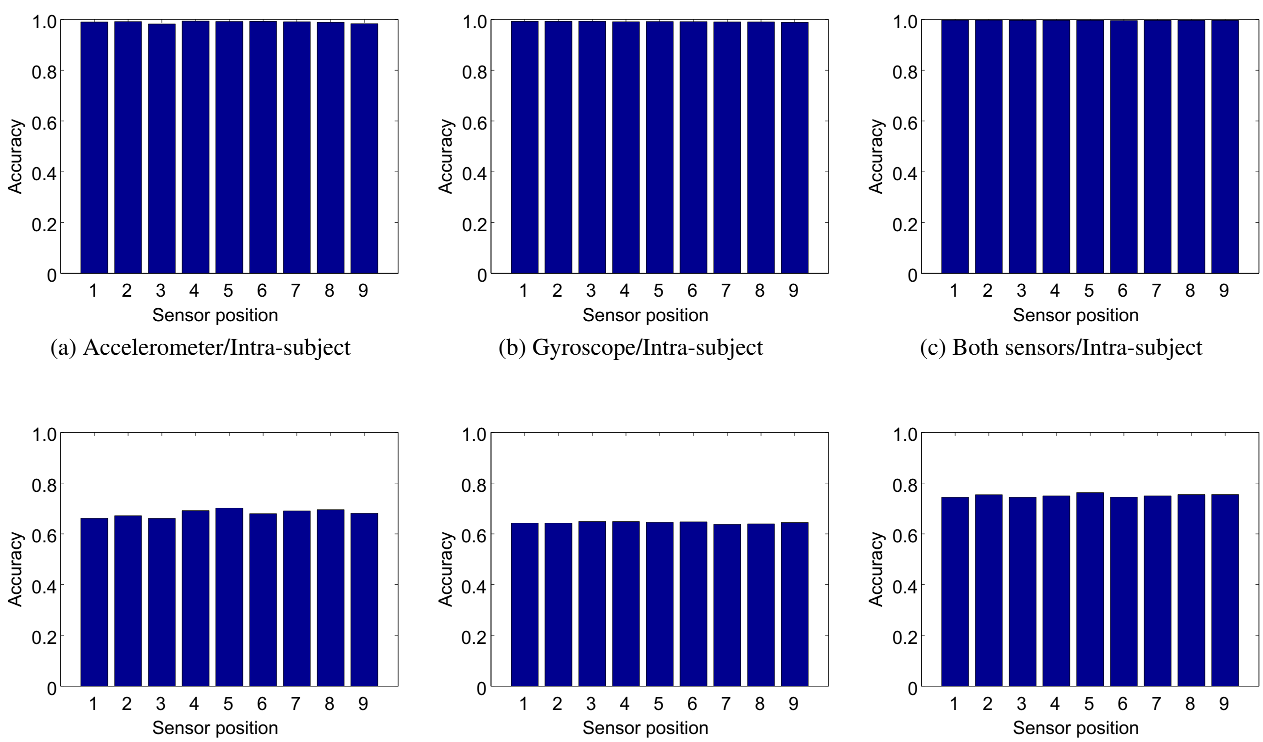
Table 3: Recognition accuracy vs. number of sensors.

(a) Intra-subject analysis

		Number of sensors								
		1	2	3	4	5	6	7	8	9
Accelerometer	Ave.	0.989	0.992	0.992	0.993	0.993	0.993	0.994	0.994	0.994
	Max.	0.994	0.995	0.995	0.995	0.995	0.995	0.995	0.994	0.994
	Min.	0.982	0.985	0.985	0.988	0.989	0.991	0.992	0.993	0.994
Gyroscope	Ave.	0.991	0.992	0.992	0.992	0.992	0.992	0.992	0.992	0.992
	Max.	0.993	0.994	0.993	0.994	0.993	0.993	0.992	0.992	0.992
	Min.	0.988	0.990	0.990	0.990	0.990	0.991	0.991	0.991	0.992
Both sensors	Ave.	0.998	0.998	0.998	0.998	0.998	0.998	0.998	0.998	0.998
	Max.	0.998	0.998	0.998	0.998	0.998	0.998	0.998	0.998	0.998
	Min.	0.996	0.997	0.997	0.997	0.997	0.998	0.998	0.998	0.998

(b) Inter-subject analysis

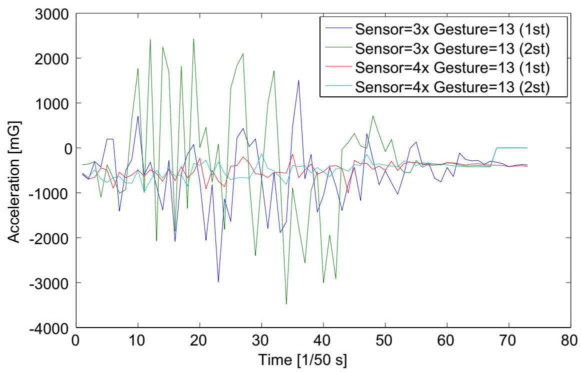
		Number of sensors								
		1	2	3	4	5	6	7	8	9
Accelerometer	Ave.	0.682	0.692	0.694	0.696	0.697	0.697	0.698	0.698	0.697
	Max.	0.702	0.702	0.704	0.705	0.706	0.704	0.703	0.701	0.697
	Min	0.661	0.674	0.682	0.687	0.688	0.691	0.693	0.695	0.697
Gyroscope	Ave.	0.638	0.639	0.639	0.639	0.639	0.639	0.639	0.639	0.640
	Max.	0.642	0.646	0.644	0.644	0.643	0.644	0.642	0.642	0.640
	Min.	0.631	0.629	0.632	0.634	0.633	0.635	0.635	0.637	0.640
Both sensors	Ave.	0.751	0.752	0.753	0.753	0.753	0.753	0.752	0.752	0.753
	Max.	0.763	0.762	0.759	0.760	0.757	0.756	0.756	0.755	0.753
	Min.	0.744	0.740	0.744	0.746	0.747	0.748	0.749	0.750	0.753



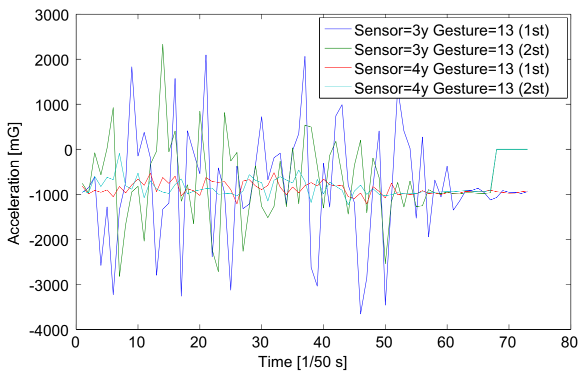
(a) Accelerometer/Intra-subject (b) Gyroscope/Intra-subject (c) Both sensors/Intra-subject

(d) Accelerometer/Inter-subject (e) Gyroscope/Inter-subject (f) Both sensors/Inter-subject

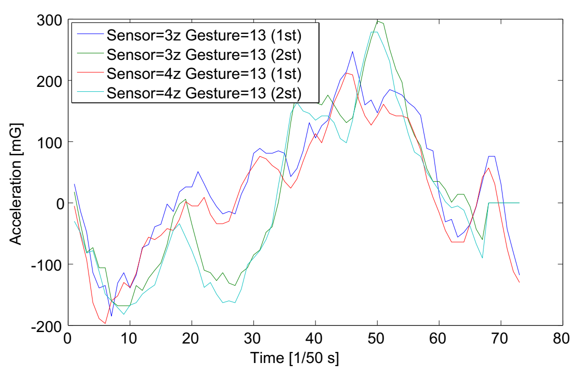
Figure 3: Recognition accuracy vs. sensor position.



(a) X-axis



(b) Y-axis

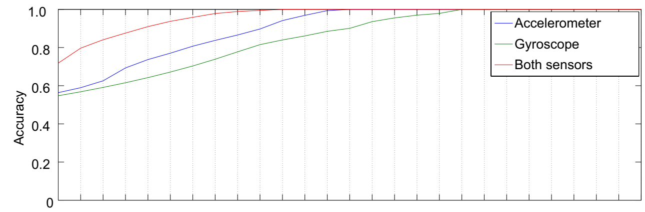


(c) Z-axis

Figure 4: Two samples of "Gather cards" gesture for sensors 3 and 4.

sor 4 were small. This noise appeared only for the "Lay cards" and "Gather cards" gestures, in which the board scraped the table. The data for sensor 9 included the same noise. This noise appeared only when the gestures were being performed. It might have been caused by the pulsation of the scraping amplified in proportion to the distance from the table surface. Further investigation will be done to clarify this.

The gyroscope results show uniform accuracy against sensor position. Examination of gyration waveforms did not reveal any noisy data. While some may think that shifting gestures do not involve gyration components, shifting gestures are actually not parallel shifting. They are sectoral shifting centered about the wrist or elbow, which produces angular values, resulting in accurate recognition with a gyroscope. Using an accelerometer and gyroscope together improves accuracy because they covers their drawbacks each other.



No.	27	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
Acc	-	5	6	8	15	20	22	23	19	21	11	7	27	16	-	-	-	-	-	-	-	-	-	-	-	-	-
Gyro	-	5	6	9	11	15	19	20	22	7	21	16	23	4	13	26	8	2	25	-	-	-	-	-	-	-	-
Both	-	8	11	19	5	23	7	22	12	4	26	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Number of gestures and excluded gestures

Figure 5: Accuracy vs. number of gestures.

4.1.3 Summary

In short, using multiple homogeneous sensors is not meaningful much for accuracy. The sensor position does not affect accuracy much. Using an accelerometer and gyroscope together slightly improves accuracy.

4.2 Number and kinds of gestures

4.2.1 Number

We calculated the overall accuracy as we reduced the number of gestures one by one, starting from one with the lowest accuracy. The results are plotted in Figure 5; the gestures excluded at each step are shown beneath the graph by sensor type for one test subject.

Specifically, 26 gestures were subject to recognition after excluding gesture 5 since gesture 5 had the lowest accuracy of the 27 gestures. (In this evaluation, sensor 5 was used for the accelerometer, gyroscope, and both sensors.) The accuracy for the accelerometer, gyroscope, and both sensors reached 1.0 when the number of gestures was respectively 13, 8, and 16.

The results for the accelerometer show that, though the average accuracy for 27 gestures was 0.5+, an average accuracy of 0.9+ could be obtained for as many as 16 gestures by selecting ones that are easily recognized, and for as many as 22 by also using a gyroscope. The gestures excluded for the accelerometer, gyroscope, and both sensors were the tapping and shifting gestures. These gestures are characterized by inconsistency, as mentioned in the previous section. These results are for one test subject, but the same tendency was observed for the other subjects.

4.2.2 Kinds

The recognition accuracies for each gesture with sensors are shown in Table 4.

From the intra-subject analysis, the accuracies of all gestures were virtually 1.0 regardless of sensor kind. This is because the training data included only data for the subject making the gestures. Moreover, the results between trials were consistent. From the intra-subject analysis for accelerometer, the accuracies for the "Lay cards" (12) and "Gather cards" (13) gestures were lower than those of the other gestures. This is because noise appeared in the data for these gestures, as discussed in Section 4.1.2.

The inter-subject accuracy was much less since the training data did not include data for the subject making the gestures. It was almost zero for gestures 5 and 6. These gestures were recognized

Table 4: Accuracy for each gesture with sensor 5.

Gesture	Intra analysis			Inter analysis		
	Acc	Gyro	Both	Acc	Gyro	Both
1	0.974	0.987	0.987	0.804	0.744	0.830
2	1.000	1.000	1.000	0.692	0.736	0.665
3	1.000	1.000	1.000	0.857	0.857	0.857
4	1.000	0.984	1.000	0.697	0.810	0.841
5	1.000	0.984	1.000	0.000	0.029	0.075
6	0.984	1.000	1.000	0.043	0.016	0.189
7	1.000	1.000	1.000	0.757	0.786	0.914
8	1.000	1.000	1.000	0.425	0.813	0.756
9	1.000	1.000	1.000	0.714	0.311	0.857
10	1.000	0.986	1.000	0.686	0.600	0.800
11	0.986	0.941	0.986	0.817	0.514	0.608
12	0.943	0.986	0.986	0.586	0.668	0.683
13	0.956	1.000	1.000	0.557	0.610	0.714
14	1.000	1.000	1.000	1.000	0.971	1.000
15	0.971	1.000	1.000	0.339	0.686	0.914
16	1.000	1.000	1.000	0.557	0.200	0.386
17	1.000	1.000	1.000	1.000	0.811	1.000
18	1.000	1.000	1.000	1.000	1.000	1.000
19	1.000	0.986	1.000	0.857	0.500	0.800
20	0.987	0.973	0.987	0.711	0.145	0.415
21	1.000	1.000	1.000	0.868	0.900	1.000
22	1.000	1.000	1.000	0.681	0.667	0.922
23	0.986	1.000	1.000	0.854	0.700	0.900
24	1.000	0.986	1.000	0.760	0.832	0.864
25	1.000	1.000	1.000	0.970	0.984	1.000
26	1.000	1.000	1.000	0.787	0.629	0.614
27	1.000	0.968	1.000	0.929	0.913	0.986
Average	0.992	0.992	0.998	0.702	0.646	0.763

as various gestures, such as "Swing twice to the left side quickly" and "Knock the board twice." This is attributed to the high degree of freedom of tapping. In short, tapping involves the tapping points, strength, and timing and the state of the hand holding the board (not moving, absorbing shock, or hitting the left hand with the board).

In addition, the accuracy for gesture 2 was less than that for gesture 1. This is because tilting to the far side is physically more difficult than tilting to the near side. The accuracy for gesture 22 was less than that for gesture 21. This is because all the subjects held the board with their right hand, and shifting right is harder than shifting left. Gestures involving rotation and shifting are simple and easy to use as interfaces, but they are affected by the orientation of the device and the state of holding it.

The gyroscope accuracies for gestures 19 and 20 were less than those of gestures 21 to 24. This is because shifting up and down is done parallel to the face of the board while shifting right, left, diagonally up, and diagonally down involve sectoral movement, which produces a rotation component that appears in the gyration values. Combining an accelerometer with a gyroscope improves accuracy by approximately 6% since each sensor obtained different components.

4.2.3 Summary

In short, some gestures had low reproducibility, which degraded recognition accuracy. The training for these gestures should thus be

Figure 6: A gesture-based browser.

done using data captured from many users or the user who is to use the device. An adequately high degree of accuracy can be achieved without excluding many gestures by selecting gestures that do not conflict with each other and that have high reproducibility.

5. APPLICATION

An example of an application that was enhanced with gesture recognition technology is the gesture-based browser shown in Figure 6. This browser uses several of the gestures used in our evaluation. For example, tilt invokes *page-up/down*, swing invokes *go back/forward*, rap invokes *stop loading/reload*, shift invokes *scroll/zoom*, and circle invokes *add bookmark*.

We do not think all the functions should be assigned to gestures. However, assigning many functions to the limited number of buttons forces users to select functions from a pop-up menu after pressing a "menu" button. A first level has six to eight choices at most, forcing users to a select "more" button to go to the next pop-up to select other functions. Such functions can be invoked with a single action by using a gesture.

Recent smartphones have a 5-inch display; for example, the GALAXY S III by SAMSUNG[2] has a 4.8-inch display and the ONE X by HTC[3] has a 4.7-inch display. Users sometimes use both hands to touch the display since the thumb usually does not reach the far side of the display. Even in this case, gesture-based interaction can be done with one hand (except for tapping and knocking) and does not interrupt other activities, such as eating and writing.

6. CONCLUSION

[2] http://www.samsung.com/uk/consumer/mobile-devices/smartphones/android/GT-I9300MBDBTU
[3] http://www.htc.com/uk/smartphones/htc-one-x/

We investigated the effect on gesture recognition accuracy of the number and positions of sensors and of the number and kinds of gestures for mobile devices. We experimentally evaluated the accuracies for 27 kinds of gestures measured using a board on which nine accelerometers and nine gyroscopes had been placed.

Using multiple homogeneous sensors did not improve accuracy by much. On the other hand, using both an accelerometer and gyroscope improved accuracy. Recognition accuracy did not depend on sensor position so much, but some gestures had noisy data depending on the sensor position.

Gestures involving shifting the side on which the device was held and tilting the far side of the device, which produced a small amount of movement due to physical constraints, and tapping, which had less consistency due to the high degree of freedom, were barely recognized with an accelerometer and gyroscope together. However, 0.9+ accuracy was achieved for 16 kinds of gestures with an accelerometer and 22 kinds of gestures with both an accelerometer and gyroscope by selecting gestures that do not conflict with each other and do have high reproducibility.

We plan to conduct further investigations of gestures.

7. ACKNOWLEDGMENTS

This research was supported in part by a Grant-in-Aid for Scientific Research on Priority Areas (21013034) of the Japanese Ministry of Education, Culture, Sports, Science and Technology.