Activity Recognition System For Mobile Phones Using The MotionBand Device

Ákos Fábián
Budapest University of Technology and Economics
Budapest, Hungary

Norbert Győrbíró
Nokia Siemens Networks
Budapest, Hungary

Gergely Hományi*
Nokia Siemens Networks
Budapest, Hungary

ABSTRACT

In this paper we present a novel system that recognizes and records the motional activities of a person using a mobile phone. Wireless sensors measuring the intensity of motions are attached to body parts of the person. Sensory data is collected by a mobile application that recognizes the prelearnt activities in real-time. For efficient motion pattern recognition of gestures and postures, feed-forward backpropagation neural networks are adopted. The design and implementation of the system are presented along with the records of our experiences. The recognized activity is used as an additional retrieval key in our extensive mobile memory recording and sharing project.

Categories and Subject Descriptors

I.2.6 [**Artificial Intelligence**]: Learning—*connectionism and neural nets*; H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*Information filtering*; H.2.8 [**Database Management**]: Database Applications—*Data mining*

General Terms

Life logging

Keywords

Motion recognition, motion sensors, mobile computing, neural networks

1. INTRODUCTION

Many systems have been developed to collect data from a person's everyday life for later retrieval (see e.g. [5, 15]). Their main purpose is to enlarge human memory utilizing the capabilities of computers. They can record e-mail and chat conversations, documents, location information, video and audio content using cameras and microphones and many other types of personal and environmental data.

In our research project, Sharme [14] we have been focusing on data that can be acquired on mobile phones. In recent years mobile phones have become our very personal devices. Besides storing our contacts, text messages and calendar appointments, the newer generation of mobile devices called *smartphones* can be used for taking photos, recording audio clips, browsing the Internet and accessing e-mail. Therefore enormous amount of personal data can be collected even during a short period of time in which the recall of a desired event is very difficult. Storage of additional metadata can help navigation in the recorded content. The activity of a person can be a useful cue for retrieving memories. When it is used for annotating content data, we can build smart search queries - for example "*Who called me while I was jogging in the park?*"

In this paper we present a mobile system that recognizes the motional activities of a person. This information is recorded along with the personal memory archive. We have focused attention on low resource usage of the solution since it needs to be running on mobile devices offering real-time motion recognition capabilities.

2. RELATED WORK

The idea of enhancing memory retrieval with additional cues is not new. Lamming and Flynn [7] utilized physical context information such as location, phone calls and the interaction between different PDAs as retrieval keys. Kern *et al.* [6] annotated meeting recordings with motional activities - for instance standing and sitting to differentiate between presentation and discussion sessions. In order to recognize postures, they utilized body-worn accelerometer sensors.

The usage of accelerometers for physical activity recognition is widely established in the literature. Randell and Muller [12] used a single biaxial accelerometer and calculated the RMS and integrated values over the last two seconds for both axes. This input data was used for classifying six activities (walking, running, sitting, walking upstairs, downstairs and standing) utilizing a neural network. Mäntyjärvi *et al.* [11] also applied neural networks for human motion recognition. Here the feature vector was created with PCA (Principal Component Analysis) and ICA (Independent Component Analysis) from two triaxial sensors attached to the left and right sides of the hip. Lee and Mase [9] developed an activity and location recognition system using a combination of a biaxial accelerometer, a compass and a gyroscope. Their classification technique was based on a

*Corresponding author: gergely.homanyi@nsn.com

Mobilware'08 February 12-15, 2008, Innsbruck, Austria.

fuzzy-logic reasoning method.

The above studies relied on wired sensors which could be uncomfortable to wear. Therefore it may be difficult to perform outdoor or long term experiences. Lately wireless accelerometers became available, enabling measurements in more comfortable settings. Bao and Intille conducted an extensive study [1] with twenty subjects using five wireless biaxial sensors. Sensory information was processed with FFT to extract mean, energy, frequency-domain entropy and correlation features. Recognition of twenty different everyday activities on these features was performed using decision table, instance-based learning, C4.5 decision tree and naive Bayes classifiers found in the Weka Machine Learning Algorithms Toolkit. Decision tree classifier delivered the best results on the gained feature vector. Utilizing the same toolkit but only a single triaxial accelerometer worn at the pelvic region, Ravi *et al.* [13] studied the performance of the base-level classifier algorithms as well as the meta-classifiers including the voting, stacking and cascading frameworks. Plurality voting performed the best in those settings.

One of the important areas of motion recognition applications is healthcare. Chen *et al.* [3] have implemented a mobile phone-based system for multiple vital signs monitoring. The system could detect if the monitored patient falls using a wireless accelerometer and can alert the supporters. In [10] accelerometers were applied to detect symptoms of Parkinson's disease.

Feature extraction techniques such as FFT, PCA or ICA are rather computation intensive and better suited for desktop computers. Mobile phones are more limited in terms of processing power. They also lack the constant supply of power and have to rely on their scarce battery resource. Thus we are looking for a method that could provide the best possible classification rate considering the limitations of mobile devices. To support continuous recording of mobile phone events a real-time recognition system is sought after.

3. SYSTEM DESIGN

Our activity recognition system is built of the following components: wireless body sensors, a smartphone and a desktop computer. The in-house MotionBand [8] devices have been used for measuring sensory data. As Figure 1 shows they resemble a watch and can be strapped on conveniently to the limbs. A MotionBand contains an accelerometer, a magnetometer and a gyroscope for measurements. There is also a button on the top which can be used to trigger arbitrary assigned events. It can be connected to other devices via wireless Bluetooth connections.

The smartphone is responsible for collecting data from the sensors. This is a natural approach since most of the time people carry their mobile phones with themselves. A phone kept in a pocket can continuously record the motional activity of the person.

However, the supervised learning of activities could be a heavy burden for even a more powerful smartphone. Therefore we have separated the learning and the recognition processes. After collecting enough sensory data, the system is trained on a desktop computer using feed-forward neural networks. They are powerful at pattern recognition [2] and after training the classification is done quickly. This enables us to implement a neural network on the phone that is fast enough to classify the activities in real-time. The parameters of this neural network are set to be equal to the trained neural network on the PC.

Figure 1: The MotionBand device

The measurement method and the detailed description of the system are presented in the following sections.

4. MEASUREMENT METHOD

The MotionBand device developed in Nokia Research Center has been utilized for measurements. The device contains three triaxial sensors: an accelerometer, a magnetometer and a gyroscope. For every axis the accelerometer provides 16-bit data accurate to $\pm 6G$. The measured sensor value is affected by the gravity of the Earth. The magnetometer measures the deflection from the magnetic north and can be used while the device is in a still state. The gyroscope measures the angular velocity, i.e. the speed of rotation, in the range of $\pm 300 deg/s$. Sensory data is arranged into 28-byte packets containing additional bytes for synchronization and checksum purposes. Packets also include the status of the button which can be found on the top of the device. The device can be connected via wireless Bluetooth connections, if connected approximately 50 packets are transmitted in every second. Due to the limitations of Bluetooth technology, at most seven MotionBands can be connected simultaneously to a single mobile phone or a computer.

MotionBands are designed to be comfortable to wear and thin enough to be used under clothes. MotionBand is attached to the user's body part with flexible straps. The device has an internal battery with an approximate active operating time of 5 hours and 20 hours in idle mode. The battery can be recharged by a standard Nokia phone charger. The weight of the device is only 30 grams including the battery.

Using the three sensors it is possible to track both the orientation and movement of a body part. However, for activity recognition purposes the accelerometer is the most valuable sensor giving information about the forces describing the movements.

The following measurement method is based on the idea that complex activities of the human body can be described by their smaller components, i.e. the intensity of body part motions. Our goal was to capture this motional activity by deriving an appropriate intensity measure from the raw accelerometer data. We were looking for a measure that can be calculated and processed real-time on a smartphone device. The intensity of motions is directly proportional to the variation of acceleration. Building on this observation we defined the intensity value as shown in Equation 1.

$$Intensity \quad := \quad \frac{1}{N}\sum_{i=1}^{N}\left|\frac{s_i - s_{i-1}}{\Delta x}\right| \tag{1}$$

where the measured acceleration values are denoted as s_i $(i = 1 \ldots N)$ and N is the number of samples. Since the input data is discrete, a simple numerical derivative is calculated to reflect volatility. The absolute values of the derivatives are summed and normalized by the number of samples. Features were extracted using a window size of two seconds giving $N = 100$ samples. Since the 50Hz sampling frequency of the MotionBand is approximately constant, the Δx sampling interval can also be treated as constant and can simplify the calculations. In a window a single intensity value is calculated for each axis.

In the Sharme project interesting events of the mobile phone user are recorded, e.g. when a new photo is taken or a phone call is made. Every captured event is augmented with metadata information describing the person's context. The current activity of the person is one such piece of contextual information. Since we need to provide the metadata instantly, the calculation of intensity values are done continuously, i.e. the window is sliding by one sample at a time. Note that this calculation can be executed quickly in case of Equation 1. When the metadata request arrives the current intensity values are classified to produce the activity state.

In the following the recognition method is presented.

5. ACTIVITY RECOGNITION

Three MotionBand devices were attached to the test subjects to collect body-part intensity values. We wanted to differentiate between several activities in both standing and sitting positions, therefore attaching the devices to the dominant wrist, hip and the dominant ankle seemed a promising choice. These sites are also suggested by research study [1].

The six activities to be recognized were resting, typing, gesticulating, walking, running and cycling. Resting was defined as an activity with very low intensity values such as sitting on a chair or lying down on a couch. For a more realistic situation, brief movements such as stretching or changing the posture were allowed. We assigned the task of writing an e-mail on a computer to typing thus it involved typing on the keyboard and pointing with the mouse. Gesticulating involved intense hand gestures while standing still or walking slowly. The above tasks were measured in an office environment. For a natural setting, walking and running exercises were performed outdoors with various speeds and turning around corners due to street obstacles. The cycling activity was recorded in a gym.

MotionBands were connected to a smartphone to capture data from the accelerometers. The benefit of this configuration is that it enables collecting data unobtrusively and very mobile, i.e. the subject could freely engage in either outside or inside activities. During recording, subjects used the button of the MotionBand on the wrist to mark the boundaries between the different activities. Using these markers and the noted sequence of activities, our software could partition and annotate the recording samples. Approximately one second of samples before and after the markers was cut out to eliminate transitional movements between activities.

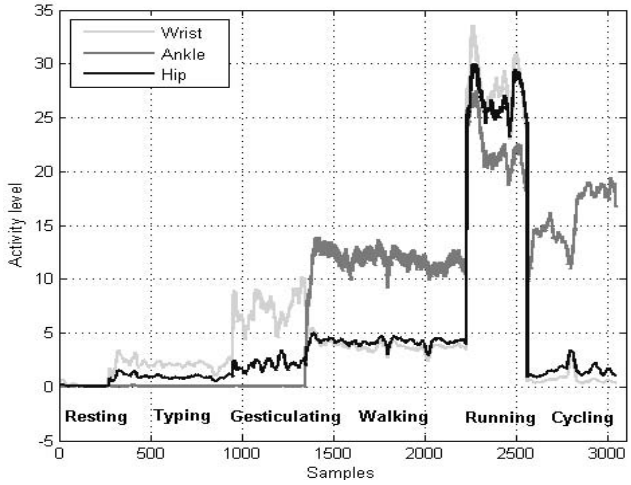

Figure 2: Sample intensity values

Figure 2 shows a sample stream of intensity values. The intensities are considerably different for the examined activities. Our recognition method is based on this observation.

Feed-forward neural networks were chosen as the tool for supervised learning of the activities. The learning phase is done offline on a desktop computer inside the Matlab environment. A design choice had to be made between a single large neural network and multiple small networks. The large network could be trained to classify all activities while the small ones could recognize one particular activity each. During classification each small network takes the current intensity values as input and calculates its output. The network producing the highest confidence value is considered as the winner.

Neural networks performing only one classification task are considered to be better than one with multiple tasks [4]. Thus we have chosen the small networks over the large one.

In theory, the training phase is computationally intensive but the recognition is fast. In practice, each perceptron classifier corresponds to an evaluation of an exponential function. However, most mobile phones do not have a floating-point unit (FPU), i.e. they can only simulate real-valued arithmetic, increasing the processing time. Therefore a network with a small number of perceptrons is employed.

The network has 9 fully-connected neurons in the input layer, 3 neurons in the hidden layer and one neuron in the output layer. The tangent sigmoid function was used as the transfer function. Six such networks were utilized, each responsible for recognizing one particular activity.

5.1 Classification performance

Teaching of activities was done using the Neural Network Toolbox in Matlab. The input set consisted of three complete sets of activities recorded by the authors of this paper. The average number of samples per activity was 1776 and all three sets had roughly the same size.

Networks were trained in multiple rounds using data from one subject at a time.

For each network the training set was composed of all the corresponding positive inputs and an equal number of negative inputs chosen randomly from the other five activities. This helped to balance the number of positive and negative training samples. The networks were trained with ten-fold cross-validation using the Levenberg-Marquardt backpropagation algorithm. This method is one of the most accurate algorithms, but due to its $O(n^2)$ memory and speed requirements it can only be used with small networks. The common mean squared error (MSE) was used as the performance function.

To evaluate recognition performance, testing was done using all data from the subjects. Here utilizing the training person's data is reasonable since only a fraction of that was presented to the network beforehand.

Classification efficiency is generally measured with accuracy, i.e. the proportion of the total number of predictions that are correct. This accuracy may not be an adequate performance measure when the number of negative cases is much greater than the number of positive cases. In our case each network was responsible for classifying only one activity, i.e. only one sixth of the samples were considered as positive. Now if all the negative cases were classified correctly, but none of the positive cases, the accuracy would still be 83.33%. Here, *F-measure* as defined in Equation 2 may be a more adequate evaluation method.

$$F = \frac{(\beta^2 + 1) \cdot P \cdot TP}{\beta^2 \cdot P + TP}, \qquad (2)$$

where P (precision) is the proportion of the predicted positive cases that were correct, TP (true positive or recall rate) is the proportion of positive cases that were correctly identified and β has a value from 0 to infinity and is used to control the weight assigned to TP and P. We set $\beta = 1$ and got the harmonic mean of P and TP, the so called F_1-measure.

After the rounds of training and testing, F-measures for each activity were aggregated using arithmetic mean. These values can be seen in the first column of Table 1.

Table 1: Classification performance.

	F-measure (%)	
Activity	9-dimensions	3-dimensions
Resting	61.92	54.29
Typing	72.14	80.28
Gesticulating	81.46	78.63
Walking	92.35	98.09
Running	94.71	99.74
Cycling	75.96	78.74
Average	79.76	81.63

The average recognition rate was 79.76%. Since some alterations in activities were allowed it hurt the performance. During resting, gentle posture changes versus resting peacefully caused the relatively low efficiency. The posture changes involving arm movements caused misclassification between resting and typing. Good measurements were achieved where the activity was nearly periodic in the window: walking, running and cycling. However, the cycling activity was carried out differently by each one of us: one kept a slow rhythm, while the other cycled in a fast pace, the third of us alternated between slow and fast periods. This noticeable variation of intensity reduced the classification performance.

During measurements the proper placement of MotionBands was important. A small misalignment may have caused changes in the distribution of intensity along the axes. In order to reduce the sensitivity of the system, as an alternative approach, the intensity vector was replaced with a scalar intensity value, i.e. the sum of the triaxial intensities. This simplification made the recognition independent of directions. The corresponding three-dimensional input set was directly derived from the nine-dimensional measurements. The same training and testing was performed and the results can be seen in the second column of Table 1. Classification rates for typing, walking, running and cycling improved, gesticulating fell slightly. Posture changes in resting reduced performance even more considerably than in the nine-dimensional case.

5.2 Recognition on mobile phone

The activity classification system was built as part of our research project, Sharme. The goal of the project is to capture personal memories by means of extensive recording. We have been focusing on data that can be acquired on mobile phones including phone calls, text messages, photos, etc. An important aspect of the research is to enhance the recollection of the recorded memories. Thus for each recorded event we store additional metadata, e.g. time, location or physical activity of the person as presented in this paper.

A prototype recording application was implemented for the Nokia 6630 mobile phone using the Symbian C++ programming environment. The activity classification was built as a module to this program. The phone has a multi-tasking operating system. This allows us to continuously record in the background during which the phone can be used for making calls or accessing other functionalities without disturbance. However, the battery of the mobile phone is a scarce resource. Mobile software draining the battery quickly could not be successful. Furthermore, the ARM mobile processor of the phone is only able to simulate floating-point operations, slowing down mathematical calculations. In order to optimize the consumption of the expensive processor and battery resources, the complexity of the system had to be reduced as much as possible. This motivated us to employ small neural networks and intensity values based on the numerical derivatives.

The training of the networks was performed in Matlab on a powerful desktop computer. The parameters of the networks, i.e. the weight matrices and the bias vectors represent the classification knowledge. Networks with the same topology were implemented for the mobile recognition system inheriting the above learnt parameter values.

Recognition on the mobile phone works as follows. The input intensity values are fed into all six activity classifiers and their outputs are compared. Since tangent sigmoid is used as the transfer function of the networks, the output values range from -1 to 1. The output is accepted if the recognition is at least 95% confident, i.e. the output is greater or equal than 0.9. If multiple outputs are above this threshold, i.e. the activity is rather ambiguous, the network with the highest output is considered to recognize the activity correctly. In other scenarios one might allow multiple activities at the same time, e.g. gesticulating while walking. This case we opted for a single best guess.

6. CONCLUSIONS AND FUTURE WORK

In this paper the architecture and implementation of a real-time mobile activity recognition system was presented. Considering the resource constraints of mobile phones, we have managed to create a system that is able to classify physical activities with acceptable confidence. In our experience the recognition was rather insensitive to the training person's characteristics thus can be used by other people without the need of re-training.

As a continuation of this work other algorithms or input processing methods could be explored. Algorithms presently suitable only for desktop computers could be soon utilized on mobile devices. This trend is driven by the increasing processing capabilities of smartphones, e.g. the new Nokia N93 phone comes with hardware FPU enabling more complex calculations.

We put emphasis on the practical applicability of the activity classification system. By using lightweight and comfortable sensors and the person's mobile phone, activities can be monitored and recorded throughout a longer period of time. Our Sharme research project aims to record all kinds of information that can be gathered on the phone for later memory retrieval. Archiving the current physical activity seems to be a useful search criterion for browsing recorded events. Sharme is designed to be easily extensible with additional sensing and measurement devices. As sensors become easily wearable, e.g. weaved into clothing, the smartphone can be thought of as the aggregator, processor and transmitter of sensory data.

7. REFERENCES

[1] L. Bao and S. S. Intille. Activity recognition from user-annotated acceleration data. *Lecture Notes on Computer Science*, 3001:1–17, 2004.

[2] C. M. Bishop. Pattern recognition and feed-forward networks. In R. A. Wilson and F. C. Keil, editors, *The MIT Encyclopedia of the Cognitive Sciences*, pages 629–632. MIT Press, 1999.

[3] W. Chen, D. Wei, S. Ding, M. Cohen, H. Wang, S. Tokinoya, and N. Takeda. A scalable mobile phone-based system for multiple vital signs monitoring and healthcare. *International Journal of Pervasive Computing and Communications*, 1(2):157–163, 2005.

[4] P. Cortez. MLP application guidelines, July 2004. Summer School NN2004 Neural networks in classification, regression and data mining, Oporto, Portugal.

[5] J. Gemmell, G. Bell, and R. Lueder. MyLifeBits: a personal database for everything. *Communications of the ACM*, 49(1):88–95, Jan 2006.

[6] N. Kern, B. Schiele1, H. Junker, P. Lukowicz, and G. Tröster. Wearable sensing to annotate meeting recordings. *Personal and Ubiquitous Computing*, 7(5):263–274, 2003.

[7] M. Lamming and M. Flynn. Forget-me-not: intimate computing in support of human memory. In *Proceedings of the Friends 21*, Feb 1994.

[8] K. Laurila, T. Pylvänäinen, S. Silanto, and A. Virolainen. Wireless motion bands, September 2005. Position paper at UbiComp'05 Workshop on "Ubiquitous computing to support monitoring, measuring and motivating exercise", Tokyo, Japan.

[9] S.-W. Lee and K. Mase. Activity and location recognition using wearable sensors. *IEEE Pervasive Computing*, 1(3):24–32, 2002.

[10] J. Mäntyjärvi, P. Alahuhta, and A. Saarinen. Wearable sensing and disease monitoring in home environment, 2004. Workshop on Ambient Intelligence Technologies for WellBeing at Home. Held in conjunction with 2nd European Symposium on Ambient Intelligence. EUSAI. Eindhoven, 8 Nov 2004.

[11] J. Mäntyjärvi, J. Himberg, and T. Seppanen. Recognizing human motion with multiple acceleration sensors. In *Proceedings of the 2001 IEEE International Conference on Systems, Man, and Cybernetics*, volume 2, pages 747–752, Oct 2001.

[12] C. Randell and H. Muller. Context awareness by analysing accelerometer data. In *Proceedings of the Fourth International Symposium on Wearable Computers*, pages 175–176, Oct 2000.

[13] N. Ravi, N. Dandekar, P. Mysore, and M. L. Littman. Activity recognition from accelerometer data. In *Proceedings of the Seventeenth Conference on Innovative Applications of Artificial Intelligence(IAAI)*, pages 1541–1546. AAAI Press, Jul 2005.

[14] Sharme project, Nokia Research Center. Homepage is at http://research.nokia.com/research/projects/sharme/.

[15] S. Vemuri and W. Bender. Next-generation personal memory aids. *BT Technology Journal*, 22(4):125–138, Oct 2004.