Experimental Evaluation of Variations in Primary Features Used for Accelerometric Context Recognition

Ernst A. Heinz[1], Kai S. Kunze[1], Stefan Sulistyo,[1]
Holger Junker[2], Paul Lukowicz[2], and Gerhard Tröster[2]

[1] Entertainment and Pervasive Computing Group, International University (IU) in Germany
{ernst.heinz, kaisteven.kunze, stefan.sulistyo}@i-u.de
http://www.i-u.de/schools/heinz/

[2] Wearable Computing Laboratory, Institute for Electronics (IfE), ETH Zürich, Switzerland
{junker, lukowicz, troester}@ife.ee.ethz.ch http://www.wearable.ethz.ch/

Abstract. The paper describes initial results in an ongoing project aimed at providing and analyzing standardized representative data sets for typical context recognition tasks. Such data sets can be used to develop user-independent feature sets and recognition algorithms. In addition, we aim to establish standard benchmark data sets that can be used for quantitative comparisons of different recognition methodologies. Benchmark data sets are commonly used in speech and image recognition, but so far none are available for general context recognition tasks. We outline the experimental considerations and procedures used to record the data in a controlled manner, observing strict experimental standards. We then discuss preliminary results obtained with common features on a well-understood scenario with 8 test subjects. The discussion shows that even for a small sample like this variations between subjects are substantial, thus underscoring the need for large representative data sets.

1 Introduction

Using simple sensors distributed over the user's body has recently emerged as a promising approach to context recognition in mobile and wearable systems. In particular, motion sensors such as accelerometers or gyroscopes have been shown to provide valuable information about user activity. Systems based on a 3 axes have been successfully trained to distinguish between everyday activities like walking, standing, and sitting [6,7,4]. It has also been shown that appropriate combinations of sensors attached to different limbs can provide information necessary to recognize specific complex activities like getting up and greeting an arriving person with a handshake [3] or operating a particular home appliance [5]. An important question that needs to be resolved before such systems can move towards real-life applications is their generalization capability. So far, most successful experiments have been performed in special settings such as a laboratory with a single or just a few randomly selected subjects. At this stage it is not clear how well those results generalize to different locations and other subjects. Thus, a system trained to differentiate between walking up or down a particular staircase with a young female subject might fail for another staircase and / or an older male subject. To a degree,

E. Aarts et al. (Eds.): EUSAI 2003, LNCS 2875, pp. 252–263, 2003.

Experimental Evaluation of Variations in Primary Features Used for Accelerometric Context Recognition

Ernst A. Heinz[1], Kai S. Kunze[1], Stefan Sulistyo,[1]
Holger Junker[2], Paul Lukowicz[2], and Gerhard Tröster[2]

[1] Entertainment and Pervasive Computing Group, International University (IU) in Germany
{ernst.heinz, kaisteven.kunze, stefan.sulistyo}@i-u.de
http://www.i-u.de/schools/heinz/

[2] Wearable Computing Laboratory, Institute for Electronics (IfE), ETH Zürich, Switzerland
{junker, lukowicz, troester}@ife.ee.ethz.ch http://www.wearable.ethz.ch/

Abstract. The paper describes initial results in an ongoing project aimed at providing and analyzing standardized representative data sets for typical context recognition tasks. Such data sets can be used to develop user-independent feature sets and recognition algorithms. In addition, we aim to establish standard benchmark data sets that can be used for quantitative comparisons of different recognition methodologies. Benchmark data sets are commonly used in speech and image recognition, but so far none are available for general context recognition tasks. We outline the experimental considerations and procedures used to record the data in a controlled manner, observing strict experimental standards. We then discuss preliminary results obtained with common features on a well-understood scenario with 8 test subjects. The discussion shows that even for a small sample like this variations between subjects are substantial, thus underscoring the need for large representative data sets.

1 Introduction

Using simple sensors distributed over the user's body has recently emerged as a promising approach to context recognition in mobile and wearable systems. In particular, motion sensors such as accelerometers or gyroscopes have been shown to provide valuable information about user activity. Systems based on a 3 axes have been successfully trained to distinguish between everyday activities like walking, standing, and sitting [6,7,4]. It has also been shown that appropriate combinations of sensors attached to different limbs can provide information necessary to recognize specific complex activities like getting up and greeting an arriving person with a handshake [3] or operating a particular home appliance [5]. An important question that needs to be resolved before such systems can move towards real-life applications is their generalization capability. So far, most successful experiments have been performed in special settings such as a laboratory with a single or just a few randomly selected subjects. At this stage it is not clear how well those results generalize to different locations and other subjects. Thus, a system trained to differentiate between walking up or down a particular staircase with a young female subject might fail for another staircase and / or an older male subject. To a degree,

E. Aarts et al. (Eds.): EUSAI 2003, LNCS 2875, pp. 252–263, 2003.

generalization issues can be addressed through additional user- and situation-specific training. However there are limits to how much training effort a user is willing to put up with. Hence, it is desirable to be able to design and train systems to work under a wide variety of circumstances. To this end, at least three issues need to be addressed.

1. For given recognition tasks, it must be determined how sensor signals change for different relevant situations and users.
2. Sets of features as insensitive as possible to the variations above must be found.
3. Standard data sets with the relevant factors varied in a controlled manner over some desired value range must be generated to allow for situation-insensitive training.

This paper describes some early results in an ongoing project addressing the above issues. The project aims to collect a representative data set for different typical context recognition tasks. The data is to be used to derive person-independent features and recognition algorithms. In addition, we intend to establish a publicly available benchmark database of standard context recognition tasks which can be used to objectively quantify and compare different approaches. Such benchmark databases are widely and successfully used in automatic speech and image recognition. In the remainder of this paper, we first outline the experimental considerations and procedures used to record the data in a controlled manner while observing strict experimental standards. We then discuss preliminary results obtained on simple scenario with 8 test subjects. While no statistically valid conclusions can be drawn from such a small sample, interesting effects can still be observed. In particular we show that variations between subjects do indeed have a strong impact on feature selection and recognition performance. This illustrates the importance of establishing standard data sets as intended by our project.

2 Experimental Procedure

The setup of the experiments was carefully planned and documented based upon trial runs performed in advance. We also explained the whole procedure in considerable detail to the test subjects beforehand, such that they could easily follow the prescribed route and tasks. The overall objective of the experiment was to gather acceleration data of selected body parts in different action contexts for multiple test subjects with varying physical characteristics. Of course, the whole experimental setup should only have a minimal effect on the normal physical movements of the test subjects if at all. The video recordings of the test subjects in action show that we fully achieved this aim.

2.1 Hardware Setup

The sensor system employed in the experiments draws its power from a standard USB connection and interfaces to any RS-232 serial port for control and data signal transmission. We used a Xybernaut Mobile Assistant IV (MA-IV) wearable computer with a 233 MHz Pentium-MMX CPU, 64 MB of RAM, and a 6 GB hard drive running SuSE Linux 6.2 to drive the sensor system. Because of the USB and serial ports needed, the full Xybernaut MA-IV system consisting of both the main unit and the port replicator had to be strapped to the test subjects. Fortunately, there was enough space left here to also

(a) on Right Leg (b) on Left Leg (c) on Right Hip

Fig. 1. Sensor Placement

affix the central components of the sensor system. Due to the professional belt packaging provided by Xybernaut for its wearable computers, this worked quite smoothly and did not hinder the test subjects in any perceivable way during the course of the experiment. Finally, three acceleration sensors were affixed to the test subjects as depicted by the three photographs shown in Figure 1:

(a) on *right leg* about 1 cm above knee cap, tightly fixed, exactly moving like the leg;
(b) on *left leg* about 1 cm above knee cap, lightly fixed, moving like the trousers;
(c) on *right hip*, tightly fixed, exactly moving like the hip.

The sensors on the legs were oriented with their z-axes in the direction of the movement and their y-axes pointing up. The sensors on the hip had their x-axes facing up and their z-axes to the right (i.e., their final orientation differed slightly from the picture above).

Fig. 2. A Single Flight of Stairs

2.2 Action Sequence

After being fully wired and connected, the test subjects were placed at the start of the experimental "trail" that led them through parts of Leibniz Hall, the building housing the School of IT at the International University (IU) in Germany. A small custom program (written in ANSI-C and running directly on the wearable computer) recorded the sensor readings throughout the whole trail by essentially dumping the raw data from the serial input port of the Xybernaut MA-IV. The list below characterizes in detail the stages and actions that the test subjects moved through.

1. Start signal: 3x tapping on right knee.
2. Walk straight for about 9.09 m.
3. Open a door with right hand, walk through, and close it (door swings inwardly with its handle on the left and the hinges on the right side).
4. Turn right and walk straight for about 6.59 m to a staircase as depicted in Figure 2.
5. Climb down 6 flights of 11 stairs each with rightward turns after the first 5 flights (stair width: 30 cm, stair height: 15 cm, stair angle: 27° from the horizontal).
6. Walk straight for about 51.34 m to opposite staircase.
7. Climb up 6 flights of 11 stairs each with left-ward turns after the first 5 flights (same stair specifications as for downward climb).
8. Walk straight at fast pace for about 44.13 m back to the door of the starting room.
9. Stop signal: 3x tapping on right knee again.

2.3 Data Recording and Processing

In order to gain a better understanding of the test subjects themselves and their physical characteristics in particular, we also requested brief individual profiles from them specifying the following details:

– age, gender, height, weight, and right- or left-handedness;
– frequency of sports activities (daily, weekly, monthly, yearly, or never on average);
– additional information such as injuries, handicaps, nationality, and other specials.

The whole procedure was supervised by at least one experiment conductor, who was also responsible for filming the whole sequence with a video camera. Post-experiment activities included verifying that the recorded files actually contained valid data, visually checking the video film for inconsistencies regarding the prescribed sequences of actions, and – last but not least – processing the recorded raw data and transforming it into meaningful visual representations. This was done using IU SENSE [1,2] and MATLAB (release 12). IU SENSE, developed by IU students and written in Java, is an extensible real-time application with graphical data displays in various views and user-configurable layouts (see Figure 3). It also provides numerous possibilities to perform off-line data visualization from files and conversion into several different file formats.

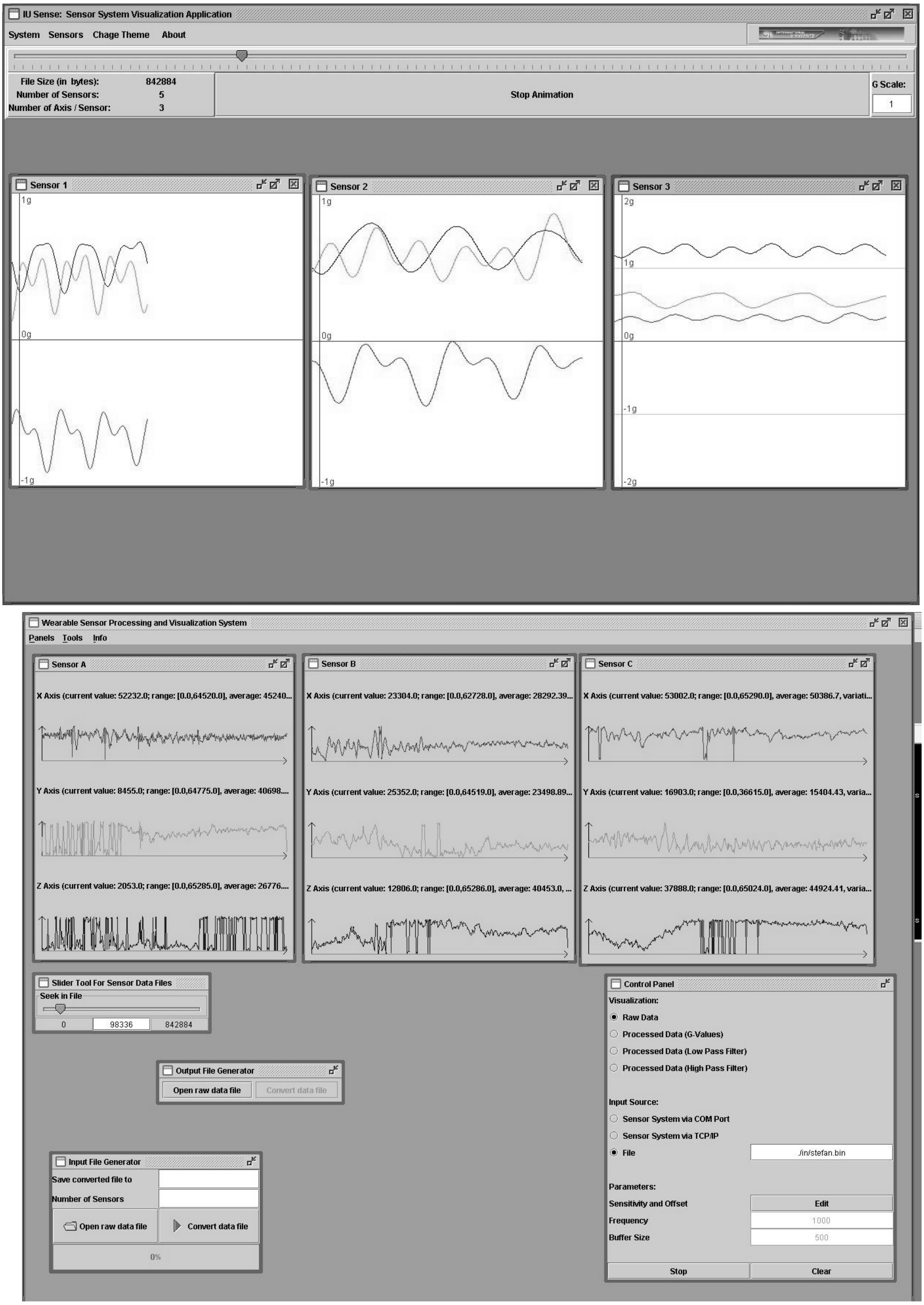

Fig. 3. IU SENSE: Graphical Data Display

2.4 Problems and Future Improvements

The trickiest task of the whole experimental setup is the proper fixation of the sensor equipment. This still takes too much time and remains somewhat sensitive to human error. In the long run, the sensors and their interconnect will be woven into some clothes directly. Until then, however, we need to be content with less professional yet nevertheless reliable alternatives. Towards this end, we have discussed and tried many different approaches based on a variety of materials and fixation schemes: strong plaster tape wrapped around the sensors and the test subjects' body parts, strings tied around the leg and knotted together, so-called "velcro straps" as used for portable music players, rubber bands shifted over the leg, etc. Solely the first of these approaches, namely the strong plaster tape, worked reliably enough in practice to be considered adequate by us. Despite its ad-hoc look, the tape solution works surprisingly well. Its main shortcomings are prolonged setup times, minor imprecisions in sensor placement, and some issues of holding strength depending on the test subjects' type of clothes. A better solution than the taping might be achieved by means of an elastic kind of bandage to be strapped over the clothes, with the sensors and their connecting cables permanently affixed to it. This is certainly worth a try.

Future experimental runs may also be improved by wireless access to the wearable computer for remote real-time control and display of the sensor data readings and recordings. Moreover, the Xybernaut MA-IV is still somewhat bulky. A smaller platform, such as a PDA like the HP-Compaq iPAQ for instance, surely impacts the test subjects even less. In fact, the iPAQ platform was our initial first choice. Unfortunately, it did not really work out in practice because some of the available iPAQ devices suffered from hardware problems connecting to the sensor system. A more light-weight setup would also allow for more precise and specific body measurements and sensor fixation within the test subjects' hip areas.

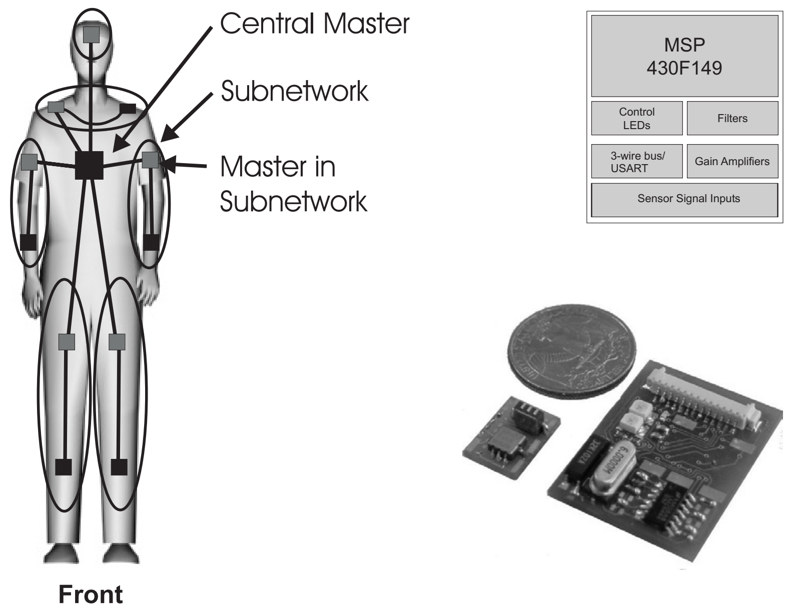

Fig. 4. The PADNET Wearable Sensor System

3 Sensor Platform

The experiments were conducted using PADNET ("Physical Activity Detection Network", see Figure 4), a sensor platform developed for user activity recognition [5]. It has been designed as a wearable system and allows for the easy distribution of multiple sensors over a person's body while being flexible. The platform consists of multiple sensor nodes interconnected in a hierarchical network. The purpose of a sensor node is to provide a physical interface for different sensor types (accelerometers, gyroscopes, magnetic field sensors, etc.), to read out the corresponding sensor signal, to provide certain computational power for signal pre-processing, and to enable communication with all other network components. Figure 4 shows such a sensor node with its logical block diagram. For the experiments, three 3D-accelerometers ADXL202E from Analog Devices were used. The analog signals from the sensors were low-pass filtered (fcutoff = 50 Hz) and A/D-converted with 12-bit resolution at a sampling rate of 100 Hz.

4 Initial Results

In the initial experimental phase of our ongoing project, we recorded data for 8 different test subjects. Of these 8, one data set has proven unusable due to a technical problem. The remaining 7 data sets were then examined with respect to three features typically used in accelerometric activity recognition:

- *root mean square* (RMS) of the signal, giving the average power of the signal;
- *cumulative sum* over the signal (sums);
- *variance* of the signal.

Of course, with just 7 subjects no statistically valid conclusions may be drawn. However, even with such a small data set useful observations can be made. In particular, proving the existence of certain phenomena such as feature variations and their consequences on classification requires isolated examples only.

4.1 Feature Variations

As a first step we investigate the statistical distribution of features for all subjects and four related context classes: walking (class #2), going down stairs (class #3), going up stairs (class #4), and walking fast (class #8). This is done for the downward-pointing axis (which is the most relevant one for all types of walking) of the hip and upper leg sensor. The results can be seen in Figures 5 to 7. The figures show the the mean value (a point) and the variance (an error bar) of each feature for all test subjects and context classes. It is interesting to note that the RMS and sums features seem to show little variations, both for each test subject individually and also between them. The only exception is test subject 6, which is totally out of range. Interestingly, from Table 1 it can be seen that nothing seems to be particularly special about this test subject except for his left-handedness which does not provide a plausible explanation for such a different walking style in our opinion. Taken together, Figures 5 to 7 indicate that variations between subjects can be expected to be a serious problem for context recognition.

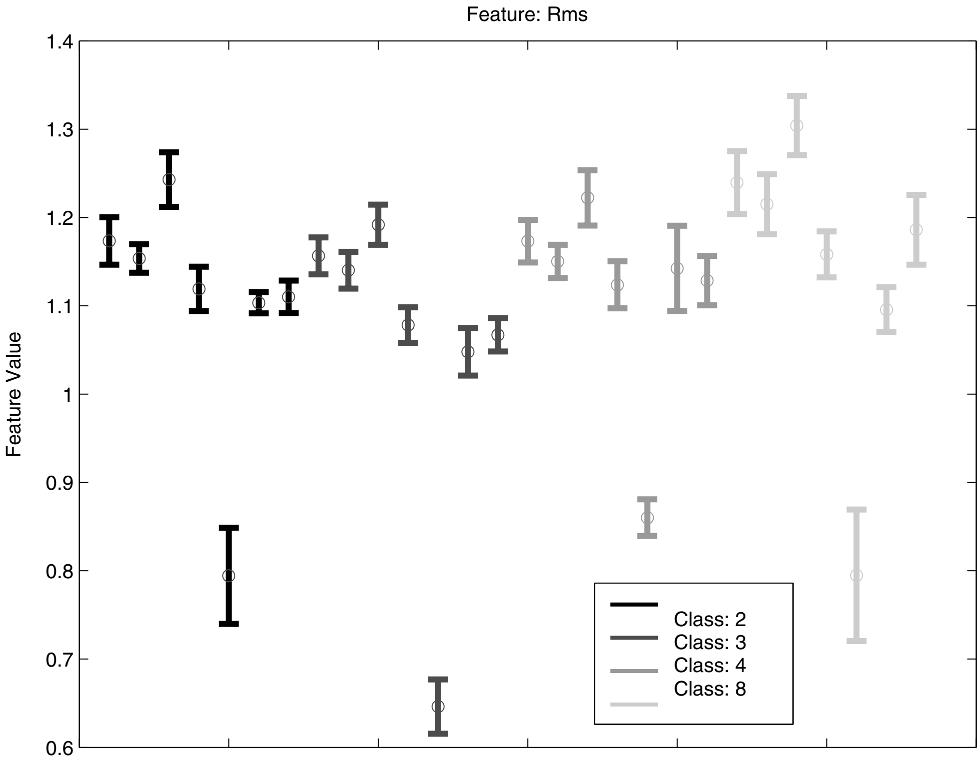

Fig. 5. Variance in the RMS Feature

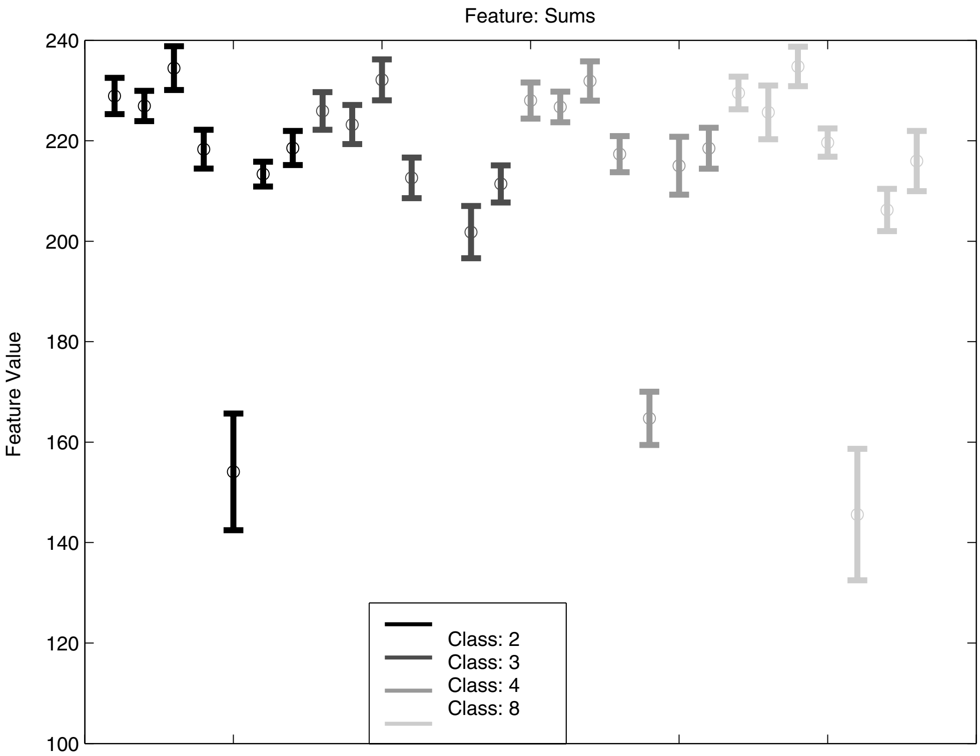

Fig. 6. Variance in the Sums Feature

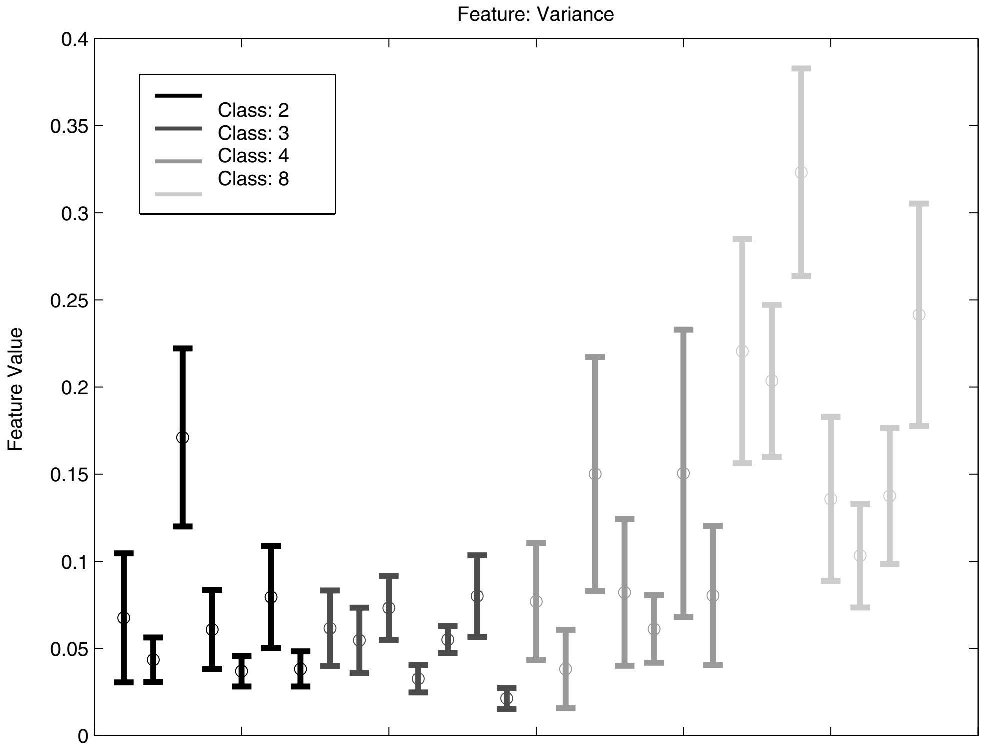

Fig. 7. Variance in the Variance Feature

Table 1. Characteristics of the 8 Test Subjects

No.	**Gender**	**Age** (Years)	**Height** (cm)	**Weight** (kg)	**Sports**	**Handed-ness**	**Special Notes**
1	male	24	173	75	weekly	right	flat feet
2	male	24	182	78	weekly	right	
3	male	36	180	90–100	monthly	right	slightly out-of-sync right foot (injury)
4	male	65	188	86	weekly	right	
5	female	23	171	62–65	weekly	right	X-like legs, asymmetrical knee-caps, wearing flip-flop sandals
6	male	21	172	70	daily	left	
7	male	23	191	110	weekly	right	wearing flip-flop sandals
8	female	23	170	50	weekly	right	wearing skirt

4.2 Usefulness for Classification

The question of usefulness of a set of features for person-independent classification involves at least two prominent aspects. First, we need to find out whether the set is universal enough to provide separation between the relevant classes for all subjects. As a stronger second requirement, we then need to see if a single separation plane can be found for all subjects enabling a person-independent system design. In our experiment, we have found that for the recognition of any two classes only (in particular, fast walking with any other) reasonable separation is achieved for all subjects by a combination of the variance and sums features. However, combining data from all subjects and then using single separation has not produced satisfactory results. When considering all 4 classes we have found the features to produce excellent separation for test subject 6. For all the others the separation was mixed to poor as illustrated in Figures 8 to 10. These figures show the values of the variance and sums features plotted in different colors using distinct shapes for each class. The main lesson of the above is that features that work very well for some people might not work at all for others. It also shows that widely used features such as RMS or variance are not directly suitable for person-independent systems. Instead, representative data sets are needed to derive more appropriate features sets.

5 Future Work

Our ongoing research project and collaboration intends to collect further data samples by repeating the exact experimental procedure described above with many more test subjects. In the end, we plan to collect enough samples to allow for statistically confident recognition schemes of the features over the whole test population. Moreover, we will make the full data and documentation of all the experiments publicly available on the Internet as soon as possible. Due to the careful design and in-depth documentation of our experimental procedure, we are convinced that our data are of common interest and value. They may even serve the purpose of a general benchmark set.

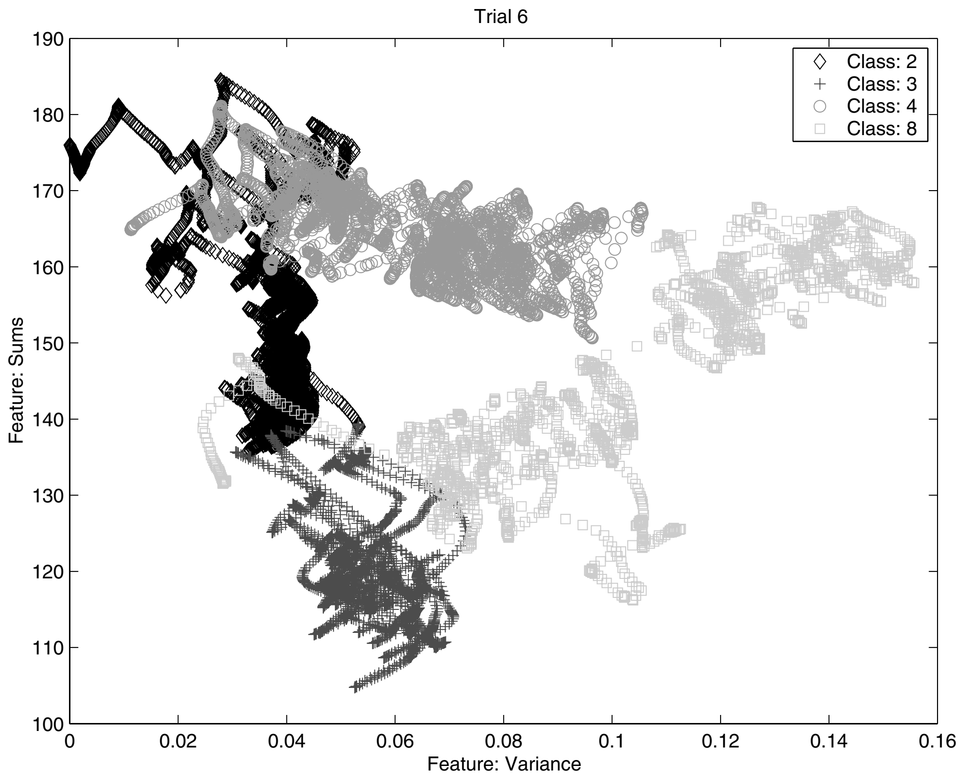

Fig. 8. Good Variance / Sums Classification

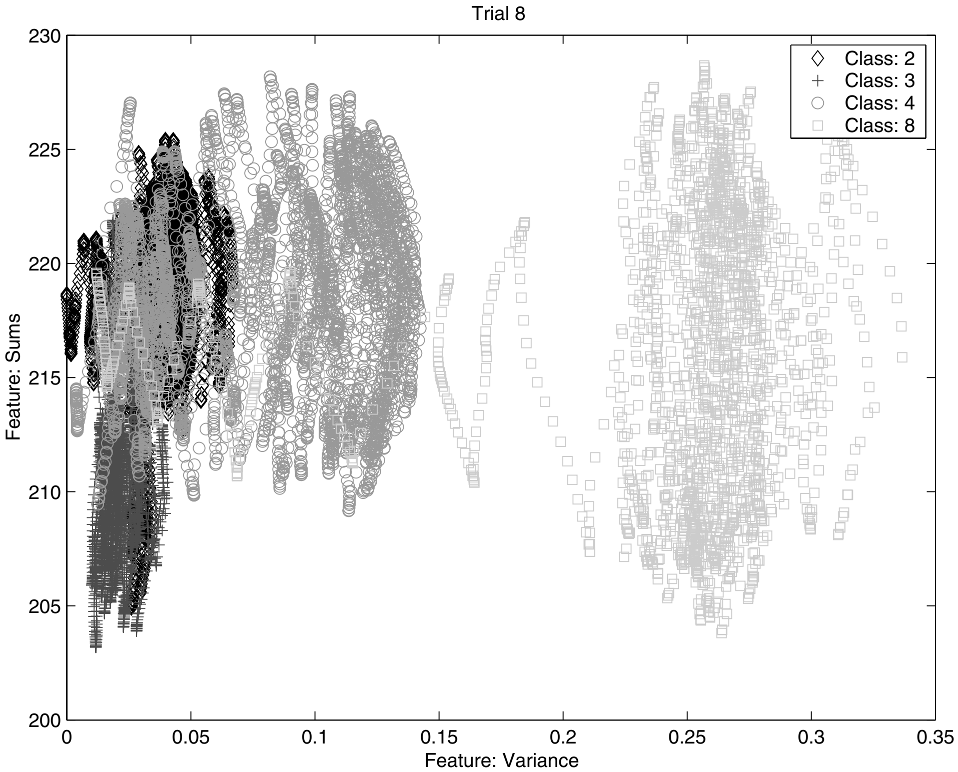

Fig. 9. Mixed Variance / Sums Classification

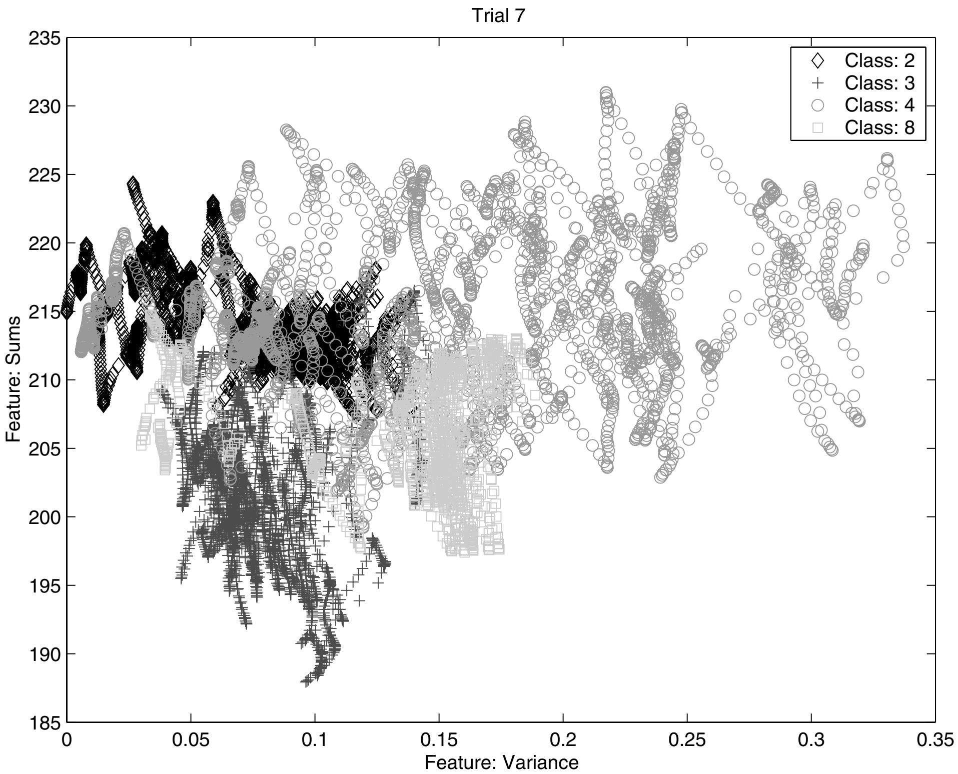

Fig. 10. Poor Variance / Sums Classification

References

1. A. Caracas, E.A. Heinz, P. Robbel, A. Singh, F. Walter, and P. Lukowicz. Real-time sensor processing with graphical data display in Java. Submitted to 3rd International Symposium on Signal Processing and Information Technology (ISSPIT), takes place in December 2003
2. H. Junker, P. Lukowicz, A. Caracas, K.S. Kunze, P. Robbel, A. Singh, S. Sulistyo, F. Walter, and E.A. Heinz. Demonstration of integrated CMOS acceleration sensors and real-time processing software with graphical data display in Java. Description of live demonstration at 1st European Symposium on Ambient Intelligence (EUSAI), November 2003
3. N. Kern, B. Schiele, H. Junker, P. Lukowicz, and G. Tröster. Wearable sensing to annotate meeting recordings. In Proceedings of the 6th International Symposium on Wearable Computers (ISWC), pp. 186–193, IEEE Press, October 2002
4. K. van Laerhoven and O. Cakmakci. What shall we teach our pants? In Proceedings of the 4th International Symposium on Wearable Computers, pp. 77–83, IEEE Press, October 2000
5. P. Lukowicz, H. Junker, M. Stäger, T.v. Büren, and G. Tröster. WearNET: A distributed multi-sensor system for context-aware wearables. In Proceedings of the 4th International Conference on Ubiquitous Computing (UbiComp), G. Borriello and L.E. Holmquist (eds.), pp. 361–370, Springer-Verlag, LNCS 2498, September 2002
6. A. Madabhushi and J. Aggarwal. Using head movement to recognize activity. In Proceedings of the 15th International Conference on Pattern Recognition, Vol. 4, pp. 698–701, IEEE Press, September 2000
7. C. Randell and H. Muller. Context awareness by analysing accelerometer data. In Proceedings of the 4th International Symposium on Wearable Computers, poster paper, pp. 175–176, IEEE Press, October 2000