A Motion Recognition Method by Constancy-Decision

Kazuya Murao, Tsutomu Terada
Graduate School of Engineering, Kobe University, Japan
murao@ubi.eedept.kobe-u.ac.jp, tsutomu@eedept.kobe-u.ac.jp

Abstract

Many context-aware systems using accelerometers have been proposed. Contexts that have been recognized are categorized into postures (e.g. sitting), behaviors (e.g. walking), and gestures (e.g. a punch). Postures and behaviors are states lasting for a certain length of time. Gestures, however, are sporadic or once-off actions. It has been a challenging task to find gestures buried in other contexts. In this paper, we propose a method that classifies contexts into postures, behaviors, and gestures by using the autocorrelation of the acceleration values and recognizes contexts with an appropriate method. We evaluated the recall and precision of recognition for seven kinds of gestures while five kinds of behaviors; The conventional method gave values of 0.75 and 0.59 whereas our method gave 0.93 and 0.93. Our system enables a user to input by gesturing even while he or she is performing a behavior.

1 Introduction

The downsizing of computers has led to wearable computing devices, and context-aware systems using various sensors. Context-aware systems have many applications, such as to healthcare [6]. Most of the contexts that can be dealt with in these applications are *postures* (e.g. *sitting*) and *behaviors* (e.g. *walking*), which are *states* of human activities lasting for certain length of time. They are generally recognized with a classifier such as SVM (support vector machine)[7] operating on extracted feature values such as the mean, variance, and FFT (fast Fourier transform) power spectrum that express body orientation and exercise intensity. Other important activities in daily life include *gestures* (e.g. *a punch*). Gestures are not states but once-off actions, and they can be recognized with a template matching algorithm such as DTW (dynamic time warping) [5] after trimming the waveform of the gesture. The feature values do not have information on how the user has moved, and they are not good for discriminating similar gestures like rotating one's arm clockwise or anticlockwise. Conventional systems force users to explicitly indicate the starting point and endpoint of a gesture, say, by them standing still before and after the gesture or by them pushing a button while performing the gesture.

It is a challenging task to automatically find gestures buried in other behavioral contexts not explicitly given starting points and endpoints. Continuing recognition regardless of the starting points and endpoints may miss out on gestures buried in other contexts or may misrecognize them at a temporal boundaries of a gesture and some other context. These problems have made it difficult for a single system to recognize *postures*, *behaviors*, and *gestures* together.

We developed a system that exploits the fact that behavioral contexts such as *walking* consist of certain data iterations. The system judges the constancy of a context by computing the autocorrelation of the data it senses. When the autocorrelation plot of a periodical wave has clear peaks, we define that the data has constancy; and they indicate certain behaviors. On the other hand, when a gesture occurs during such a behavior, the periodical wave breaks down and the autocorrelation plot no longer has peaks. When there is a peak, our system uses SVM; when there is none, it uses DTW. Postures are recognized when acceleration wave is stationary. By employing autocorrelation, the user's gestures can be automatically spotted even during behavioral activities like *walking*.

2 Related Work

Ten studies on activity recognition are listed in [1], but most of them focus on recognizing ambulation and posture. One study presents a technique using accelerometers for recognizing gestures for the purpose of video annotation. The technique recognizes *cut*, *elbow*, and *punch* [2]. However, these gestures are not while the user is moving. Another study attempts to spot gestures with sensors worn on the body [4]. The data are partitioned into motion segments by using the sliding-window and bottom-up algorithm. The study targeted ten different gestures and although all of them were once-off actions such as *phone up* and *push button*, they had to be captured while the user re-

978-1-4244-9044-8/10/$26.00

mained stationary. It is unclear how that study's method would work when the gestures occur in behavioral contexts. Our approach in this paper is to classify data sequences into three context types (postures, behaviors, and gestures).

3 System Structure

The procedure of our system consists of two phases, as shown in Figure 1. The first phase detects movements of the user. When a movement is detected, the second phase classifies the movement into a behavior or a gesture by determining whether the movement is periodic or not.

3.1 First Phase: Displacement Detection

The system checks for displacements in the sensed data. If $|x(t) - \overline{x}(t)|$ where $x(t)$ is a sensed data and $\overline{x}(t)$ is a moving average exceeds a threshold ϵ, our system detects a movement. Otherwise, the system judges that the user is maintaining a posture. The region of $\overline{x}(t) \pm \epsilon$ is called the epsilon tube; it is used to remove small displacements. We set ϵ to 200 mG since the fluctuation of the data while the subjects were stationary was up to 100 mG.

Since the current value $x(t)$ might temporarily go into the epsilon tube even while the user is moving, the posture begins only after $x(t)$ has been within the epsilon tube for more than 0.25 second. This interval was chosen from our pilot studies. While the data is within the epsilon tube, the system judges that the user is maintaining a posture. At that time, the mean value of the data in the window is calculated as a feature value and the posture is recognized with SVM, which has learned only postures. Since the variance of postures is almost zero, only the mean is used for recognition. When the data indicates movement, this process goes on to the second phase.

3.2 Second Phase: Constancy Decision

Basically, data on a behavior like *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 ACF (autocorrelation function) finds iterations in the user's movements. The discrete ACF $R_{xx}(\tau)$ at lag τ for a data sequence $x(t)$ is defined as $R_{xx}(\tau) = \sum_{t=0}^{N-1} x(t)x(t-\tau)$, where N is the window size for the ACF calculation (64 samples (3.2 seconds) is long enough to capture at least two iterations). In addition, since ACF is at a maximum at $\tau = 0$, all the values of ACF are normalized by $R'_{xx}(\tau) = R_{xx}(\tau)/R_{xx}(0)$ so that the range is (-1,+1).

The system has to decide whether the movement has constancy or not. As shown in Figure 2, the ACF of behaviors

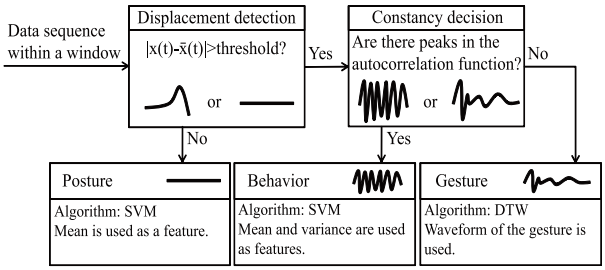


Figure 1. Recognition procedure.

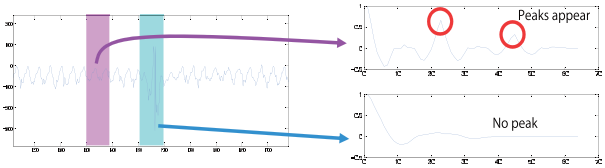


Figure 2. Accelerations of chop gestures while walking (left) and autocorrelation of walking (upper right) and chop (lower right).

shows clear peaks, whereas the ACF of gestures 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 0.8. Reason n/N is used is that the height of the first peak linearly decreases as τ increases. SVM operating on the mean and variance is used to recognize the detected behavior, whereas DTW operating on trimmed original wave is used for recognizing gestures.

Intille et al. focused on acquiring *in situ* data and mentioned that acceleration data collected in non-laboratory environments may display marked fluctuations in the gait cycle [3]. However, although such data fluctuate daily or hourly, data in the range of a second are significantly periodic.

4 Evaluation

We evaluated our system in two different situations; gestures occurring while standing and gestures sporadically occurring during a behavior.

The training and test data were taken from four male subjects aged 22 to 27 years, who wore three accelerometers [8] on their right wrist, hip, and right ankle. The sampling frequency was 20 Hz. They acted out four postures (*sitting*, *standing*, *lying*, and *kneeling*), five behaviors (*walking*, *running*, *bicycling*, *ascending stairs*, and *descending stairs*), and seven gestures (*chop*, *throw*, *punch*, *draw a clockwise circle*, *draw an anticlockwise circle*, *jump*, and *kick*). In total, one hour of posture and behavior data were recorded. Each gesture was recorded about 100 times. *Jump* and *kick* while *bicycling* were not performed for safety's sake. Since we did not assume that two contexts could occur simulta-

neously, the test subjects stopped their behavior while they performed gestures; however, there were no posing intervals to partition the data. The logged data were manually labeled, 10% of which were used for training, the remaining 90% for testing. Training data of gestures were captured while the subjects stood.

Contexts were recognized with three methods; SVM, DTW, and our SVM and DTW hybird. The first two methods are conventional methods, simply applied in steps of several samples. The last one selectively uses SVM and DTW according to the type of context.

4.1 Results

Table 1 lists the recall and precision of recognition while the subjects stood. Figure 3 shows the plots of recognition while the subjects walked.

4.1.1 Results of SVM

The recall and precision of all contexts except gestures were high. This is because the feature values have information on the orientation and exercise intensity but do not have information on the trajectory. The reason why the precision of *clockwise* is listed as N/A is that no *clockwise* result was output.

The drawback of the feature values becomes conspicuous during behaviors. Figure 3(a) plots the recognition results of gestures acted while *walking*. The horizontal axis shows elapsed time, and the vertical axis shows the contexts. Open circles ◯ mean ground truth, and crosses × mean recognition results. × overlapping with ◯ means a correct recognition, and a cross × by itself means a misrecognition. A lot of ◯s and ×s on the *walking* line are ground truths and recognition results of *walking* between gestures. Almost all gestures except for *throw* were incorrectly recognized. The *kick* gestures were buried in the *walking* behavior and were not recognized as gestures. These failures were due to the same reason mentioned above. Gestures should thus be recognized by inspecting time-series data.

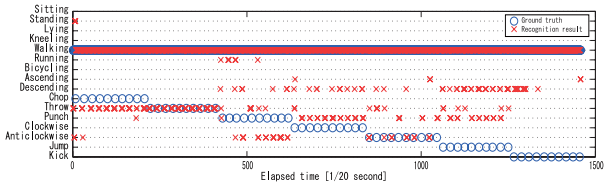
4.1.2 Results of DTW

DTW had high recall and precision for all contexts, as shown in Table 1. This is because DTW used time-series data that can distinguish gestures that were not recognized from the feature values. However, even DTW did not work well while the subjects were *walking*, as shown in Figure 3(b). Correct results as well as incorrect ones were output at the same time. In short, only ×s are shown in Figure 3(b). Figure 4 shows detailed example of subjects *throwing* while *walking*. The recognition results are shown above

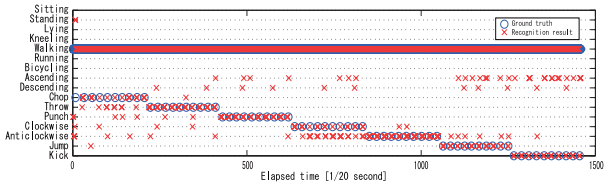
Table 1. Recall and precision of recognition.

Contexts	SVM		DTW		SVM+DTW	
	Re[1]	Pr[2]	Re	Pr	Re	Pr
Sitting	1.00	1.00	1.00	1.00	1.00	1.00
Standing	1.00	0.99	0.97	0.99	0.99	1.00
Lying	1.00	1.00	1.00	1.00	1.00	1.00
Kneeling	1.00	1.00	1.00	1.00	1.00	0.99
Walking	1.00	0.98	1.00	0.99	1.00	1.00
Running	1.00	1.00	1.00	1.00	1.00	1.00
Bicycling	1.00	1.00	1.00	1.00	1.00	1.00
Ascending	0.99	1.00	1.00	0.99	1.00	1.00
Descending	0.98	0.99	0.98	1.00	1.00	1.00
Chop	1.00	0.83	1.00	1.00	1.00	1.00
Throw	0.88	1.00	1.00	1.00	1.00	1.00
Punch	1.00	0.86	1.00	0.75	1.00	1.00
Clockwise	0.00	N/A	0.92	1.00	0.92	1.00
Anticlockwise	1.00	0.51	1.00	0.91	1.00	0.91
Jump	1.00	1.00	0.93	1.00	0.97	1.00
Kick	0.88	1.00	1.00	1.00	1.00	0.97

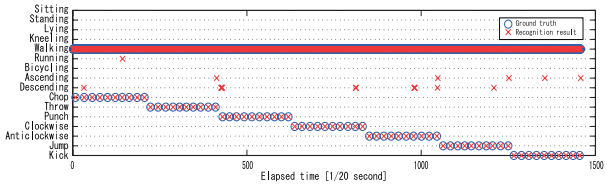
[1]Recall [2]Precision



(a) SVM.



(b) DTW.



(c) SVM+DTW.

Figure 3. Recognition of gestures while walking.

the spikes in the upper part of the figure. The first two results are correct *walking*. However, when the gesture starts, the system misrecognized the data two times (*going downstairs* and *kick*) because the window contains the *walking* and *throw* data. Although the gesture was correctly recog-

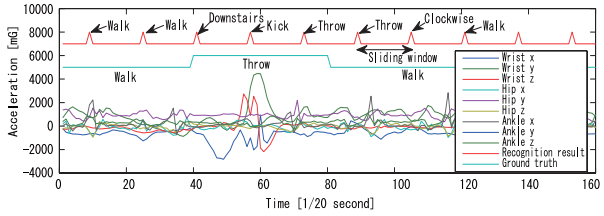


Figure 4. Detailed recognition result of DTW.

nized if a lot of *throw* data were in the window, the subsequent results were incorrect.

4.1.3 Results of the proposal

The results of the proposed method in Table 1 are almost the same as those of DTW since it recognizes gestures with DTW. Figure 3(c) shows that when the subjects performed gestures, the system outputted gestures only once thanks to the constancy-decision. However, false positives nevertheless appeared when the subjects performed behaviors and the ratios of inconstancy to the constancy decision during *walking*, *running*, *bicycling*, *ascending*, and *descending* were 6.8%, 6.9%, 9.0%, 36%, and 15%, respectively. To remove these false positives, we used a filter whereby four consecutive positives would be treated as a real positive. This filter is derived from the fact that the window was slid in steps of 1/4 the length of the window. The waveform is subject to four consecutive decisions, which in turn produce four consecutive inconstancies in motion. This filter removes small interruptions in constancy, and reduced ratios of inconstancy to the constancy decision of *walking*, *running*, *bicycling*, *ascending*, and *descending* to 0.43, 0.32%, 0.69%, 3.3%, and 0.99%, respectively. Especially for *ascending* and *descending*, worse results would stem from flat floors between ten steps, subjects skipping steps, and subjects getting fatigued during an ascent.

To make it clear that our system has an advantage over two comparative methods, Table 2 lists the recall and precision of gesture recognition during behaviors. The recall and precision of SVM are low. Our proposal has higher recall and precision compared with DTW. In addition, the average precisions for the five behaviors were 0.83, 0.85, and 0.90 for SVM, DTW, and our proposal, respectively. These results confirmed that our proposal outperforms the other methods for the three context types.

5 Conclusion

We constructed a new context recognition mechanism that classifies data into postures, behaviors, and gestures by using an autocorrelation function to make a decision about the constancy of behaviors and gestures. Accordingly, DTW is used only when gestures occur, and SVM is used when postures or behaviors occur.

As a future work, we plan to devise a way to recognize simultaneous contexts by changing algorithms for each sensor when their data sequences change.

Table 2. Recalls and precisions of gesture recognition during behaviors.

Behavior	SVM		DTW		SVM+DTW	
	Re[1]	Pr[2]	Re	Pr	Re	Pr
Walking	0.21	0.19	0.84	0.86	0.93	0.94
Running	0.33	0.43	0.74	0.58	0.89	0.92
Bicycling	0.45	0.12	0.63	0.50	0.96	0.97
Ascending	0.43	0.13	0.77	0.49	0.90	0.89
Descending	0.39	0.14	0.77	0.52	0.98	0.94
Average	0.36	0.20	0.75	0.59	0.93	0.93

[1]Recall [2]Precision

Acknowledgments

This research was supported in part by a Grant-in-Aid for Scientific Research (A) (17200006) and Priority Areas (21013034) of the 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] Bao, L. and Intille, S.S. Activity recognition from user-annotated acceleration data. In *Intl. Conference on Pervasive Computing (Pervasive 2004)*, pages 1–17, 2004.

[2] Chambers, G.S., Venkatesh, S., West, G.A.W., and Bui, H.H. Hierarchical recognition of intentional human gestures for sports video annotation. In *Intl. Conference on Pattern Recognition (ICPR 2002)*, pages 1082–1085, 2002.

[3] Intille, S.S., Bao, L., Tapia, E.M., and Rondoni, J. Acquiring in situ training data for context-aware ubiquitous computing application. In *ACM Conference on Human Factors in Computing Systems (CHI 2004)*, pages 1–9, 2004.

[4] Junker, H., Amft, O., Lukowicz, P., and Tröster G. Gesture spotting with body-worn inertial sensors to detect user activities. *Pattern Recognition*, 41:2010–2024, 2008.

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

[6] Ouchi, K., Suzuki, T., and Doi, M. Lifeminder: A wearable healthcare support system using user's context. In *Intl. Workshop on Smart Appliances and Wearable Computing (IWSAWC 2002)*, pages 791–792, 2002.

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

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