A Recognition Method for Combined Activities with Accelerometers

Kazuya Murao
Ritsumeikan University
1-1-1 Nojihigashi, Kusatsu,
Shiga, 5258577 Japan
murao@cs.ritsumei.ac.jp

Tsutomu Terada
Kobe University
1-1 Rokkodaicho, Nada, Kobe
Hyogo, 6578501 Japan
tsutomu@eedept.kobe-u.ac.jp

UbiComp '14, September 13 - 17 2014, Seattle, WA, USA

ACM 978-1-4503-3047-3/14/09...$15.00.
http://dx.doi.org/10.1145/2638728.2641304

Abstract
Many activity recognition systems using accelerometers have been proposed. Activities that have been recognized are "single" activities which can be expressed with one verb, such as sitting, walking, holding a mobile phone, and throwing a ball. In actual, however, "combined" activities including more than two kinds of state and movement are often taken place. Focusing on hand gestures, they are performed not only while standing, but also while walking and sitting. Though the simplest way to recognize such combined activities is to construct the recognition models for all the possible combinations of the activities, the number of combinations becomes immense. In this paper, we propose a recognition method for combined activities by learning single activities only. Evaluation results confirmed that our proposed method achieved 0.84 of recall and 0.85 of precision, which is comparable to the method that had learned all the combined activities.

Author Keywords
Activity recognition, Combined activity, Accelerometer

ACM Classification Keywords
H.4 [Information Systems Applications]: Miscellaneous

Introduction

Along with the progress in wearable computing, many context-aware systems with accelerometers have recently been introduced. In the procedure of making an activity recognition system, developers generally define the objective activities, collect their data, annotate them, and construct the recognition models. Therefore, recognition results are limited within the predefined activities. In other word, if we want to recognize a combined activity *holding a mobile phone while walking*, training data for the activity has to be collected and annotated so, otherwise, recognition result would be either *walking* or *holding a mobile phone*.

Here, we define two types of activities; global activity and local activity. Global activity is a bodywide movement, such as *walking* and *standing*. Multiple global activities contradict each other at a time. Local activity is a movement of a specific parts of body, such as *throwing* and *holding something*. Multiple local activities coexist unless these are on the same part. The simplest way to recognize combined activity is to construct recognition models for all the combined activities, but the number of the possible combinations of global and local activities gets immense. Supposing five global activities; *standing*, *sitting*, *walking*, *running*, and *bicycling*, and ten local activities of hand gestures, data for the 50 patters must be collected. As one kind of hand gesture is added, data for the gesture performed during five global activities must be captured, and it is a backbreaking task. Moreover, considering foot gestures, the possible patters are global activity×local hand activity×local foot activity. However, these combined activities are not negligible since they are physically possible and might occur in our daily life.

Problem of combined-activity recognition is not only taking time for collecting training data, but also annotating data. After collecting all the data through long experiment, the data have to be annotated with ground truth. Figure 1 shows acceleration of a chop gesture while running and while standing. It is easy to find starting point and end point for *chop while standing*, whereas it is hard to trim it exactly by visual check for *chop while running*. Someone might think that recognizing activity of each part of body individually and integrating them produce correct activity. However, as shown in Figure 1, a gesture while running is different to that while standing. Moreover, arm swing of running is included in the beginning and ending of the gesture unless the gesture is trimmed with a tightfitting window. Motion of running is propagated, the gesturing is slightly different, and part of running motion is included before and after the gesture, which leads to misrecognition.

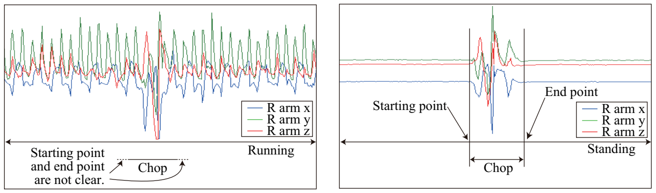

Figure 1: Waveform of an accelerometer mounted on the right wrist: a chop gesture while running (left) and while standing (right).

Our previous work tackled on gesture recognition while moving. Gestures having been recognized in the research are not combined ones, but single ones which are smoothly changed from and to global activity, such as *walking*[3]. An effective method to treat combined activity has not been reported.

We developed the system that recognizes combined activities from training data of single activities, i.e. global activities and local activities while standing. The system classifies each part of the body to posture, behavior, and gesture, from fluctuation and autocorrelation in the acceleration data. In this work, we define posture, behavior, and gesture as follows: Posture is a state of a user remaining stationary lasting for a certain length of time, e.g. *sitting* and *standing*. Behavior is a state of a user doing periodical movement lasting for a certain length of time, e.g. *walking* and *running*. Gesture is not a state but once-off action having starting point and endpoint that sporadically occur, e.g. *punch* and *draw a circle in the air*.

In general, autocorrelation plot of a periodical wave shows high peaks. When parts of sensors are showing high peaks, the corresponding body parts are meant to have constancy and are classified to behavior, otherwise the parts are classified to gesture. Finally, the system outputs conclusive recognition result from the recognition results of each part. By using our system, combined activities such as *throw while walking* and *holding a mobile phone while running* are recognized only from single activities; *walking*, *running*, *throwing*, and *holding a mobile phone*.

Related work

Studies on activity recognition are listed in Table 1, however most of them focus on single activities, such as ambulation and posture. One study recognizes eight activities including *vacuuming* and *brushing teeth* with an accelerometer attached to the pelvic region[4]. Other study employs twenty-two kinds of sensors to recognize *lying*, *rowing*, *running*, *Nordic walking*, *bike*, *walking*, *sitting*, and *standing*[1]. Though *walking carrying items* activity seems combined activity, it cannot be separated and recognized by combining other activities, i.e. *running carrying items* activity and *holding items* activity cannot be recognized in this architecture.

Table 1: Activities recognized and sensors used in past work on activity recognition.

Ref.	Activities recognized	# of sensor	Sensor kind	Sensor position
[1]	ambulation, posture, scrubbing, vacuuming, folding laundry, brushing teeth, bicycling, eating or drinking, working on computer, walking carrying items	5	2D acc × 5	left elbow right wrist torso left knee right ankle
[2]	ambulation, posture, putting clothes, eating, rowing, bicycling, respiratory	22	air pressure, mic, pulse, ball sensor, light, switch, EKG, humidity, 3D acc × 2, temperature, respiratory, 3D compass × 2, GPS, heart rate × 3, SaO2 × 2, skin temperature, skin resistance	wrist upper back below neck finger armpit chest forehead shoulder
[4]	ambulation,posture, sit-ups, vacuuming, brushing teeth	1	acc	pelvic region
[5]	hammer, file, sand, saw, screw, vise, drill, clap, use driver, grind	2	acc × 2	wrist elbow
[6]	push button, handshake, phone up, phone down, cutlery, door, coin, drink, spoon, handheld	5	gyro × 5	wrists upper arms upper torso
[8]	draw '>' mark, square shift left to right, shift right to left, shift bottom to up, shift up to bottom, clockwise circle, anticlockwise circle	1	wii remote	hand

Georgia Tech Gesture Toolkit[5] as a tool to support gesture recognition has been proposed by Westeyn et al. This is a toolkit that enables ordinary users who do not have enough knowledge on speech recognition to use existing HMMtoolkit[1] with ease. In the literature, four applications are presented, one is gesture recognition with two 3-axis accelerometers positioned at wrist and elbow, and achieves 93.3% accuracy for ten kinds of gestures such as *grinding*, *sawing*, and *screwing*. The system proposed by Junker et al.[6] recognizes ten daily short actions, such as *pushing a button* and *drinking*, and achieves approximately 80% precision and recall.

[1]HTK Hidden Markov Model Toolkit home page. http://htk.eng.kcam.ac.uk/.

Innovative point of this study is that it partitions the stream of sensor into several segments that represent atomic human movement by using the sliding-window and bottom-up (SWAB) algorithm[7]. The method proposed by Liu et al. recognizes eight gestures such as *drawing a line* and *a circle*, which are recommended by Nokia laboratory, with one 3-axis accelerometer[8]. This research captures more than 4,000 samples from eight test subjects for a long period. They use Dynamic Time Warping algorithm (DTW)[9] as a recognition algorithm and achieve 98.6% accuracy by successively renewing training data.

Activities that have been targeted in these works are single activities. Combined activities have to be defined one by one. However, the number of possible combinations increases in multiple order of the number of global and local activities, causing much time to capture training data. Capturing ten global activities and ten 10-second local gestures for five times takes 5,000 seconds. This is an actual time of movement and more than double or triple time is needed considering interval and rest.

System structure

Posture and behavior are state lasting for a certain period of time and consist of periodic pattern of acceleration waveform. These activities are generally recognized with a classifier such as SVM[13] after converting raw data into mean, variance, fast Fourier transform (FFT) coefficient over a time window. This approach enables high-speed recognition since not all data in the window but feature value is used. Moreover, one of the advantages of using feature value is also that the recognition process does not have to consider which part of the movement is included in the window, e.g. beginning of the window does not have to fit the specific motion of steps of walking, since feature value discards temporal information.

On the other hand, gesture is a once-off action that has starting point and endpoint, which is different to posture and behavior. Feature-based approach cannot distinguish similar gestures such as *rotating arm clockwise* and *anticlockwise* since feature value does not have information on how it moved, therefore gesture must be recognized in a different way. In general, gestures are recognized with a template matching algorithm such as dynamic time warping (DTW) or a statistical model such as hidden Markov models (HMM), after trimming an actual movement from stream data. The conventional method forced the users to indicate gesture interval by pushing a button of a device or by standing still before and after the gesture[10]. It is hard to stop motions or perform specific gestures to indicate starting point. Taking out a device or keep holding a device to push button while performing gestures is also unrealistic. Recognizing a gesture without considering starting point and endpoint, misrecognition occurs or miss out the gesture buried in behavior. Our system classifies user activities into posture, behavior, and gesture for each body part, then apply DTW to gesture or apply SVM to posture and behavior.

Our system consists of three phases, as shown in Figure 2. The first phase classifies activities of each part of body into three types; posture, behavior, and gesture. The second phase recognizes activities according to the activity type. The third phase integrates the recognition results and outputs conclusive result.

In this paper, we assume that user attaches five accelerometers on their both wrist, hip, and both ankle. Activities are four postures (*sitting*, *standing*, *lying*, and *kneeling*), five behaviors (*walking*, *running*, *bicycling*, *descending stairs*, and *ascending stairs*), five hand-gestures (*chop*, *throw*, *punch*, *draw a clockwise circle*, and *draw a anticlockwise circle*), and two hand-postures (*holding a mobile phone*, and *raising a hand*). The sampling frequency is 20 [Hz], which is sufficient for activity recognition as reported in [11].

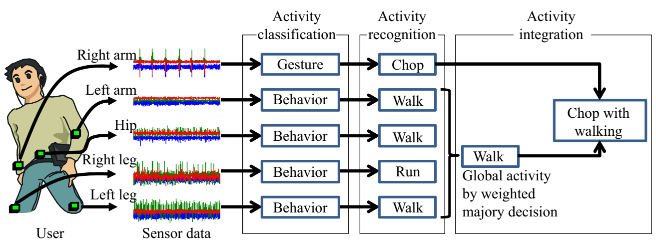

Figure 2: Recognition flow.

Activity classification
Displacement detection
Activity classification phase checks for displacements in the sensed data. Supposing time $t = T$ now, if a difference of moving average over 20-sample (1-second) sensed data $\overline{x}(T)$ and current value $x(T)$ runs over a threshold ε, our system detects a movement as a following equation. Otherwise, our system judges that the user is maintaining a posture.

$$\begin{array}{lcl} \text{if } |\mathrm{x(T)} - \overline{\mathrm{x}}(\mathrm{T})| > \epsilon & \Rightarrow & \text{Behavior or Gesture} \\ \text{otherwise} & \Rightarrow & \text{Posture} \end{array} \quad (1)$$

The region of $x(t) \pm \varepsilon$ is called the epsilon tube, which removes displacements. In this paper, $\varepsilon = max\{ave\{std(Leg)\}, 200\}$, where $Leg = \sqrt{leg(x)^2 + leg(y)^2 + leg(z)^2}[mG]$. std is a standard deviation over a window and ave is an average over both legs. The vibration of movement of legs is propagated to hand, which produces not-small values. While the movement of legs is not intense, ε is set to 200 [mG] since fluctuation produced while being stationary was up to 100 [mG]. Since the current value $x(t)$ might temporarily goes into the epsilon tube even while moving, posture begins only after $x(t)$ stays within the epsilon tube for more than 0.25 second. These values are obtained from our pilot studies. As shown in Figure 3, while the data is within the epsilon tube, the system judges that the parts of body are maintaining a posture. When the data indicates movement, this process goes on to constancy decision phase.

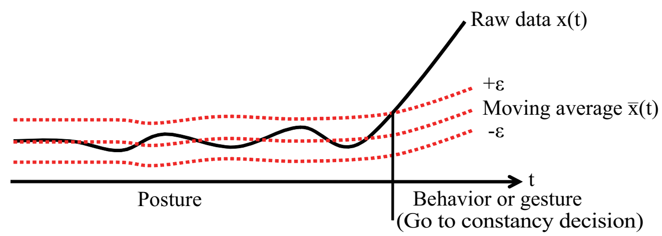

Figure 3: Displacement detection.

Constancy decision
Basically, data on walking include iterations in rhythm with the steps. On the other hand, gestures are once-off actions and do not have iterations. Note that we consider that the iterations of once-off actions are behaviors. In this phase, the autocorrelation function (ACF) finds iterations in the user's movements, and classifies the movement into behavior or gesture. The discrete ACF $Rxx(\tau)$ at lag τ for a data sequence $x(t)$ is defined as $Rxx(\tau) = \sum_{t=0}^{N-1} x(t)x(t-\tau)$, where N is the window size for the ACF calculation and set to 64 samples (3.2 seconds), which is long enough to capture at least two iterations. In addition, since ACF shows a maximum at $\tau = 0$, all the values of ACF are normalized by $R'xx(\tau) = Rxx(\tau)/Rxx(0)$ so that the range is $(-1, +1)$. The system has to decide whether the

movement has constancy or not. Figure 4 shows acceleration waveform and its ACF of *walk, chop, then walk again* activity. As shown in the figure, the ACF of walking shows clear peaks, whereas the ACF of chop does not have high peaks. Constancy is detected when the height of the first peak $R'xx(n)(n > 0)$ exceeds $\alpha \cdot (1 - n/N)$, where α is a coefficient set to 0.6. Reason n/N is used is that the height of the first peak linearly decreases as τ increases.

$$\begin{array}{lcl} \text{if } \mathrm{R}'_{\mathrm{xx}}(\mathrm{n}) \geq \alpha \cdot (1 - \mathrm{n}/\mathrm{N}) & \Rightarrow & \text{Behavior} \\ \text{otherwise} & \Rightarrow & \text{Gesture} \end{array} \tag{2}$$

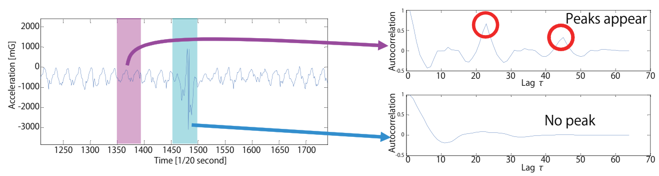

Figure 4: Accelerations of *chop gesture while walking* (left) and autocorrelation of *walking* (upper right) and *chop* (lower right).

Intille et al. focus on acquiring in-situ training data and mention that acceleration data of walking in laboratory displays consistent gait cycle. On the other hand, acceleration data of the same person outside the laboratory may display marked fluctuation in gait cycle and length[12]. Also we think that is true and the data obtained from same subject on different days are different. Although acceleration data fluctuate in the range of day or hour, data in the range of few seconds or shorter are significantly periodic, which could produce constancy. Moreover, since our approach is unsupervised and does not require training, influence of differences among individuals and users' conditions are small.

Activity recognition

Activity recognition phase recognizes activities of each part of body according to the activity types. For posture data, the mean value of the data in the window is calculated as a feature value and the posture is recognized with SVM[13] that has learned only postures. Since the variance of postures is almost zero, only the mean is used for the recognition. SVM operating on the mean and variance is used for behavior data, whereas DTW[9] operating on trimmed original wave over a window is used for recognizing gestures. SVM and DTW has learned behaviors only and gestures only, respectively.

Activity integration

Since each part of the body outputs activity recognition result individually, these have to be integrated in order to decide conclusive recognition result. Even if the user is just walking, recognition results of all parts are not necessarily *walking*. This section describes the method to integrate them.

In this paper, we assume that if recognition result for hand is either posture, hand-posture or hand-gesture, it will be a local activity, otherwise local activity will be null. Then global activity is decided by a weighted majority vote. We use recall of recognition result for training data as a weight. Each body part has one voting right, but it is weighed, then the weighted votes are summed up over the body except for local activity. For example, in Figure 2, the recognition results of the left arm, hip, right leg, and left leg are *walk*, *walk*, *run*, and *walk*, respectively. Suppose recall of *walk* recognized with a sensor on the left arm, hip, and left leg are 0.7, 0.8, 0.4, respectively, and recall of *run* recognized with a sensor on right leg is 0.4, poll to *walk* is 1.9 and poll to *run* is 0.4, resulting in walking as global activity. Finally, combined activity is output by merging global and local activity. Even if a part of body is doing different activity to global activity, user activity is correctly recognized. For *holding a mobile phone while walking* activity, the hand holding a phone is classified into posture and recognized as *holding a phone*, whereas remaining parts would be classified into behavior and recognized as *walking*.

Evaluation

In this section, we evaluate our system on the basis of recall and precision.

Setup

We evaluated our system for the data while *standing*, *sitting*, *walking*, *running*, and *bicycling*. The training and test data were taken from three male subjects aged 22 to 28 years, who wore five accelerometers[2] on their right wrist, left wrist, hip, right ankle, and left ankle. The sampling frequency was 20 [Hz]. Activities are as listed in Table 2: four postures (*sitting*, *standing*, *lying*, and *kneeling*), five behaviors (*walking*, *running*, *bicycling*, *descending stairs*, and *ascending stairs*), five hand gestures (*chop*, *throw*, *punch*, *draw a clockwise circle*, and *draw an anticlockwise circle*), and two hand postures (*holding a mobile phone* and *raising a hand*). The subjects were acted these hand gestures and hand postures while *sitting*, *standing*, *walking*, *running*, and *bicycling*. Each gesture was recorded 10 times for each global activity. All local activities while ascending and descending, and jump and kick while bicycling were not performed for safety's sake. The logged data were manually labeled. In this evaluation, activities were recognized with three method; SVM only, DTW only, and our method. As shown in Table 3, the first two methods are comparisons, simply trained with all the possible combinations. 20% of which were used for training, the remaining 80% were used for testing. The last one selectively uses SVM and DTW which learned single activities only and integrate the results. Most part of the test data consists of global activity and gestures sporadically occur. Correct recognition results for gestures which is output in one second from the gesture ends are accounted for true positives.

[2]Wirelss Technologies Inc.: http://www.wireless-t.jp/.

Results

Table 4 shows the recall and precision of recognition for hand-gestures and hand-postures. "-" in the table means that the subjects were doing global activity only. The recall and precision of gestures recognized by SVM were quite low. This is because the feature values have information on the orientation and exercise intensity but do not have information on the trajectory. Combined gestures of behavior with hand gesture or hand posture are almost misrecognized as single behavior. Because of the same reason, the results for gestures while *standing* or *sitting* are also low. In addition, *sitting* is not correctly recognized. This is because the subjects leaned back in order not to hit the armrest when performing a gesture, resulting in the large difference to the training data for just *sitting*. Also, gesture motion while *sitting* is slower than that while *standing*, resulting in small value of variance, and average body orientation during the gesture is similar to just *sitting*. Moreover, all of *hold mobile* and *raise hand* while running were misrecognized as *running* since difference of hand orientation is absorbed by that of vibration of running. From these results, it is hard for feature-based recognition to recognize a lot of combined activities.

DTW, in the contrary, had high recall and precision for all activities. It, however, is remarkable that the performance of our proposal is comparable to that of DTW that has learned all the possible combined activities (9 global×7 local=63 activities), while our proposal has learned single activities only (9 global+7 local=16 activities).

Table 2: Single activities

Kind		Activity
Global	Posture	Sitting
		Standing
		Lying
		Kneeling
	Behavior	Walking
		Running
		Bicycling
		Descending stairs
		Ascending stairs
Local	Hand gesture	Chop
		Throw
		Punch
		Draw a clockwise circle
		Draw a anticlockwise circle
	Hand posture	Hold a mobile phone
		Raise a hand

Table 3: Proposed and comparison methods

Method	# of activities trained	Activities trained
Proposal	16	4 postures
		5 behaviors
		5 hand gestures while standing
		2 hand postures while standing
SVM	63	Combinations of 9 global × 7 local activities
DTW	63	Combinations of 9 global × 7 local activities

Table 4: Recall and precision of recognition.

Activity		Proposal		SVM		DTW	
Global	Local	Recall	Precision	Recall	Precision	Recall	Precision
Stand	Chop	1.000	0.917	0.925	0.633	1.000	0.950
	Throw	1.000	1.000	0.839	0.942	0.996	1.000
	Punch	1.000	1.000	0.874	0.989	0.964	0.996
	Clockwise	1.000	1.000	0.397	0.426	1.000	1.000
	Anticlockwise	1.000	0.882	0.938	0.661	1.000	0.920
	Hold mobile	0.996	-	1.000	-	1.000	-
	Raise hand	1.000	-	0.996	-	1.000	-
	Null	0.999	-	0.919	-	0.993	-
Sit	Chop	1.000	1.000	0.736	0.866	0.993	0.908
	Throw	0.955	1.000	0.840	0.296	0.842	0.983
	Punch	1.000	1.000	0.880	0.680	0.920	0.625
	Clockwise	1.000	1.000	0.831	0.503	0.805	0.914
	Anticlockwise	1.000	0.955	0.855	0.473	1.000	0.471
	Hold mobile	1.000	-	1.000	-	1.000	-
	Raise hand	1.000	-	1.000	-	1.000	-
	Null	1.000	-	0.000	-	0.522	-
Walk	Chop	1.000	1.000	0.000	0.000	0.994	1.000
	Throw	1.000	1.000	0.000	0.000	0.946	1.000
	Punch	0.929	0.833	0.000	0.000	0.983	1.000
	Clockwise	0.944	0.944	0.000	0.000	0.993	0.995
	Anticlockwise	1.000	0.944	0.000	0.000	0.990	0.990
	Hold mobile	0.705	-	0.975	-	1.000	-
	Raise hand	0.442	-	0.956	-	0.999	-
	Null	0.988	-	0.956	-	0.537	-
Run	Chop	0.917	0.826	0.000	0.000	0.986	1.000
	Throw	1.000	0.975	0.000	0.000	0.894	0.997
	Punch	0.429	0.429	0.000	0.000	0.966	1.000
	Clockwise	0.500	0.500	0.000	0.000	0.978	0.916
	Anticlockwise	0.000	0.000	0.000	0.000	0.975	0.942
	Hold mobile	0.000	-	0.000	-	0.886	-
	Raise hand	0.000	-	0.000	-	0.734	-
	Null	0.981	-	0.875	-	0.309	-
Bike	Chop	0.900	1.000	0.000	0.000	0.729	0.956
	Throw	0.913	0.917	0.500	0.500	0.872	0.843
	Punch	1.000	0.742	0.000	0.000	0.930	0.711
	Clockwise	0.622	1.000	0.000	0.000	0.780	0.908
	Anticlockwise	0.500	0.450	0.000	0.000	0.937	0.686
	Hold mobile	1.000	-	1.000	-	1.000	-
	Raise hand	0.999	-	1.000	-	1.000	-
	Null	0.991	-	0.997	-	0.878	-
Average		0.843	0.857	0.510	0.279	0.908	0.908

The drawback of our proposal can be seen from the results of *hold mobile* and *raise hand* while *running*. These low recall and precision were caused by the fact that vibration of running is stronger than we assume, therefore the hand is not classified into posture. Though our method set the threshold according to the intensity of leg as stated in the previous section, it is set based on walking from our pilot study. Employing flexible threshold is our future work. From our extra experiment, however, *holding mobile phone* and *raising hand* are correctly recognized if the hand is classified into posture.

Processing time

This section discusses the processing time of the proposed method. Table 5 shows the processing time for the displacement detection, constancy decision, and recognition with SVM or DTW. Computer used for the evaluation is SONY VAIO VGN-US90PS (Inter CoreSolo Processor 1.2 [GHz]). Simulation program is implemented with Visual C++. The evaluation result is time per one processing calculated based on the processing time for 100,000 trials. The reason the processing time for SVM and DTW differs in the proposal and comparisons is that the number of activities learned is different. Most of the recognition with SVM is occupied with feature extraction, therefore the effect of the number of activities to be recognized is small. The recognition with SVM for the proposal is smaller than that for the comparison method because the proposal extracts only mean as feature value, while the comparison method extracts mean and variance. The number of template for DTW is one per one activity, and the recognition with DTW takes processing time in response to the number of templates. For the results, the processing time for the displacement detection and constancy decision is shorter than that for recognition and processing interval, therefore the proposed method can be applied to real-time applications.

Table 5: Processing time for comparison methods and proposed method [msec].

	SVM	DTW	Proposal		
			Posture	Behavior	Gesture
Displacement detection	-	-	0.00141		
Constancy decision	-	-	-	0.0452	
Recognition with SVM	0.0531	-	0.00203	0.0514	-
Recognition with DTW	-	137	-	-	34.9
Total	0.0531	137	0.00344	0.0980	34.9

Recognition of complex activities

Local activities are confined to right-handed ones in this work. This section discusses recognition of gestures that consist of multiple local activities over the body. Even when multiple parts of body perform local activities, procedure of the proposed method does not change. However, all gestures and postures for each part have to be collected since gesture data for right hand cannot be used for learning the same gesture for left hand.

Combination of local activities and global activities would be the following six patterns: 1) Local activities are all postures, and global activity is also posture. Ex) *Standing with holding cell phone with the right hand and raising the left hand to hold on to a strap.* 2) Local activities are all postures, and global activity is behavior. Ex) *Walking with holding cell phone with the right hand and raising the left hand to cross at the crosswalk.* 3) Local activities are all gestures and global activity is posture. Ex) *Standing with punch with the right hand and chop with the left hand.* 4) Local activities are all gestures and global activity is behavior. Ex) *Walking with punch with the right hand and chop with the left hand.* 5) Local activities are posture and gesture, and global activity is posture. Ex) *Standing with punch with the right hand and raising the left hand.* 6) Local activities are posture and gesture, and global activity is behavior. Ex) *Walking with punch with the right hand and raising the left hand.*

In the cases of 1) and 2), local activities are all postures, which means no part of the body is performing gesture. The parts of the body making a posture are classified into posture and are recognized with SVM that has learned postures only. The activity of the right hand is recognized as *standing with holding a mobile phone*, left hand is recognized as *standing with raising the hand*. The remaining parts are recognized as *standing* or *walking*, therefore such combined activities can correctly be recognized. In the cases of 3) through 6), gesture is included as a local activity. Vibration of the gesture is propagated to the opposite hand that is not performing gesture. In this paper, since the part of body that performs gesture has been limited to the right hand, the body parts other than the right hand are not classified to gesture. For example of *walking with rotating the right arm* activity, vibration of the right hand is propagated to the left hand and the left hand is classified into gesture. In case that local activities are all gestures like in 3) and 4), it is not problem that the left hand is classified to gesture. Though the recognition may be incorrect due to the propagated vibration, the system has a room to recognize it correctly. However, in the cases of 5) and 6), recognition results of the hand making a posture when the leg performing gait are forced to be any of gestures. To contend with the problem, recognition result whose confidence is low is not output by setting threshold to the DTW calculation. However, depending on the setting of the threshold, correct recognition results may be discarded.

Conclusion

We constructed an activity recognition mechanism for combined activities that classifies each part of body into posture, behavior, and gesture, then recognizes individual activities and integrates them. Evaluation results confirmed that our proposed method achieved 0.84 of recall and 0.85 of precision, which is comparable to the method that had learned all the combined activities: 0.90 of recall and 0.90 of precision. As a future work, we plan to separate and integrate activities in more primitive level, that is to say to delete gesture component from raw data of combined activity and to extract pure global activity.

References

[1] Bao, L. and Intille, S.S.: *Activity recognition from user-annotated acceleration data*, in Pervasive 2004, pp. 1–17 (2004).

[2] Pärkkäa, J., Ermes, M., Korpipää, P., Mäntyjärvi, J., Peltola, J., and Korhonen, I.: *Activity classification using realistic data from wearable sensors*, in Trans. on Information Technology in Biomedicine, Vol. 10, pp. 119–128, (2006).

[3] Murao, K. and Terada, T.: *A Motion Recognition Method by Constancy-Decision*, in Proc. of ISWC 2010, pp. 69–72 (2010).

[4] Ravi, N., Dandekar, N., Mysore, P., and Littman, M.L.: *Activity Recognition from Accelerometer Data*, in Proc. of AAAI 2005, pp. 1541–1546 (2005).

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

[6] Junker, H., Amft, O., Lukowicz, P., and Tröster G.: *Gesture spotting with body-worn inertial sensors to detect user activities*, in Journal of Pattern Recognition Society, pp. 2010–2024 (2008).

[7] Keogh, E., Chu, S., Hart, D., and Pazzani, M.: *An online algorithm for segmenting time series*, in Proc. of ICDM 2001, pp. 289–296 (2001).

[8] Liu, J., Wang, Z., Zhong, L., Wiekramasuriya, J., and Vasudevan, V. *uwave: Accelerometer-based personalized gesture recognition and its applications*, in Proc. of PerCom 2009, pp. 1–9 (2009).

[9] Myers, C.S. and Rabiner, L.R. *A comparative study of several dynamic time-warping algorithms for connected word recognition*, in The Bell System Technical Journal, vol. 60(7), pp. 1389–1409 (1981).

[10] Fujinami, K., Jin, C., and Kouchi, S.: *Tracking On-body Location of a Mobile Phone*, in Proc. of ISWC 2010, Late Breaking Results, pp. 190–197 (2010).

[11] Junker, H., Lukowicz, P., and Tröster, G.: *Sampling Frequency, Signal Resolution and the Accuracy of Wearable Context Recognition Systems*, in Proc. of ISWC 2004, pp. 176–177 (2004).

[12] Intille S.S., Bao, L., Tapia, E.M. and Rondoni, J.: *Acquiring in Situ Training Data for Context-aware Ubiquitous Computing Application*, in Proc. of CHI 2004, pp. 1–9 (2004).

[13] Vapnik, V.: *The Nature of Statistical Learning Theory*, Springer (1995).