proceedings

MDPI

Proceedings

Data Labeling for Participatory Sensing Using Geature Recognition with Smartwatches †

Luis A. González-Jasso [1,2] and Jesus Favela [2,*]

[1] INIFAP, 20660 Aguasclaientes, Mexico; gonzalez.luis@inifap.gob.mx
[2] Computer Science Department, CICESE, 22860 Ensenada, Mexico
* Correspondence: favela@cicese.mx; Tel.: +52-646-175-0500
† Presented at the 12th International Conference on Ubiquitous Computing and Ambient Intelligence (UCAmI 2018), Punta Cana, Dominican Republic, 4–7 December 2018.

Published: 22 October 2018

Abstract: Supervised activity recognition algorithms require labeled data to train classification models. Labeling an activity can be performed trough observation, in controlled conditions, or thru self-labeling. The two first approaches are intrusive, which makes the task tedious for the person performing the activity, as well as for the one tagging the activity. This paper proposes a technique for activity labeling using subtle gestures that are simple to execute, and that can be sensed and recognized using smartwatches. The signals obtained by the inertial sensor in a smartwatch are used to train classification algorithms in order to identify the gesture. We obtained data from 15 participants who executed 6 proposed gestures in 3 different positions. 208 characteristics were computed from the accelerometer and gyroscope signals and were used to train two classification algorithms to detect the six proposed gestures. The results obtained achieve a precision of 81% for the 6 subtle gestures, and 91% when using only the first 3 gestures.

Keywords: gesture recognition; data labeling; smartwatch; activity recognition

1. Introduction

Activity, and behavior recognition has become one of the most active areas of research in ubiquitous computing. The field makes use of data gathered using mobile, wearable and/or environmental sensors to create models capable of inferring the activity being performed by an individual [1]. These models are often obtained by training supervised classification algorithms. Thus, the data needs to be labeled in order to have "ground-truth" for training and testing the model or classifier [2].

Labeling mobile sensing data is a task that requires considerable effort and is error-prone, particularly if the data is captured in naturalistic conditions. One approach is to capture data in controlled conditions, such as in a laboratory environment. In such a setting labeling can be done by the research team with the help of specialized equipment. For instance, the individual can be made to walk at a certain speed on a treadmill; asked to prepare tea in an instrumented kitchen; or be monitored while sleeping using polysomnography with specialized equipment.

Requiring subjects to attend a lab to capture and label data can be expensive. Furthermore, ecological validity can be compromised by the fact that the subjects are being observed or the instrumentation can interfere with the manner in which they perform the activity. For instance, sleeping in a lab with sensors attached to the body vs. sleeping in your own bed and subject to normal environmental conditions (noise, light, the presence of others, etc.).

Another common approach is to ask the subject under study to perform her own labeling. This could be done with the help of mobile systems programed to capture labeling data. One such example would be a smartphone app that requires the subject to press a button just before going to

sleep and right after waking up. This approach has some drawbacks, including the fact that the subject could forget to indicate that the activity is being performed, or he can do so at a later time. The labeling could also interfere with the activity, for instance in the case of a subject who stops cycling to take his phone out and indicate that he is riding a bicycle.

This article proposes an approach to data labeling based on self-report from the individual conducting the activity using the recognition of simple gestures. The approach follows the experience sample method [3], in which the subject under study wears a smartwatch, and receives a notification, either in the form of vibration or sound, requesting him for a label about his current activity or mood. The subject responds with a subtle gesture that demands little cognitive and physical effort and it is also discreet and thus can be used in various social settings without others being aware. For instance, the system could at random times query the subject about his current social setting (i.e., alone; with work peers; with family and/or friends; with strangers), or when the device infers from accelerometer data that the subject is moving it can ask him to confirm this and indicate the mode of transportation (i.e., not moving; walking; running; in a car/bus). The rest of the paper describes the subtle gestures proposed for labeling, the methodology used to classify the gesture, the classification results, and a discussion of the results and application of the proposed approach.

2. Materials and Methods

We describe the criteria proposed to define the gestures to be recognized, those that were selected and the approach proposed for their recognition as well as to assess its accuracy.

2.1. Defining the Body Gestures to Label Activities

Wearable computers are increasingly being used to infer activities. Some of these devices, notably smartwatches, already detect gestures to activate diverse functions, for instance turning the screen on when the wearer raises and turns his arm in the direction of his face, signaling that he wants to look at the watch.

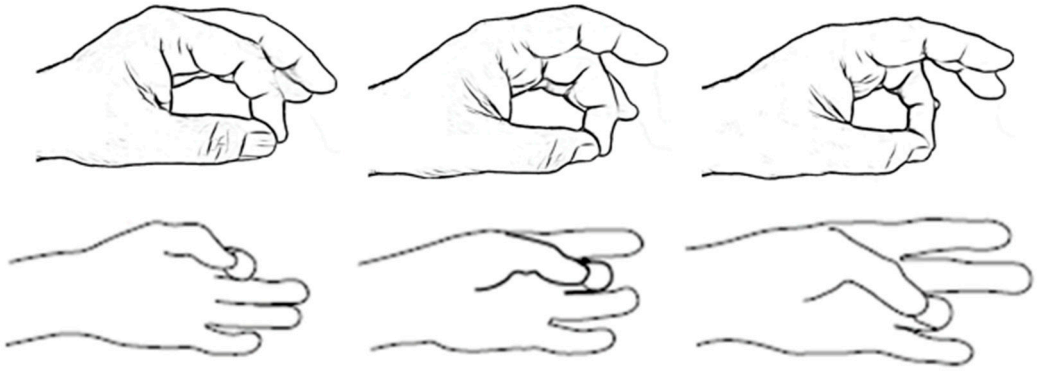
Most gesture recognition using smartphones use the accelerometer and/or gyroscope in the device. For instance, Kalatrian et al. report on their effort to predict activities such as food intake or opening a medication jar using a smartwatch [4], while Costante, G. et al., proposed an approach to detect personalized gestures to assist visually impaired users [5].

Our first task was to define the gestures to be recognized for labeling. The gestures had to fulfill the following criteria:

- Between 4 and 8 gestures that could be adequately discriminated. This number could allow labeling applications to answer binary queries (Yes/No), ternary queries (Yes/No/I don't know) and 5 or 7 likert-scale queries.
- Users should require little effort to perform the gestures.
- The gestures should be easy lo learn and differentiate by those performing them.
- The proposed gestures should be different from those currently used in smartwatches to operate the device.
- The user should be able to perform the gesture in different locations, while performing a variety of activates, while in different body postures (laying down, sitting, running, etc.) and in different social circumstances (in a meeting, at home, walking in the street, etc.). Thus, the gestures should be discrete and subtle; the user should be able to make the gesture without others noticing it.

After considering several alternatives we settled for the six gestures shown in Figure 1. These gestures require a small movement of the hand to tap once, or twice, the thumb with the index, middle or ring finger. We originally considered an additional two gestures that involved the little finger, but our initial efforts showed that it was difficult to discriminate these gestures with those using the ring finger. These gestures fulfill the criteria defined above, but their subtlety makes it challenging to recognize them. Dementyev and Paradiso report on their efforts to recognize the

gesture made by taping the thumb with the index finger [6]. We next describe our proposed approach to recognize these gestures.



(**a**) Tap with index (**b**) Tap with middle (**c**) Tap with ring

Figure 1. The six gestures selected. They require users to tap their thumb, once or twice, with either the index, middle or ring finger.

2.2. Data Sources, Signal Processing and Feature Extraction

Two sensors in the smartwatch will be used to characterize the movement in the wrist associated to each gesture: accelerometer and gyroscope. The accelerometer measures the acceleration at which the sensor moves in three axes, while the gyroscope measures orientation and angular velocity in the same axes. One advantage of using a smartwatch is that it is normally worn in the wrist of the left hand of the individual and the X-axis coincides with the direction of the arm with positive values towards the fingers, while the Z-axis is perpendicular to the screen of the device.

The acceleration signal is first processed to eliminate the acceleration due to gravity, which depending on the position of the hand could affect all axes. A low-pass filter is used to reduce the effect of this component.

The next step is to segment the signal that includes the gesture. The recording of the gestures initiates right after the query is given (through vibration of the device or sound), but the subject might wait a short amount of time before it performs the gesture. From the analysis of signals of the gestures we estimated the maximum amount of time required to perform the more complex gesture (double-tap with the ring finger) and add 1.5 s to consider the time it takes the user to react after perceiving the query. This time window guarantees that the gesture will be contained in the signal, except in cases were there is considerable delay by the user in performing the gesture. The Dynamic Time Wrapping (DTW) algorithm is used [7] to make an initial assessment of the presence of a gesture by comparing the signal with that of a sample gesture. We found the signal from the gyroscope to be more stable and thus, the comparison is made with the signal registering angular velocity. The comparison is made with total angular velocity, which adds the contribution from the 3 axes, to account for differences in orientation when performing the gesture [8].

Once the signal is classified as having a gesture using DWT we proceed to determine where the gesture starts and ends. The signal (magnitude of angular velocity) is divided in windows of approximately 240 ms each. The average angular velocity of each window is calculated and a new, compressed signal is produced with each average value per segment. A low and a high threshold were empirically determined to estimate the presence of a gesture in the signal. When the magnitude exceeds the high threshold we establish the possible presence of the gesture and a low threshold is used to establish where the signal starts and ends. Figure 2 illustrates this process with the signal of a double-tap with the index. A value above the high threshold is initially detected in data point 15, indicating the presence of a gesture. The window for the gesture is established between data point 9 and 45, the first and last values below the low threshold.

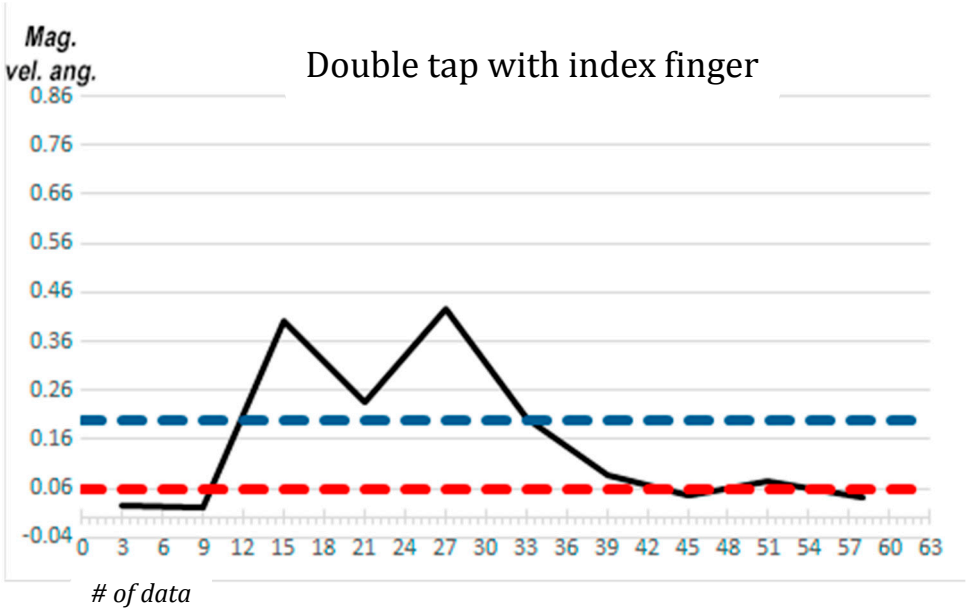


Figure 2. Determining.

If the length of the potential gesture segmented is less than 200 ms the signal is considered too short to include a gesture and the user is queried again to repeat the gesture. This minimum length of time was estimated empirically. To eliminate potential peaks in the signal produced by sporadic movements we confirm that the start and end of the signal is no less than 220 ms away from a threshold above 0.4 in angular velocity. If this is not the case, the length of the signal is adjusted as shown in Figure 3. Similarly, if the signal lasts more than 1300 ms it will be rejected.

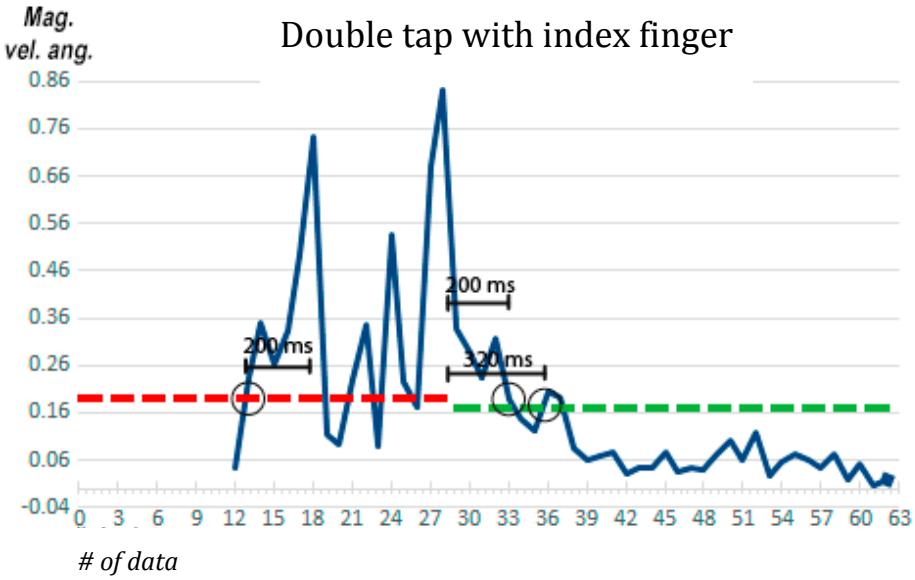


Figure 3. Determining the final length of the signal. A peak is detected and eliminated at the end of the signal to avoid movements not associated with the gesture. The final signal, of 800 ms, goes from data 13 to 33.

Once the signal with a potential gesture is detected we proceed to extract relevant features to be used for gesture classification. The signals from each axis obtained from each sensor are combined to produce 14 signals (7 per sensor). This includes the original signals X, Y and Z, combinations of two of them XY ($\sqrt{X^2+Y^2}$), XZ ($\sqrt{X^2+Z^2}$), and YZ ($\sqrt{Y^2+Z^2}$), and the total magnitude XYZ ($\sqrt{X^2+Y^2+Z^2}$). Seven features are extracted from each of these signals: average, maximum,

minimum, standard deviation, and the 3 quartiles, for a total of 98 (14 × 7) features per signal. An additional 12 features are calculated from the area under the curve of the signals corresponding to the three components of each sensor. Finally, 98 additional characteristics are obtained by applying a Fourier transform to the signals and estimating the same 7 measures mentioned above. This results in a total of 208 features.

2.3. Gesture Classification

As classifiers we propose the use of two supervised machine learning algorithms: Support Vector Machine (SVM) and Sequential Minimal Optimization (SMO). Initial experiments were also performed using a backpropagation neural network, but this technique was later eliminated due to long training times. For SVM we use the two most common kernels: *linear* and a *radial basis function* and tried several parameters of cost and gamma (γ) as suggested in [9]. The best results were obtained with a cost of C = 4 and γ = 0.0078125. For SMO we used Polykernel with the best results obtained for a C = 11.

The 208 features obtained from each signal are normalized in the range of [–1, 1] to avoid for features of greater magnitude to dominate the classification.

2.4. Evaluation of the Approach for Gesture Recognition

An experiment was designed to evaluate the approach proposed and compare the performance of the two classifiers selected (SVM and SMO) using WEKA 3.9.1.

The study design is within-subjects, in which participants are asked to perform the 6 gestures using a smartwatch that records the signals from the accelerometer and gyroscope. Two android smartwatches were used by each participant: an LG G100 and an ASUS ZenWatch 2.

The individuals performed the gestures while in three different postures: standing with the arms facing down; standing with the arm bent and the watch facing the user; and siting down with the arm resting on a pile of books on top of a table (to provide support).

A total of 15 participants were recruited to gather data to train the classifiers. Inclusion criteria for participants included: age between 10 and 60; no previous experience using a smartwatch; and being right-handed. As criteria for exclusion we considered having known motor problems that could cause excessive movement in their arms/hands.

Each participant is asked to perform each gesture twice in each posture, for a total of 36 gestures per individual. Sound was used to indicate to the user which gesture to perform next. A researcher used a smartphone connected via Bluetooth to the smartwatch to initiate and control the intervention.

3. Results

To evaluate the approach estimating the precision of the approach using the two classifiers: SVM and SMO.

3.1. Gesture Detection and Signal Segmentation

Of the 90 sample recorded for each gesture the approach described in Section 2.2 to detect the presence of gesture eliminated approximately 4.5% of them, with a similar number of samples reject per gesture (SD = 0.013). Of those signals that were not excluded (N = 515) the DWT estimated that in 94.7% of them a gesture was present.

Figure 4 shows an example of the presence of a gesture being detected using DWT. Figure 4a shows the signal used as reference, which corresponds to the magnitude of angular velocity of an individual performing a tap with the index finger. Figure 4b shows the signal of another gesture that is accurately recognized as including a gesture using DWT. Finally, Figure 4c shows the signal produced by the circular movement of the wrist, which the algorithm correctly identifies as not having one of the gestures of interest.

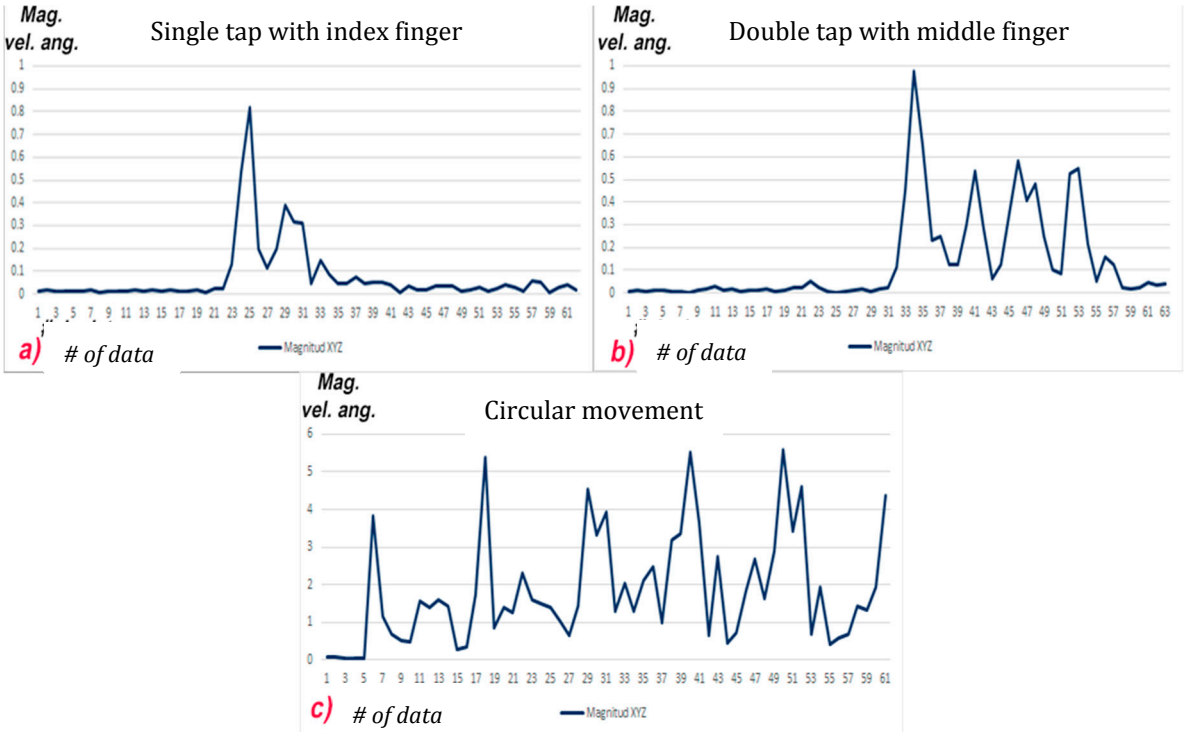


Figure 4. Recognizing the presence of a signal using DWT. (**a**) Signal of the magnitude of angular velocity used as a reference (single-tap with index finger). (**b**) Signal in which the presence of a gesture is correctly inferred (double-tap with ring finger). (**c**) Signal from a hand performing another movement (circular movement with the wrist).

3.2. Gesture Recognition

To estimate the accuracy of the approach a 10-fold cross validation was performed. The average precision using SVM was of 81.55, with tap with the index finger having the highest precision of 89.4%. The lowest precision was obtained with the double tap with the index finger (77.1%). Table 1 shows the confusion matrix for these results. Of the 85 gestures of a single tap with the index finger, 9 were incorrectly classified (false negatives) and 6 were identified as this gesture when in fact they were not (false positive). False negatives are less important than false positives, since the system could ask the user to repeat a gesture it doesn't recognize.

Table 1. Confusion matrix for 6 gestures using SVM and 10-fold validation.

a	b	c	d	e	f	<-- Classified as
76	0	2	5	2	0	a= index
1	64	2	1	5	6	b= middle
2	2	66	3	7	2	c= ring
3	0	1	64	12	3	d= double index
0	2	0	2	65	12	e= double middle
0	1	2	2	10	63	f= double ring

The SMO classifier had similar results, with an average precision of 81.15%. Similarly, the best accuracy was obtained with a single tap with the index finger (89.41%) and the lowest with the double tap with the index finger (77.1%).

We performed an additional evaluation using just the 3 gestures that require only a single tap. The average accuracy improved to 91.46% using SVM.

WristFlex is an array of force sensitive resistors worn around the wrist, which was developed to infer gestures similar to the ones proposed here [6]. A test was conducted to recognize 6 gestures, including the 3 we propose, plus taping also the little finger, a relaxed hand, and an open hand with fingers spread, with 10 individuals. While the approaches could not be directly compared given the slight differences in gestures used the accuracy obtained was similar. Yet, our approach does not require the use of specialized hardware, but rather uses standard commercial smartwatches.

5. Conclusions

We described an approach to data labeling for activity recognition thru gesture recognition with a smartwatch. Data labeling remains an open problem in activity and behavior recognition, which often use supervised classifiers. We proposed 6 subtle gestures that take little time and effort to perform and that could be enacted in a wide variety of postures and social settings. With 6 different labels participatory data labeling can be done for a variety of mobile sensing applications which might require the user to confirm her current activity/behavior, or respond to binary, tertiary or likert-type queries.

The precision obtained for 6 gestures (81%) seems is low for the application, it would generate a significant number of false positives. However, considering only 3 gestures the precision improved to 91% which seems practical for several circumstances, particularly considering that participatory sensing is often done with high frequency and a few data points incorrectly labeled might have little impact. In addition, the application could decide to query the user a second time if it estimates that the answer provided is inconsistent with the readings from the sensor.

The approach proposed could be improved if the model could be tailored trough additional learning with a few sample gestures from each individual.

A limitation of this work is that the gestures were performed in control conditions; we expect accuracy to decrease when the gestures are performed under naturalistic conditions.

References

1. ReproBulling, A.; Blanke, U.; Schiele, B. A tutorial on human activity recognition using body-worn inertial sensors. *ACM Comput. Surv.* **2014**, *46*, 33.
2. Spanias, J.A.; Perreault, E.J.; Hargrove, L.J. A strategy for labeling data for the neural adaptation of a powered lower limb prosthesis. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* **2014**, *2014*, 3090–3093.
3. Larson, R.; Csikszentmihalyi, M. The experience sampling method. *New Directions Methodol. Soc. Behav. Sci*. **1983**, *15*, 41–56.
4. Kalantarian, H.; Alshurafa, N.; Sarrafzadeh, M. Detection of Gestures Associated with Medication Adherence using Smartwatch-based Inertial Sensors. *IEEE Sens. J.*, **2015**, *16*, 1054–1062.
5. Costante, G.; Porzi, L.; Lanz, O.; Valigi, P.; Ricci, E. Personalizing a smartwatch-based gesture interface with transfer learning. In Proceedings of the 2014 22nd European Signal Processing Conference (EUSIPCO), Lisbon, Portugal, 1–5 September 2014; pp. 2530–2534.
6. Dementyev, A.; Paradiso, J.A. WristFlex: Low-Power Gesture Input with Wrist-Worn Pressure Sensors. *Annu. ACM Symp. User Interface Softw. Technol.* **2014**, *14*, 161–166.
7. Senin, P. Dynamic Time Warping Algorithm Review. *Science* **2008**, *80*, 1–23.
8. Wen, H.; Ramos Rojas, J.; Dey, A.K. Serendipity: Finger Gesture Recognition using an Off-the-Shelf Smartwatch. In *Proc. 2016 CHI Conf. Hum. Factors Comput. Syst*; ACM: New York, NY, USA, 2016; pp. 3847–3851.
9. Hsu, C.W.; Chang, C.C.; Lin, C.J. A Practical Guide to Support Vector Classification. *BJU Int.* **2008**, *101*, 1396–400.