

Ascending and Descending Walking Recognition using Smartphone

Vu Ngoc Thanh Sang

Department of Computer and
Information Sciences, Tokyo
University of Agriculture and
Technology, Japan

vsang@livingsys.lab.tuat.ac.jp

Truong Quang Dang Khoa

Department of Computer and
Information Sciences, Tokyo
University of Agriculture and
Technology, Japan

khoa@cc.tuat.ac.jp

Shiro Yano

Department of Computer and
Information Sciences, Tokyo
University of Agriculture and
Technology, Japan

syano@cc.tuat.ac.jp

Toshiyuki Kondo

Department of Computer and
Information Sciences, Tokyo
University of Agriculture and
Technology, Japan

t_kondo@cc.tuat.ac.jp

ABSTRACT

In the recent decades, activity recognition for human behavior monitoring has been one of the major interesting research subjects. This paper investigates fractal dimension (FD) as an effective feature for human activity recognition. The target activities are descending and ascending stairs, which are performed by six male participants. To testify the validity of feature, the classification methods are k-nearest neighbors (kNN) and support vector machine (SVM). The best result achieved in this results is using SVM classifier and obtaining 95.30% for walking downstairs and 88.56% for walking upstairs with window of 4s and equalized data. Another aim of this paper is to examine the impact of database on the classification results. The finding in the paper provides evidence that the classification accuracy is affected insignificantly by the ratio of descending and ascending data.

Categories and Subject Descriptors

I.5.2 [Design Methodology]: Design Methodology – *Feature evaluation and selection*; I.5.4 [Pattern Recognition]: Applications – *Signal Processing*;

General Terms

Algorithms, Measurement, Performance, Design, Experimentation, Human Factors

Keywords

Fractal dimension, descending, ascending, human activity recognition, smartphone

1. INTRODUCTION

Human activity recognition is an important topic in the field of mobile Health (mhealth) [1]. They enable many applications such as daily life monitoring, personal biometric signature, elderly and youth care [2]. Mobile devices are becoming increasingly sophisticated and embedded many diverse and powerful sensors. These sensors include accelerometer, gyroscope, GPS sensors, camera, microphone, etc. It is important to realize that smartphones are an ideal factor for the development of mHealth area because of their embedded sensors, computational power, communication capabilities and ubiquity.

2. LITERATURE REVIEW

The comprehensive survey of the recent advances in human activity recognition with embedded sensors in smartphone was introduced recently [2]. The paper reviewed basic concepts of sensors and their functionality. A smartphone application processed the features extracted from raw sensor readings and predicted user's activities. The common features used in activity recognition were also reported in both time and frequency domain. The activities were categorized in different types such as simple activities, complex activities, living activities and health activities.

Another research attempted to recognize six human activities using multiclass support vector machine (MC-SVM) [3]. The experiment was carried out by a group of 30 subjects wearing the smartphone on their wrist. The recognition process was based on the fixed-width sliding windows of 2.56 second and 50% overlaps. From each window, a vector of 17 features in time and frequency domains was extracted from the accelerometer readings. The classification results using MC - SVM was obtained an average accuracy of 89.3%.

With the same objective, a system was conducted by numerical experiments on identifying the physical activities of a user [4]. The data was collected from 29 users as they performed daily activities. In the next step, informative features were generated based on the 200 raw accelerometer readings (10 seconds). Although there were six basic features, the total features generated from three axes of an accelerometer were forty-three features. The

classification methods in this experiment were J48, logistic regression, multilayer perceptron and straw man. The most average accuracy result was 91.7%, which was derived from multilayer perceptron classification. However, it appeared that the walking downstairs and walking upstairs had the results of 44.3% and 61.5% accuracy because they confused each other. As results, the authors suggested a solution that considering both activities as one set, which had the classification result is 77.6%.

At the moment this paper is written, many commercial smartphone applications have been developed to recognize human activities in daily life. The first example is Google Fit, which is compatible with Android smartphones [5]. The indexes of daily activities such as number of steps, traveled distance, jogging, biking, etc. are recorded automatically and the consumed energy is calculated at the same time. The second example is Sony Lifelog application, which is developed by Sony Corporation [6]. Sony Lifelog is an application that allows users to draw a framework of the activities carried out in daily life. One of many benefits of this application is the recording of the total number of steps a user walked in the day. The consumed energy corresponding to the step counts is estimated simultaneously. However, the energy is estimated based on the number of steps counts, which is inappropriate calculation in walking-like activities. For example, it is required smaller amount of energy for walking downstairs in comparison with walking upstairs. Although the Health App of Apple has similar functions with the two applications in collecting user's exercise data and keeps track of health issues [7], it provides more information about flights of stairs, which is counted as about 3 meters (16 steps) of elevation gain.

In this research, ascending stairs and descending stairs walking are considered because they are similar activities and performed regularly by many people in their daily life. Additionally, fractal dimension is investigated as a new feature, which is useful and effective in human activity recognition. It is benefit not only in classification but also in reducing the curse of dimensionality. Another key point of this research is the trade off between utilizing balance data to secure the mathematical meaning of classification and utilizing unbalance data to maintain the physical meaning of the activities.

3. METHODOLOGY

3.1 Data Acquisition

In the experiment, six male subjects were asked to perform 10 times with each time including walking descending then walking ascending five floors of a building. The smartphone is located portrait in the left pant's pocket of a subject. A Sony Xperia Z3 smartphone has been exploited for the experiment because it contains an accelerometer and a gyroscope for measuring three-dimension linear acceleration and angular velocity respectively at a constant rate of 25 Hz, which is sufficient for capturing human daily activities. In order to achieve a practical data of the activity, each subject was allowed to walking freely. However, this allowance leads to an unbalance in data of walking downstairs and upstairs of each subject and among different subjects. The consequence is that there is a trade off in constructing training set and testing set. Although equalizing the descending and ascending data in each set avoid bias in classification mathematically, it is regardless the physical meaning of going up and down of five floors. The two cases were analyzed in this research to find a suitable data set for online classification.

3.2 Feature Extraction

3.2.1 Fractal Dimension (FD)

FD presents the irregular characteristic of a time series based on the power law index [8]. Although the FD can be considered similar to Euclidian dimension, an important difference between them is FD may be a non-integer. The FD demonstrates how a set in micro-scale is similar to its counterpart in the macro-scale. The closest integer value of a FD shows its equivalence to the Euclidian dimension. For example, if the FD is close to zero, it presents a point (0-dimension in Euclidian). It describes lines when it approaches one (1-dimension in Euclidian) and surfaces when it approaches two (2-dimension in Euclidian), and so on.

3.2.2 Higuchi's Algorithm

Higuchi's algorithm is an effective method to calculate the FD in time domain [9]. It is the most accuracy algorithm to estimate FD [10]. It is based on the measurement of the length of the curve in time series.

From a given time series: $X(1), X(2), \dots, X(N)$, the algorithm constructs k new time series:

$$X_k^m : X(m), X(m+k), X(m+2k), \dots, X\left(m + \text{int}\left[\frac{N-m}{k}\right]k\right) \quad (1)$$

where k is the interval time, m is initial time ($m = 1, 2, \dots, k$), $\text{int}[\cdot]$ is the integer part of a real number. The length $L_m(k)$ of the curve with k -segments is calculated by the following formula:

$$L_m(k) = \frac{1}{k} \left[\sum_{i=1}^{\text{int}\left[\frac{N-m}{k}\right]} |X(m+ik) - X(m+(i-1)k)| \right] \frac{N-1}{\text{int}\left[\frac{N-m}{k}\right]} \quad (2)$$

$L_m(k)$ is not the length of the curve in Euclidean sense, it stands for the normalized sum of absolute values of difference in ordinates of pair of points distant k (with initial point m). The length of curve for the time interval k , $L(k)$ is calculated as the mean of the k values $L_m(k)$ for $m = 1, 2, \dots, k$:

$$L(k) = \frac{\sum_{m=1}^k L_m(k)}{k} \quad (3)$$

If the average length $L(k)$ is estimated approximately as followings:

$$L(k) \sim k^{-D} \quad (4)$$

the curve is said to has fractal dimension D , which is calculated by least-square linear best-fitting procedure.

According to the result of practical experiments, in this experiment k has the value of 10, which means the data in investigated window is divided in 10 segments.

3.3 Signal Processing

In this experiments, horizontal, vertical and forward/backward movement of the user's leg are captured in x -axis, y -axis and z -axis respectively. The activities are analyzed in the window's length of 2 and 4 seconds, which is the expected duration for classification in online implementation. From each window, a vector of 6 features is obtained by calculating FD from three accelerometer variables and three gyroscope variables.

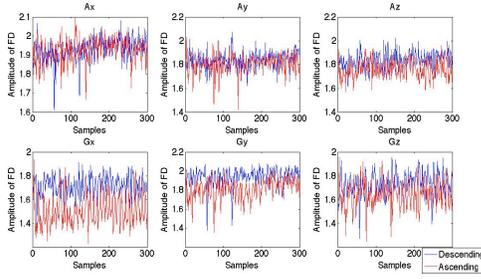


Figure 1. Sample of FD (window = 2s, k = 10)

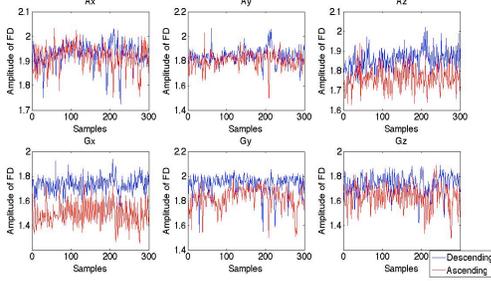


Figure 2. Sample of FD (window = 4s, k = 10)

As the aforementioned information, there is a difference in data length. For going downstairs the shortest time is about 32 seconds (about 800 samples) and longest time is 1 minute 11 seconds (about 1780 samples). On the other hand, for walking upstairs, the minimum and maximum durations are 25.8 seconds (about 645 samples) and 2 minute and 9 seconds (about 3200 samples). Therefore, this paper attempts to investigate the effects of dividing training set and testing set equally and unequally. Each data is equalized based on the smallest duration, which is 25.8 seconds in walking upstairs. In contrast, the unequal data is the raw data, which have different lengths of ascending and descending walking. A 6-fold cross-person activity recognition model is applied to eliminate the effect of user sensitivity. Particularly, in each case, a subject is set as a training set while the others five subjects are treated as a training set. The final result is the average of six above circumstances. The classification is concluded by k-nearest neighbor and support vector machine classifier.

3.4 Classification

3.4.1 k-nearest neighbor (kNN)

kNN is a type of instance-based learning and is the simplest of all machine learning algorithms. In this research, the neighbors were taken from two classes (ascending stairs, descending stairs). The Euclidean distance is typically computed for classification. In the next step, the distance from each sample in the testing set to all the samples in training set is calculated and the results is arranged in ascending order. A sample is classified by a majority vote of its neighbors, then the object is assigned to the class most common among its k nearest neighbors (k is a positive integer). The kNN classification in this paper is concluded by $k = 5$ and $k = 15$ then the value of k that rises a better accuracy will be reported in this paper.

3.4.2 Support Vector Machine (SVM)

In SVM classification, the inputs are transformed to high dimensional feature spaces in which they become more separable in comparison to the original input spaces. At the beginning, a

hyperplane is constructed to optimally separate the training data into two classes. Then, two parallel hyperplanes (or support vectors) are constructed on each side of the hyperplane. By maximizing the distance between the two parallel hyperplanes, the data is classified into two classes.

4. RESULTS AND DISCUSSION

4.1 Fractal Dimension

The FD of the accelerometer (A_x , A_y , A_z) and gyroscope (G_x , G_y , G_z) in different window lengths are shown in the Figure 1 and Figure 2. It is possible to draw several quantitative evaluations from the two figures. Firstly, although descending walking and ascending walking are visually distinguishable in channel A_z , G_x and G_y , the 4-second window shows clear difference than 2-second window. Among all channels, G_x is a channel that manifests the clarity of the difference between the two activities. The reason is when a subject moves his left leg to step forward in the ascending stairs, there is a moment the smartphone is fixed in the subject pant's pocket. When he moves his right leg forward, the smartphone is released in the left pant pocket and vibrate around x -axis. In descending walking, this phenomenon does not happen, which leads to the difference in FD of the two activities.

When FD of a curve approaches 2, it manifests the filling capacity the whole plane or the complexity of the curve. According to the two figures, the information of accelerometer is more complex than information of gyroscope because its FD is near to 2. The figures also reveal that descending walking has higher degree of complexity than ascending walking.

4.2 Classification

4.2.1 Statistical Results

The differences between descending walking and ascending walking are highlighted in Table 1. Although the slight different in activities are shown in the A_x channel, the most noticeable difference could be found in G_x channel. Furthermore, there are significant differences in the channels A_z , G_x and G_y . Therefore, in the following experiments, these channels will be combined to form feature vectors for classification.

Table 1. Statistical results

Window	t-test	A_x	A_y	A_z	G_x	G_y	G_z
2s	t-value	14.37	30.25	46.12	65.56	41.02	28.58
	p-value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
4s	t-value	13.41	29.13	45.15	60.60	60.56	36.55
	p-value	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

Table 2. Classification Accuracy for kNN with k = 15

Window	Activity	A_{xyz}	G_{xyz}	A_zG_{xy}	$A_{xyz}G_{xyz}$
2s	D_eq	76.10	88.62	91.05	88.67
	D_uneq	72.41	86.12	87.98	85.89
	A_eq	74.16	81.08	80.11	82.76
	A_uneq	76.55	83.35	83.81	83.66
4s	D_eq	84.40	92.29	94.50	93.79
	D_uneq	82.69	90.32	93.24	91.46
	A_eq	81.03	87.32	89.54	87.94
	A_uneq	82.37	88.36	91.02	89.24

Table 3. Classification Accuracy for SVM

Window	Activity	A_{xyz}	G_{xyz}	A_zG_{xy}	$A_{xyz}G_{xyz}$
2s	D_eq	77.82	89.59	89.86	89.95
	D_uneq	76.41	89.55	89.45	89.76
	A_eq	74.16	81.08	81.97	82.54
	A_uneq	74.98	82.79	83.24	83.80
4s	D_eq	85.02	93.17	95.30	95.21
	D_uneq	84.70	93.10	94.74	94.03
	A_eq	81.38	87.23	88.56	88.56
	A_uneq	82.72	89.09	90.39	90.77

4.2.2 *k*-nearest neighbors (*k*NNs)

The Table 2 shows the accuracy of *k*NN classification for different feature vectors. For example, the column *Axyz* presents the classification results when the inputs are the three axes of accelerometer. Similarly, the data in column *AzGxy* is the accuracy of classification when inputs are the vectors of accelerometer *x*-axis, gyroscope *x*-axis and gyroscope *y*-axis. It is apparent from the Table 2 that there is no significant difference between equal (*X_eq*) and unequal (*X_uneq*) data in descending (D) and ascending (A) walking. However, data from this table reveals that long window (4 sec) provides more accuracy in classification in comparison with short window (2 sec). It can be seen from the table that the accuracy using accelerometer data is lower than the results using gyroscope data. Among several methods of channel selection, the *AzGxy* option raises the highest accuracy with the same window lengths.

4.2.3 Support Vector Machine (SVM)

According to the Table 3, the data affirms that the results of SVM classifier have similar findings with *k*NN classifier. To begin with, although the equal data classification results in higher accuracy than unequal data, the difference between them is insufficient to draw conclusion which one is better the other. Secondly, 4-sec window yields a better recognition results than 2-sec window. Additionally, gyroscope data has been proved that it provides a better result than accelerometer in both *k*NN and SVM classification. Finally, selecting the *AzGxy* option bringing the best classification is also true with SVM classifier. The most accuracy of SVM classification is 95.30% for walking downstairs and 88.56% for walking upstairs with window of 4 seconds and equalized data.

In comparison with other works, this research has several advantages. Firstly, only one feature is used to recognize human similar activity such as descending and ascending stairs. This advantage decreases the cost of dimensional curse in comparison with other researches. For example, the research in [4] needs 43 features for activity classification but the result is quite small for descending stairs (44.3 %) and ascending stairs (61.5 %), which leads to the consideration of combining the two activities as one action. In another research, the classification precisions of walking downstairs and walking upstairs are 87.2% and 72.6% respectively [3]. However, it requires a vector of 17 features to draw that classification result. It is important to notice that the feature in this research is calculated directly from the raw data in time domain, which reduces the calculating complexity. Another good point of this research is an activity could be recognized in a short interval of time such as 2 or 4 seconds. The research [4] applies an approximate window size of 2.56 sec, while the study [3] divides data in 10 seconds window segments. A short window time allows the flexible in recognition because sometimes the subjects may perform an activity less than the required duration for recognition. This paper also recommends that the data of the two activities should be equalized in the preprocessing stages. Although there is no significant difference between equalized and unequalized data, this implementation facilitates the signal processing as well as mitigates the complexity of calculation.

5. CONCLUSION AND FUTURE WORK

In this research, FD is introduced as a valuable feature for human activity recognition. To evaluate its validity, it is used to classify similar activities such as descending and ascending stairs with *k*NN and SVM classifier. According to the results, it is

recommended that ones should utilize equal raw data for classification when the difference ratio is insignificant. This utilization leads to the easy in programming application and processing data. This study focuses on classify activities of the subjects with the smartphone in their left pant's pocket. The final goal of this research is develop a complete human activity recognition application on smartphone. More broadly, research is also needed to determine different locations of the smartphone and more target activities in the future. Equally important is the implementing of completed smartphone application that could classify human activities online.

6. ACKNOWLEDGMENTS

The research was partially supported by MEXT Fund for Smart Space Technology toward Sustainable Society, and JSPS KAKENHI, (Grant number 26120005).

7. REFERENCES

- [1] World Health Organization, *mHealth: New Horizontal for Health Through Mobile Technologies*, Global Observatory for eHealth Series, 3rd ed, 2011.
- [2] Xing Su, Hanghang Tong and Ping Ji, *Activity Recognition with Smartphone Sensors*, Tsinghua Science and Technology, vol.13, 6/2014.
- [3] Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra, Jorge L.Reyes-Ortiz, *Human Activity Recognition on Smartphones using a Multiclass Hardware-Friendly Support Vector Machine*, Ambient assisted living and home care. Springer Berlin Heidelberg, 2012, pp.216- 223.
- [4] Kwapisz, Jennifer R., Gary M. Weiss, and Samuel A. Moore, *Activity recognition using cell phone accelerometers*, ACM SigKDD Explorations Newsletter 12, no. 2, 2011, pp.74-82.
- [5] Google Fit [Online]. Available: <https://developers.google.com/fit/>
- [6] Sony Lifelog [Online]. Available: <http://www.sonymobile.com/gb/apps-services/lifelog/>
- [7] Health App [Online]. Available: <http://www.apple.com/ios/whats-new/health/>
- [8] T. Higuchi, *Relationship between the fractal dimension and the power law index for a time series: a numerical investigation*, Physica D: Nonlinear Phenomena 46(2), 1990, pp.254-264.
- [9] T. Higuchi, *Approach to an irregular time series on the basis of the fractal theory*, Physica D: Nonlinear Phenomena 31(2), 1988, pp.277- 283.
- [10] Rosana Esteller, George Vachtsevanos, Javier Echauz, Brian Litt, *A comparison of waveform fractal dimension algorithms*, Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on 48(2), 2001, pp.177-183.
- [11] B. E. Boser, I. M. Guyon, and V. N. Vapnik, *A training algorithm for optimal margin classifiers*, Proceedings of the fifth annual workshop on Computational learning theory, Pittsburgh, Pennsylvania, USA, 1992
- [12] D. Olson and D. Delen, *Support Vector Machines*, Advanced Data Mining Techniques, Springer Berlin Heidelberg, 2008, pp. 111-123.