Evaluating Gesture Recognition by Multiple-Sensor-Containing Mobile Devices

Kazuya Murao, Tsutomu Terada
Graduate School of Engineering, Kobe University
1-1 Rokkodai-machi, Nada-ku,
Kobe-shi, Hyogo 657-8501, Japan
{murao@ubi., tsutomu@}eedept.kobe-u.ac.jp

Ai Yano, Ryuichi Matsukura
Fujitsu Laboratories Ltd.
4-1-1 Kamikodanaka, Nakahara-ku,
Kawasaki-shi, Kanagawa 211-8588, Japan
{yano.ai, r.matsukura}@jp.fujitsu.com

Abstract

Mobile phones or video game controllers using gesture recognition technologies enable easy and intuitive operations. However, usually only one of each type of sensor is installed in each device, and the effect of multiple 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 captured data for 27 kinds of gestures by using a mobile device with 9 accelerometers and 9 gyroscopes, we then experimentally investigated the effects on recognition accuracy of changing the number and positions of sensors, and the number and kinds of gestures. The results showed that the use of multiple sensors and of sensors positioned at specific positions affects accuracy. It was also shown that gestures are interdependent and selecting specific gestures improves recognition accuracy.

1. Introduction

Many kinds of devices containing small sensors have been released, and their applications have attracted a great deal of attention. In particular, an accelerometer is installed in most current mobile phones, such as the iPhone and Android-powered devices, and video game controllers for Wii or PS3, 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 deteriorates usability of the interface. Though use of multiple sensors would improve accuracy, the effect on accuracy of changing the number of sensors has not been investigated. In addition, whether recognition accuracy is affected by the kinds and number of gestures has not been studied.

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

(a) iPod (b) iPhone (c) Wii remote (d) T-Mobile G1

Figure 1. Positions of sensors installed in commercial devices.

2 Factors in gesture recognition accuracy

2.1 Number and positions of sensors

The *iPod* and *iPhone* by Apple Inc. and the *Wii remote* and *Nunchuk* by Nintendo use a 3-axis accelerometer. *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. A sensor with a 3-axis accelerometer and 3-axis geomagnetic sensor is installed in *T-Mobile G1*, an Android-powered phone. One device usually has only one sensor of each type, and as far as we know, no device with multiple homogeneous sensors has been released. The positions of sensors installed in the iPod, iPhone, Wii remote, and T-Mobile G1 are shown in Figure 1 with ■ marks, respectively. These marks show there is no consistent positioning for sensors in these devices. Since sensor positions are not coordinated, especially in the iPod and iPhone, we suggest the positions of sensors extensively depend on the limitations of the hardware implementation.

One study[2] has measured the recognition accuracies of daily movements for all sensor combinations of five 3-axis accelerometers attached to both wrists, both ankles, and the hip of a subject. The results show that recognition accuracies change according to the number of sensors and the sensor combinations.

1550-4816/11 $26.00 © 2011 IEEE
DOI 10.1109/ISWC.2011.20

Table 1. List of gestures

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

2.2 Kinds and number of gestures

Many systems using gesture recognition technology with accelerometers have been proposed. The system proposed by Liu et al. recognizes 8 gestures, such as drawing a line and a circle (recommended by Nokia laboratory), with one 3-axis accelerometer, resulting in 98.6% accuracy by successively renewing training data[1]. The system proposed by Holger et al. recognizes 10 daily short actions, such as pushing a button and drinking, and achieves approximately 80% precision and recall[1]. The Georgia Tech Gesture Toolkit[5] has been proposed as a tool to support gesture recognition. One of its applications is gesture recognition with a 3-axis accelerometer on the wrist and on the elbow, and 93.3% accuracy is achieved for 10 kinds of gestures such as grinding and sawing. In these studies, the number of gestures to be recognized is not many and the influence of the number and kinds of gestures has not been reported.

3 Experiment

3.1 Experimental environment

Data on 27 kinds of gestures (Table 1) that we assume are performed when using a mobile tablet such as the iPad were captured 10 times for each gesture from 7 male and 1 female subjects aged 21–26 years. A total of 2160 samples were collected with 9 sensors placed on a board as shown in Figure 2. The board was W117×H155×D16 (mm) and weighed 200 g. The sensors used in the experiment were WAA-006, made by Wireless-T Inc.[1], in which a wireless 3-axis accelerometer and 3-axis gyroscope are installed. All the subjects were right-handed and performed the gestures while holding the lower right of the board with their right hands. The sampling frequency was 50 Hz. The subjects were instructed in how to perform the gestures by watching the actual movements demonstrated by one of the authors in order to reduce error due to individual interpretation.

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

Figure 2. Experimental board with sensors.

3.2 Preprocessing and recognition

The test subjects stood still before and after each gesture to indicate the starting and end points of the gestures. Then, the mean and variance over the sensor value of the gesture were extracted as feature values. In a comparison of the resultant feature values over the duration of the whole gesture and over the first 1 second, the latter were better since the durations of the gestures ranged from 1.2 to 2.0 seconds, and movements around the end points added noise. Therefore, the results of the features over the first 1 second are shown in this paper. The collected data were manually labeled. Dynamic time warping (DTW)[3] is often used as a gesture recognition algorithm since DTW measures the similarity of two time series in detail. DTW, however, scans all the data in a time series, which causes delay and disables immediate recognition. Therefore, we used a support vector machine (SVM)[4] after feature extraction in this work. In the evaluation, recognition accuracy was measured in two ways: intra-subject and inter-subject analysis. Intra-subject analysis measures the averaged accuracy over eight test subjects after four-fold cross-validation is conducted for each subject. Inter-subject analysis, in contrast, conducts four-fold cross-validation after the data for the eight subjects are merged.

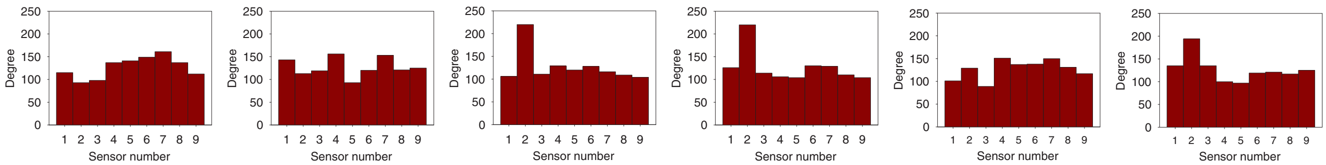
4 Results and consideration

4.1 Number and positions of sensors

Number. Recognition accuracy on average over the gestures for each number of sensors is shown in Table 2.

Table 2. Accuracy (mean±sd) vs. number of sensors.

Type	Sensor	Number of sensors 1	2	3	4	5	6	7	8	9
Intra	Acc	0.849±0.0130	0.891±0.00614	0.902±0.00379	0.907±0.00304	0.909±0.00256	0.910±0.00236	0.910±0.00157	0.910±0.00146	0.910±N/A
	Gyro	0.702±0.00983	0.737±0.0182	0.753±0.0174	0.761±0.0186	0.770±0.0188	0.778±0.0166	0.785±0.0140	0.792±0.0108	0.797±N/A
	Both	0.901±0.00628	0.908±0.00298	0.912±0.00243	0.913±0.00182	0.912±0.00169	0.912±0.00161	0.911±0.00188	0.911±0.00139	0.912±N/A
Inter	Acc	0.584±0.0294	0.701±0.0259	0.747±0.0161	0.775±0.0115	0.794±0.00769	0.806±0.00554	0.815±0.00409	0.824±0.00345	0.829±N/A
	Gyro	0.469±0.0168	0.521±0.0401	0.550±0.0417	0.574±0.0407	0.592±0.0387	0.609±0.0344	0.625±0.0297	0.642±0.0232	0.650±N/A
	Both	0.799±0.0118	0.847±0.0114	0.874±0.00947	0.893±0.00833	0.906±0.00634	0.915±0.00489	0.924±0.00449	0.930±0.00291	0.936±N/A



(a) Acc/Intra-subject (b) Acc/Inter-subject (c) Gyro/Intra-subject (d) Gyro/Inter-subject (e) Both/Intra-subject (f) Both/Inter-subject

Figure 3. Degree of sensor positions in combinations where accuracy was over median.

From the results, accuracy slightly increases but this growth is throttled as the number of sensors increases. In a comparison of intra-subject and inter-subject results, the intra-subject ones are much better since the training data include only the data of the subject for testing. However, inter-subject analysis achieves equivalent accuracy to that of intra-subject analysis when multiple homogeneous sensors are used. This means that multiple homogeneous sensors can capture the diversity among the subjects while one sensor cannot. Moreover, with regard to kinds of sensors, a gyro is inferior to an accelerometer. Detailed discussion on this point is in Section 4.2. The calculation cost also increases as sensors are added. In measuring the operation time to recognize 27 kinds of gestures with SVM, the recognition of 1 gesture including feature extraction took 0.3–0.4 msec with 1 sensor and 6.7 msec with 9 accelerometers and 9 gyroscopes. The device used was an Android-powered phone (HTC Desire; OS is Android 2.2 and CPU is Qualcomm QSD 8250 1 GHz). Though it may lengthen if other applications are running in the background or the CPU is poor, such a degree of delay would not be significant in this environment.

Position. To evaluate the positions of sensors, the appearance of sensors in the combinations where accuracy was over the that of median was counted for one to nine sensors. For the example of one sensor (nine potential positions), the appearance of sensors in the top four positions is counted. If sensor 1 is the best, the count for sensor 1 is incremented. This procedure is iterated for all nine sensors. A sensor that is counted frequently contributes to the improvement in accuracy. The results are shown in Figure 3. Though there is no clear trend for the accelerometer, sensors far from the holding position (corresponding with the position of sensor 3) were frequently assigned to the top 50%. This makes it clear that sensors near the points where shock and impact are given by the gestures are effective for recognition. The results for gyroscope 2 show that it plays an important role. This would be due to the rotation axes of each gesture crossing at the point of sensor 2, and the rotation of each gesture is sensed well at that point.

Summary. Using multiple homogeneous sensors improves accuracy when training with data from multiple users. A position further from the holding position is better for an accelerometer. A gyro placed on the cross-point of the axes of gyration for each gesture works well. Though using multiple sensors would be even more accurate, realistically we recommend placing accelerometers on the opposite side to the holding position and placing one gyro at the cross-point of axes of gestures, in the design of sensors in devices.

4.2 Kinds and number of gestures

Recognition accuracies for each gesture are shown in Table 3. The sensors used were those combinations that performed best.

Kind. From the intra-subject results, the accuracies of gestures 12 and 13 were quite low with any kind of sensor, even though the system was trained with data for the same subject for testing. This is because each trial for these gestures is not consistent. Moreover, the accuracies of gestures 19 to 24 with an accelerometer were inferior to those of the other gestures since these gestures involve small parallel displacements in common that rarely appear in the acceleration value. Focusing on only the gyro results, gestures 19 to 24 were not recognized well. This is because parallel displacements do not appear in angular velocity. In contrast, gestures 14 to 16, which include a shock, were recognized well by an accelerometer compared to a gyro since these gestures produce a large acceleration and do not produce a large angular velocity. The inter-subject accuracy dropped by a large margin, but the accuracies of gestures 12 and 13 were improved. It appears that the intra-subject training data did not capture the consistency among the gestures due to the training data being for one test subject only. On the other hand, the inter-subject results were consistent through

Table 3. Accuracy for each gesture.

Gestures	Intra analysis			Inter analysis		
	Acc	Gyro	Both	Acc	Gyro	Both
1	0.990	0.880	0.964	0.933	0.733	0.987
2	0.990	0.875	0.990	0.897	0.387	0.974
3	0.964	0.875	1.000	0.987	0.352	0.986
4	0.990	0.917	0.974	1.000	0.857	0.974
5	1.000	0.870	1.000	0.896	0.895	0.988
6	0.974	0.615	0.979	0.907	0.241	0.933
7	0.990	0.927	0.990	0.874	0.949	0.988
8	0.948	0.943	0.979	0.829	0.461	0.974
9	1.000	0.932	1.000	0.897	0.545	0.949
10	0.979	0.927	1.000	0.857	0.779	0.974
11	0.943	0.849	0.927	0.816	0.829	0.921
12	0.219	0.229	0.229	0.670	0.848	0.936
13	0.219	0.250	0.250	0.507	0.883	0.948
14	1.000	0.797	1.000	1.000	0.847	1.000
15	0.979	0.792	0.979	0.961	0.597	0.948
16	0.979	0.719	0.979	0.913	0.413	0.963
17	0.990	0.958	0.974	0.987	1.000	0.987
18	1.000	0.994	1.000	1.000	1.000	1.000
19	0.917	0.531	0.953	0.861	0.356	0.861
20	0.990	0.891	0.984	0.598	0.481	0.909
21	0.911	0.802	0.958	0.663	0.169	0.832
22	0.924	0.755	0.938	0.438	0.713	0.675
23	0.974	0.745	0.990	0.937	0.570	0.949
24	0.891	0.661	0.896	0.558	0.455	0.778
25	0.979	0.906	0.938	0.845	0.869	0.961
26	1.000	0.932	0.990	0.848	0.772	0.962
27	0.969	0.953	0.979	0.722	0.658	0.924
Average	0.915	0.797	0.920	0.830	0.654	0.936

Sensors used; Intra/Acc:2,5,6,7,8,9 Intra/Gyro:1,2,4,5,6,7,8,9 Intra/Both:4,5,6 Inter/Acc:1,2,3,4,6,7,8,9 Inter/Gyro:1,2,3,4,6,7,8,9 Inter/Both:1,2,3,4,5,6,7,8,9

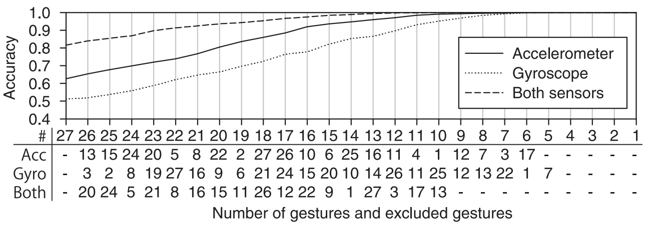


Figure 4. Accuracy vs. number of gestures.

training with the data of seven test subjects. Especially in the gyroscope results, the accuracies of gestures 12 and 13 were improved by a large margin since these gestures involve rotation.

Number. We measured the accuracy by reducing the number of gestures one by one starting from the gesture whose accuracy was the lowest overall. The accuracies are plotted in Figure 4, and excluded gestures at each step are shown beneath the graph, according to sensor type. Specifically, 26 gestures were recognized after excluding gesture 13 since gesture 13 had the lowest accuracy of the 27 gestures. In this evaluation, only one sensor was used: sensor 5 for the accelerometer, sensor 2 for the gyroscope, and sensor 5 for the both sensors. The accuracy reached 1.0 when the numbers of gestures were 6, 5, and 10 for the accelerometer, gyro, and both sensors, respectively. From the results for the accelerometer, though the average accuracy for 27 gestures is 0.6+, an average accuracy of 0.9+ can be obtained for as many as 16 gestures by selecting ones that are easily recognized. Focusing on the excluded gestures of the accelerometer and both sensors, one of two gestures that move in the opposite direction to each other is excluded, then the accuracy of the remaining gesture goes up. For example, with the accelerometer, the accuracy of gesture 23, which involves an opposite movement to gesture 24, is improved from 0.785 to 0.974 after excluding gesture 24 when the number of gestures is reduced from 25 to 24. These results indicate recognition accuracy does not simply depend on the number of gestures.

Summary. Gestures involving parallel displacement are rarely recognized. A gyro is not good at recognizing gestures including a shock, such as tapping. Some gestures have low reproducibility, which leads to deterioration in recognition accuracy. These gestures should be trained with data captured from many users. An adequately high degree of accuracy can be achieved without excluding a lot of gestures by selecting gestures that do not conflict with each other.

5 Conclusion

We investigated the effects on gesture recognition accuracy of the number and positions of sensors, and the kinds and number of gestures. We experimentally evaluated the accuracies of 27 kinds of gestures measured using a board on which nine accelerometers and nine gyroscopes were placed. Accelerometers and gyroscopes in devices are currently placed together, but we plan to investigate more effective positions by placing these sensors separately.

Acknowledgements

This research was supported in part by a Grant-in-aid for PRESTO from Japan Science and Technology Agency and Scientific Research(A)(20240009) from Japanese Ministry of Education, Culture, Sports, Science and Technology, and by a Grant-in-Aid for JSPS Fellows (21·249) of Japan Society of the Promotion of Science.

References

[1] Liu, j., Wang, Z., Zhong, L., Wiekramasuriya, J., and Vasudevan, V. uwave: Accelerometerbased personalized gesture recognition and its applications. In *PerCom 2009*, pages 1–9, 2009.

[2] Murao K., Terada T., Yoshinari T., and Shojiro N. A context-aware system that changes sensor combinations considering energy consumption. In *Pervasive 2008*, pages 197–212, 2008.

[3] Myers, C., and Rabiner, L. R. A comparative study of several dynamic time-warping algorithms for connected word recognition. In *The Bell System Technical Journal*, volume 60, pages 1389–1409, 1981.

[4] Vapnik, V. *The Nature of Statistical Learning Theory*. Springer, 1995.

[5] Westeyn, T., Brashear, H., Atrash, A., and Starner, T. Georgia tech gesture toolkit: Supporting experiments in gesture recognition. In *ICMI 2003*, pages 85–92, 2003.