Gesture Signature for Ambient Intelligence Applications: A Feasibility Study

Elisabetta Farella[1], Sile O'Modhrain[2], Luca Benini[1], and Bruno Riccó[1]

[1] DEIS University of Bologna, v.le Risorgimento 2,
40136 Bologna, Italy
[2] Sonic Arts Research Centre,
Queens University, Belfast, Ireland

Abstract. This work investigates the feasibility of a personal verification system using gestures as biometric signatures. Gestures are captured by low-power, low-cost tri-axial accelerometers integrated into an expansion pack for palmtop computers. The objective of our study is to understand whether the mobile system can recognize its owner by how she/he performs a particular gesture, acting as a gesture signature. The signature can be used for obtaining access to the mobile device, but the handheld device can also act as an intelligent key to provide access to services in an ambient intelligence scenario. Sample gestures are analyzed and classified using supervised and unsupervised dimensionality reduction techniques. Results on a set of benchmark gestures performed by several individuals are encouraging.

1 Introduction

Is it possible to use gesture as an integral part of a personal identification - authentication system? Is there something in how we perform movements which is unique and personal? If we think about handwriting, it is evident that each of us has a different calligraphic identity. Is it possible that there also exists for each of us a calligraphy of gestures? The present work investigates the feasibility of a personal verification system using gestures as biometric signatures, given a constrained scenario. We imagine a user holding her/his Personal Digital Assistant (PDA) or her/his mobile phone and unlocking or locking it through a simple gesture, a kind of gesture signature, which gives the device the ability to recognize its owner. In this example the mobile device is the target: the user wants to interact with the mobile device and access private data, such as their address book, personal notes, files and programs. The PDA can also act as a bridge to allow an individual to be identified in a more general ambient intelligence scenario. Imagine arriving home and being recognized by your house using a personal gesture signature. By performing a simple gesture, all the services you pre-programmed in your house are delivered to you, e.g. your personal mail is read to you or your favourite music is turned on.

Having selected as our target a mobile scenario, some constraints are immediately apparent [1]. Vision and optical systems for motion tracking are not

K.P. Fishkin et al. (Eds.): PERVASIVE 2006, LNCS 3968, pp. 288–304, 2006.

suitable in this context because such systems are either fixed and thereby cumbersome or they require the user to stop moving in order for an entire gesture to be captured. However inertial sensors, due to their small form factor, low-cost and power consumption characteristics appear to provide a viable alternative solution since they are suitable for integration into mobile systems such as PDAs and mobile phones [2]. Moreover gestures enable interactions with a device that do not necessarily require visual support or support from other input devices such as pens, keyboards and joysticks. The use of gesture as an input modality for mobile systems has been considered in many studies [5][15] as a suitable alternative solution to the mentioned usual interfaces. This current work takes a step forward, by exploring the challenge of exploiting inertial sensors embedded in mobile devices for personal identification and authentication. The recognition of a gesture signature is targeted, focusing only on the gesture chosen by the user as her/his personal gesture to authenticate her/himself in the system.

Our analysis was carried out using a prototype of the Mesh platform [1], an expansion pack for palmtop computers integrating 3-axes accelerometers. The prototype is a handheld box equipped with inertial sensors, able to collect accelerations along three orthogonal axes. Results and observations from analysis based on gestures collected from a sample of individuals are presented. Feature extraction from data is implemented using two well known reduction techniques, namely Principal Component Analysis [18] and Locally Linear Embeddings [19]. We demonstrate that results are sufficiently robust to proceed with this line of investigation (e.g. with refining the analysis, increasing the data set and using additional sensors).

Gestures collected in the form of accelerations through low-power, low-cost inertial sensors can be used to authenticate people within a small group, such as to distinguish among members of a family or colleagues who are members of the same group at a workplace. In the following sections we describe prior related work in this area and position our work in the context of biometrics and interaction techniques for mobile devices. Moreover, we describe the chosen dimensionality reduction techniques applied to the data collected from a sample of users. Finally, the results of the feasibility experiments will be presented and discussed.

2 Background

2.1 Interaction Techniques for Mobile Devices

Mobile devices (PDAs, mobile phones, etc.) present unique and specific challenges in terms of interaction design and usability [1]. Designing interfaces for mobile computers is complicated in a mobile setting where the users attention is not fixed on the computer, but on real-world tasks. A limited amount of screen area is available in such devices and the users visual attention is often focused on negotiating their surroundings rather than on the interface [3]. Alternative interaction techniques for mobile devices, which do not use standard pens and touch panels, are based on use of embedded sensors [5]. In particular, inertial sensors are used to exploit changes in position and orientation of the PDA or the mobile phone as input [4]. Tilt and motion based interfaces enable single-handed operation. Interaction is thus minimally disruptive and demanding of cognitive and visual attention. Over the course of a day, a mobile device is picked up many times and typical natural gestures, which regularly occur when using a mobile phone or MP3 player can become an integral part of interaction with the device.

2.2 Gesture-Based Biometrics

The use of inertial sensors in building alternative gestural interfaces has been extensively explored. Because of their reduced size and weight, however, they are also suitable for applications such as signature capture [26] and gesture recognition [14] [6], where detecting movements can be of great help. More innovative is the suggestion that movements collected through inertial sensors can be used for biometric purposes. A biometric is a physiological or behavioural feature that can be used to identify people [16]. In physical biometrics, biological features (e.g. fingerprints, hand geometry, retina or facial characteristics) are examined in order to identify an individual. Behavioural recognition examines the mannerisms of an individual, including signatures, handwriting, voice and keystroke patterns and so on (references can be found in [27]). More generally, techniques for authentication can be based on one of several possible attributes:

– Something you are (a biometric)
– Something you know (a password or PIN)
– Something you have (a key, token card, etc)

Gestures have been used as authentication techniques based on "something you know", that is the gestures or the sequence of movements performed is chosen by the user as he/she might choose a password or a code number. In [7] personal identification is proposed using hand gesture patterns expressing a code number, captured by a CCD camera. A sensor-based authentication mechanism for mobile devices has been presented in [8]. This work explores the problem of verifying a user identity when accessing the public infrastructure, e.g. when he wants to annex his device to I/O resources encountered in the local environment. The recognition mechanism is based on a sequence of shake and pause actions detected by inertial sensors integrated into the mobile device. The sequence is sent to the public infrastructure by the user's mobile device after a discovery procedure has identified the presence of the device.

The aim of our work is somewhat different: using gestures and arm movements as biometrics. In fact, we do not simply want to distinguish among gestures, but to investigate the feasibility of a system, appropriately trained, to distinguish/identify the gesturer from the gesture made. As a consequence, gestures are treated as a behavioural biometric ("something you are"). Thus, the question we pose is whether people can be identified by the way they move. In this area literature is scarce. In [9], simple filters are used to extract features from a gesture captured in the form of still frames, with the purpose of introducing a biometric measure based on hand gestures. Unfortunately, a complete description of this work is not publicly available. In work by Gupta [10] the same work is cited and the authors further state that the algorithm applied could not perform recognition accurately enough to use gestures as biometrics.

Encouraging results come from gait recognition, a relatively new area of study, receiving growing interest, within the realms of computer vision [11]. Gait recognition is the process of identifying an individual by the way in which she/he walks. Early psychological studies into gait by Murray [12], suggested that gait was a unique personal characteristic, with clear cadence and was cyclic in nature. Johansson [13] carried out studies by attaching moving lights onto the principal joints of a group of participants who were than asked to walk across a darkened room. He then showed movies of these "light-point walkers" to a second group of observers. The observers could recognize the biological patterns of gait from the moving light displays, even when some of the markers were removed, once again indicating gait as a potential biometric. Even if conducted in the field of computer vision, these studies suggest that the way we move is personal. Thus it is worth investigating whether it is possible to use data collected with sensors other than cameras.

3 Apparatus and Analysis

Biometric systems typically involve several stages of processing. Data derived from behavioural or physiological characteristics are converted into templates, which are used for subsequent matching and decision-making processes. The work flow in our case is described in Figure 1.

The first decision was the choice of the data acquisition device and consequently the nature of data was determined. We chose the Mesh platform (described later) equipped with inertial sensors. Thus, data collected are accelerations along three orthogonal axes. As a second step, data must be collected from the user and submitted to the system. As it is impossible to collect all the possible samples, a selection of a given amount and kind of sample, considered meaningful and representative for the purposes of the investigation, must be made. A set of four different gestures was selected. For each gesture, a number of examples were collected for each person in the sample group. Data collected were then windowed and grouped in matrices. Rescaling was applied to prepare data for dimensionality reduction. Two different unsupervised dimensionality reduction techniques were applied (PCA and LLE), as described in section 3.4. Being an exploratory study we performed two kinds of analysis. A first qualitative analysis consisted in using data from the initial dimensionality reduction phase to obtain a graphical representation of relationships among gestures performed by different people. Original data are variable, high-dimensional and complex to analyze. As will be described later, PCA and LLE are feature extraction techniques that can be used for dimensionality reduction, to eliminate data redundancy and extract representative vectors from a large amount of data. Reducing data to the first two or three feature vectors, it is possible to plot them and qualitatively identify data clusters, evaluate distances among data representing the same

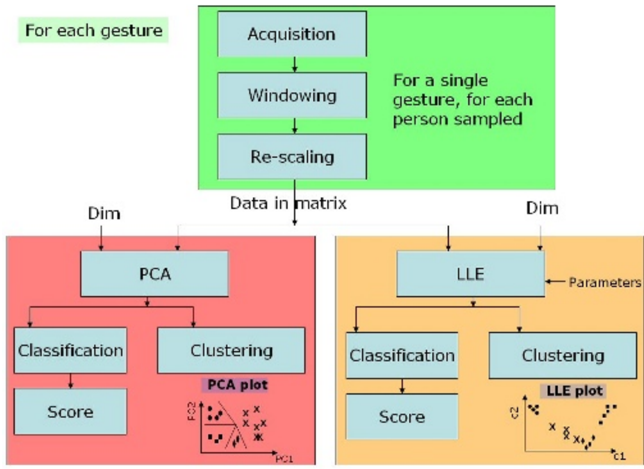

Fig. 1. Steps of the feasibility analysis

gesture performed by different people plotted in a bi-dimensional graphical space. A plot provides a fast but only a qualitative idea with respect to the separation of gestures performed by different individuals. Therefore, a second quantitative evaluation phase was necessary to validate the analysis. For the purpose, we applied a k-Nearest Neighbor method to obtain scores indicating how much a given gesture was distant from another gesture of the same type performed by the same person.

3.1 Acquisition Device

A prototype of the Mesh platform (Figure 2.a) was used to acquire gesture samples. Mesh [24] is an expansion pack for IPAQ handheld computers featuring vibrotactile output and input in the form of motion sensing. The prototype is equipped with a 3-axes accelerometer and can be connected to other devices through the serial port. The accelerometers are two biaxial sensors (ADXL202E), each mounted along and in line with the principal axes of the box prototype, i.e. orthogonally to each other. Thus, gestures are collected as three arrays of samples representing the accelerations referred to axes x, y, z of the box (see Figure 2.a).

The frequency response of the device extends to DC, allowing the acceleration due to gravity to be monitored. Their bandwidth stretches to 100 Hz, yielding sufficient temporal resolution to capture data to drive gesture recognition algorithms. For the work described here, the data is gathered from the sensors at 100Hz, and transmitted over an RS232 serial link to the Personal Computer (PC), where data analysis was performed using Matlab software. The data rate is sufficient for the purpose of this work because human movement frequency is predominantly in the range from 0 to 30Hz.

a. b.

Fig. 2. a. Acquisition device (Mesh platform) and b. The four gestures chosen

3.2 Data Collection

Two groups of tests were performed, using the apparatus already described. The first was under more controlled conditions than the second. Indeed, the first test involved collecting gestures from a participant who was imitating the gesture performed by a second person. The aim was to obtain intra-personal consistent data, reducing the variability in gesture duration and shape. For the second test no guidance was provided, thus the gestures which resulted were more realistic and hence affected by more variability.

First group of tests. This set of tests involved collecting four different gestures of the type described in figure 2.b from a small group of people. In particular we collected the right arm opening and closing horizontally (gesture 1), the rotation of the wrist (gesture 2), a gesture similar to answering a phone (gesture 3) and a gesture consisting of touching the left shoulder (gesture 4).

Each of the 4 gestures was collected 20 times. The first ten times people were asked to repeat the gesture as consistently as possible, especially with respect to the duration of the gesture. Visual feedback about gesture duration was provided by a clock visible on the PC desktop. People were also asked to pay attention to the inclination of the box during the movement, especially when the direction of movement changed. We obtained a relatively high degree of intra-personal consistency and repeatability among gestures. For the second part of this test we modified the procedure slightly, asking people to imitate a gesture performed by a second person, hereafter called the "target". In practice, *the target* is performing the role of the device owner, and the other participants are trying to imitate his gestures in order to access his personal device. The *target* performed the gesture in synchrony with the people trying to imitate it. In this way the duration of the gestures was the same and each person had immediate visual feedback about the gesture while performing it.

The second group of tests. The second group of tests focused only on gestures 1 and 2 from the initial group of four gestures. A group of 10 people, none of whom had participated in the first test, were asked to perform the gestures and less help was provided to guarantee intra-repeatability of the gesture. This time the gesture was shown only once and the person proposing the gesture did not perform it in synchrony with the participants. Thus, each gesture was repeated by each participant only ten times. We expected that the gestures collected in this second tests would result in less intra-repeatable, but also more personal.

Qualitative results. A major problem in conducting experiments is supporting users for gesture repeatability. In our case we distinguish between performing recognition of a person by the way she/he generally moves and targeting a specific gesture chosen and performed by that person as her/his "gesture signature". From this perspective the same problem can be encountered in other behavioural biometric studies, such as signature recognition. Signature recognition is different from handwriting recognition, mainly for repeatability of data [25]. Handwriting has more intra-personal variability than signatures do. While performing signatures people try to be more consistent with some prototype. They have a real or mental image to imitate each time. Handwriting is generally more variable in time. In this sense our work is aimed at finding a "gesture signature", not a "gesture-writing style". Unfortunately a gesture does not have the same feedback as a signature, which is still visible while and after being performed. This suggests that a way of providing feedback must be found also for gestures: the gesture signature must be experienced while and after it has been performed. Further, it must be easy to remember in order to increase its repeatability.

Instructions given to users about how to perform the gestures were both verbal and visual: in the first test the gesture was explained and simultaneously demonstrated. In the second test, less external support was provided. The gestures were performed only once by *the target*. Despite the invitation to carefully observe and imitate him, people performed the gestures clearly in their own way, especially in the second group of tests, as evidenced by their different physical characteristics and postures and by the different speed of gestures and orientation of the box during the trajectory path (e.g. uncertainty in the initial position of the device). Moreover people did not pay attention to the instruction to repeat each gesture with the same duration. This affected especially gesture 1, which is wider and which therefore takes longer to perform. Moreover since it is wider it is more difficult to control and repeat in the same way. We can also observe that the gestures chosen have different characteristics. The first one is more dependent on physical characteristics, such as the length of the arm of the person performing the gesture. Moreover, this gesture was also reported to be uncomfortable to perform, because it draws attention to the user from passers-by. The

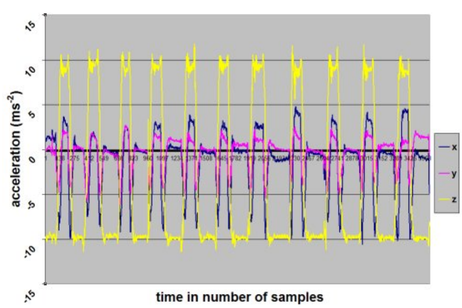

Fig. 3. Accelerations vs time collected along the 3 axes of right-wrist rotation for one person

second gesture is smaller, easy to repeat and was performed with approximately the same duration by many users. The waveform resulting from the acquisition of the three acceleration data streams over time is thus a periodic wave (Figure 3). This gesture is more intra-personally consistent but at the same time we expected more extra-personal similarities, because it is less dependent on individual body characteristics. The last two gestures can be said to provide physical reference to the body. It is expected that allowing the body to act as a frame of reference in this way will improve repeatability, because the gesture is spatially defined having a starting and final position on two precise points of the body.

3.3 Data Pre-processing: Creation of a Biometric Sample

The data collected from a single participant were arranged into a 10 column matrix by gesture. Since people did not perform each of the ten gestures at the same speed, the resulting columns were of different lengths. Variability in data duration was greater in the second group of tests. Thus, to process data with Matlab functions some columns were truncated and others padded. This was done by finding an average duration value (ADV) for each participant and then padding data where a column length did not reach the ADV or cutting extra-samples where a columns length exceeded it. In our opinion this operation is not critical since the differences between columns belonging to the same person and their ADV is in general not significant (e.g. 20-40 samples for a whole duration of 300-400 samples at a sampling rate of 100Hz). A single column vector $\bar{g}$ is shown in Figure 4. The number of samples is averaged to a given ADV' related to the participant and to the length of the gesture performed. Samples s_i ($i = 1 \ldots ADV'$) are related to the temporal window WIN_j with $j = 1 \ldots 10$. Accelerations a_x, a_y, a_z along the three axes are consecutively filled in the column vector. The matrix containing 10 gestures from a single participant is matrix $\Gamma^{P\varphi}$, where $P\varphi$ is an identifier for a given participant ($\varphi = 1 \ldots N$, where N=number of participants) represented in Figure 4.

The differences between individuals, as represented by their ADV, are far more significant (e.g. 100-200 samples), especially for the second group of tests. Thus many possible solutions can be found to adjust data in preparation for applying dimensionality reduction techniques. One strategy could be to compare only people whose gestures have the same ADV or an ADV that is not significantly different. Otherwise a duration value corresponding to the maximum, the minimum or an average duration value among different users could be fixed and user's data padded. In this case, different choices can be made for the value to use for filling the columns. In our case the value chosen to complete the columns was the data value corresponding to the rest position. This choice was verified as the option that would minimally influence the data processing and thus have less impact on the analysis. In conclusion, we identified subgroups of P participants having for a given gesture similar ADVs, (i.e. a difference in length smaller then 10%). In Figure 4 we identified as ADV" the average of the ADVs related to the group of participants selected. Each column vector $\Gamma'^{P\varphi}$

$$\bar{g} = \begin{bmatrix} a_x(s_1) \\ \cdots \\ a_x(s_{ADV'}) \\ a_y(s_1) \\ \cdots \\ a_y(s_{ADV'}) \\ a_z(s_1) \\ \cdots \\ a_z(s_{ADV'}) \end{bmatrix}_{WTNj}$$

$$\Gamma^{P\varphi} = \left[\bar{g}_{WTN1} \quad \bar{g}_{WTN2} \quad \cdots \quad \bar{g}_{WTN10}\right] = \begin{bmatrix} a_x(s_1)_{WTN1} & a_x(s_1)_{WTN2} & \cdots & a_x(s_1)_{WTN10} \\ \cdots & \cdots & \cdots & \cdots \\ a_x(s_{ADV'})_{WTN1} & a_x(s_{ADV'})_{WTN2} & \cdots & a_x(s_{ADV'})_{WTN10} \\ a_y(s_1)_{WTN1} & a_y(s_1)_{WTN2} & \cdots & a_y(s_1)_{WTN10} \\ \cdots & \cdots & \cdots & \cdots \\ a_y(s_{ADV'})_{WTN1} & a_y(s_{ADV'})_{WTN2} & \cdots & a_y(s_{ADV'})_{WTN10} \\ a_z(s_1)_{WTN1} & a_z(s_1)_{WTN2} & \cdots & a_z(s_1)_{WTN10} \\ \cdots & \cdots & \cdots & \cdots \\ a_z(s_{ADV'})_{WTN1} & a_z(s_{ADV'})_{WTN2} & \cdots & a_z(s_{ADV'})_{WTN10} \end{bmatrix}$$

$$G = \left[\Gamma'^{P1} \quad \Gamma'^{P2} \quad \cdots \quad \Gamma'^{PP}\right]_{ADV''}$$

Fig. 4. General structure of the column vector $\bar{g}$ for a single gesture, of the matrix $\Gamma^{P\varphi}$ and of the biometric dataset G

corresponds to $\Gamma^{P\varphi}$ padded to ADV" instead of ADV' (with $\varphi = 1 \ldots P$). The biometric dataset created is G, which consists of a $[3 \times\ ADV'',\ P \times\ 10]$ matrix where ten consecutive columns correspond to a single user.

3.4 Dimensionality Reduction: PCA and LLE

Because this is an early stage of this investigation it may be valuable to perform exploratory data analysis to gain insight into the nature or structure of the data. Unsupervised methods are good for this purpose, because they provide a form of data-dependent "smart pre-processing" or "smart feature extraction". The discovery of distinct subclasses - clusters or groups of patterns whose members are more similar to each other than they are to other patterns - or of major departures from expected characteristics is an important input when designing the classifier. Dimensionality reduction is a useful operation for data clustering and pattern recognition. High-dimensional data can contain a lot of redundancies and correlations hiding important relationships among data. The purpose of dimensionality reduction techniques, which can be based on both linear and nonlinear methods, is to ease the analysis of data, eliminating redundancies and reducing the amount of data to be processed.

Here, we will briefly describe Principal Component Analysis (PCA) [18] and Locally Linear Embedding (LLE) [19] [23]. In this work, we apply both methods to the exploratory analysis and visualization of data sets. Both methods are unsupervised procedures for mapping high-dimensional data to a lower-dimensional space. We chose PCA because it is a powerful linear method, widely and traditionally used in many different application fields. LLE is an example of a nonlinear approach that, even if perhaps less tried, has demonstrated robustness and has produced a number of interesting results as shown in other fields [23] [21].

Principal Component Analysis. Principal Component Analysis [18] is a linear method for dimensionality reduction that projects the data into the subspace with a minimum reconstruction error. PCA is a useful statistical technique that has found application in fields such as face recognition and image compression, and it is a common technique for finding patterns in data of high dimensionality. Data processed with the PCA technique are expressed in such a way that their similarities and differences are highlighted. Since patterns can be hard to find in data of high dimensionality, where the luxury of graphical representation is not available, PCA is a powerful tool for analyzing such data. PCA is often used also for compressing data, since it helps to reduce the number of dimensions, without much loss of information. Basically PCA transforms input data so that they are expressed in terms of the patterns between them, where the patterns are the lines that most closely describe the relationships between the data. The result of this simple algebraic technique may be seen from several points of view, either as a variance preserving projection, or a minimal reconstruction error projection, or as a distance preserving projection.

Locally Linear Embedding. Though widely used for its simplicity, PCA is limited by its underlying assumption that the data lies in a linear subspace. Recently, several algorithms for nonlinear dimensionality reduction (i.e. [17]) have been proposed that overcome this limitation of PCA. Like PCA, these algorithms are simple to implement, but they compute nonlinear embeddings of high dimensional data. So far, these algorithms have mainly been applied to data sets of images and video, where they have revealed low dimensional manifolds not detected by purely linear methods. One of these algorithms is Locally Linear Embedding.

LLE is a recent method for data analysis, an unsupervised learning algorithm that computes low dimensionality, neighborhoods preserving embeddings of high dimensional data [23] [21]. LLE attempts to discover nonlinear structure in high dimensional data by exploiting the local symmetries of linear reconstructions. Notably, LLE maps its inputs into a single global coordinate system of lower dimensionality, and its optimizations, though capable of generating highly nonlinear embeddings, do not involve local minima. Because LLE is a new method, it is not yet well known. Thus, it is worth describing it briefly.

As an input, the LLE algorithm requires N points, for example corresponding to N samples of gestures from the same person. We define D as the dimensionality of the original sample and d the embedding dimensionality after applying LLE. In our case for example D is $3 \times \ ADV$, while d is 2. Each point is a X_i, where $i \in [1, N]$, $Xi \in R^D$. As an output, it gives N points, again the N gestures, re-mapped in a new vector space with lower dimensionality. Thus, each output point is an Y_i, $Y_i \in R^d$, where i$\in [1, N]$, and $d \prec\prec D$. The output is such that geometrical proprieties of the input set of points are locally best preserved. The algorithm consists of three steps:

- Step 1. For each X_i find its K nearest neighbors $X_{i1} ... X_{iK}$.
- Step 2. Measure the reconstruction error resulting from the approximation of each X_i by its nearest neighbors and compute reconstruction weights W_{ij} minimizing this error.

– Step 3. Compute low-dimensional embeddings best preserving the local geometry represented by the reconstruction weights.

In Step 1 the Euclidean distances are used to determine a neighborhood around each X_i, though other definitions of "closeness" are possible as well. Step 2 assumes that the manifold is well-sampled, i.e., there are enough data, each data point and its nearest neighbors lie on or close to a locally linear patch of the manifold. Hence, we can approximate each sample X_i by a linear combination of its neighbors. This is equivalent to approximating the nonlinear manifold in the vicinity of X_i by the linear hyperplane passing through $X_{i1}...X_{iK}$. To do so, we need to minimize the reconstruction error. In Step 3 the low-dimensional embeddings are found which best preserve the high-dimensional neighborhood geometry represented by the weights W_{ij}.

Unsupervised clustering techniques provide a first step analysis in order to evaluate separation of data. Afterward, we applied to the processed data a basic classification algorithm, the *k-nearest neighbor classifier*, described below. This additional investigation is carried out in order to provide a quantified evaluation of the results coming from PCA and LLE dimensionality reduction. Their application to gesture data will be described in detail in Section 4.

4 Analysis of Results After Feature Extraction

To summarize, tasks accomplished to this point are shown in the upper part of the Figure 1. As already stated, for each kind of gesture and for each person, data are acquired, segmented and re-scaled. Data are then grouped into a matrix, referred to in the following paragraph as X, where each column vector is $3\times\ ADV$ in length, since each column contains time samples of the acceleration along three orthogonal axes. Subgroups of ten consecutive columns represents a given gesture performed by the same person. The PCA script receives as input the matrix A and the parameter d, which is the number of principal components requested (Figure 6 left side). In practice we reduced data dimensionality to $d = 2$ (or 3). Thus, as output, we obtain a matrix B still containing $P\times\ 10$ columns, but each column is only d samples in length (in our case two or three). In B data are still organized so that subgroups of ten consecutive columns represents a gesture belonging to a given user, with a reduced dimensionality w.r.t. the original input matrix A. Thus, associating a different symbol with each ten columns, it is possible to plot data belonging to the same user with the same symbol. Each point in the plot represents a gesture mapped in the space defined by the first two principal components ($d = 2$). Results are satisfying, as can be seen in Figure 5 a and b. Here clusters are easily distinguished. In the plots we outlined the separation in clusters of data belonging to different people dividing the plane with approximate lines. The plot represents only gestures 1 and 2, but similar results were also observed for gestures 3 and 4 when a similar process was used for the LLE method (Figure 6 right side). Again, the LLE script used [19] receives as input matrix A. Moreover, two parameters are requested: d the embedding

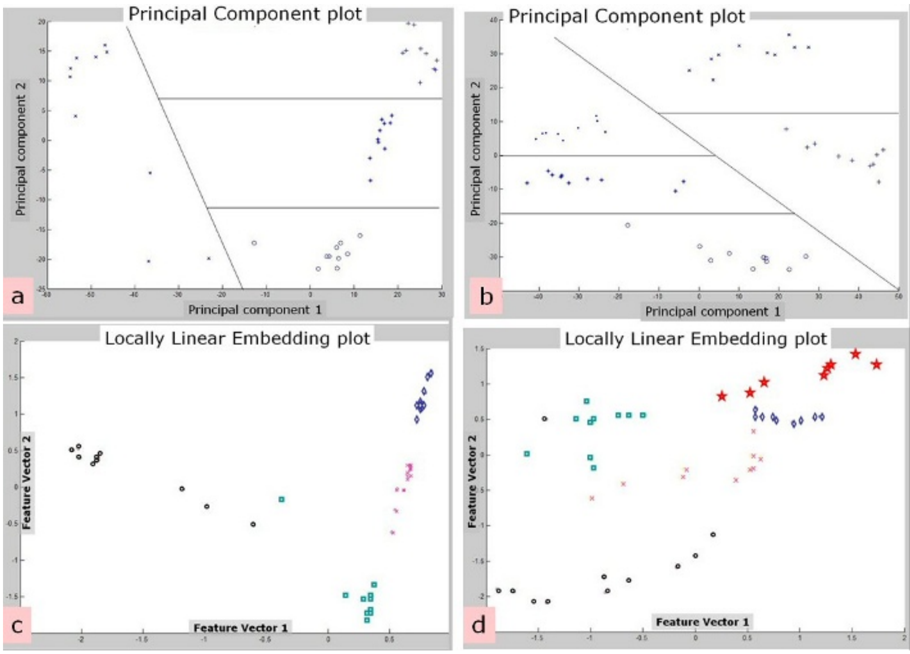

Fig. 5. a) and b) PCA applied to gesture 1 and 2 respectively. In **a** data are collected from 4 different people; the input matrix is 750x40. In **b** data are collected from a further 5 different people. The input matrix is 540x50. In **c** and **d** LLE is applied to gesture 1 (with the same input data as plot **a**; $K = 12$) and 2 (the same input data as plot **b**; $K = 13$) respectively.

dimensionality and K the number of nearest neighbors. While d was fixed at 2, we tested different values of nearest neighbors K. In [22] it is observed that the results of LLE do not depend considerably on the choice of the number of nearest neighbors. However several criteria are indicated to help the choice of K. One of them is based on the fact that the algorithm can only be expected to recover embeddings whose dimensionality, d, is strictly less than the number of neighbors, K, and some margin between d and K is desirable to improve the algorithm's robustness. For us the choice of a value for K around 5 or lower means that we are using as neighbors almost always gestures from the same person, with the result that data would seem falsely divided. In fact our matrices have ten by ten gestures from different people. If we chose a too large value for K (e.g around 20) we thereby loose the advantage of the algorithm in terms of showing local properties and thus enhancing differences among data. The algorithm, in fact, is based on the assumption that a data point and its nearest neighbors can be modeled as locally linear; the more the manifold is curved, the more choosing K too large will violate this assumption. That is why we preferred a value for K between 11 and 15. Low variability of results is experienced while choosing one of the value in this range. We can conclude that results are stable over a middle

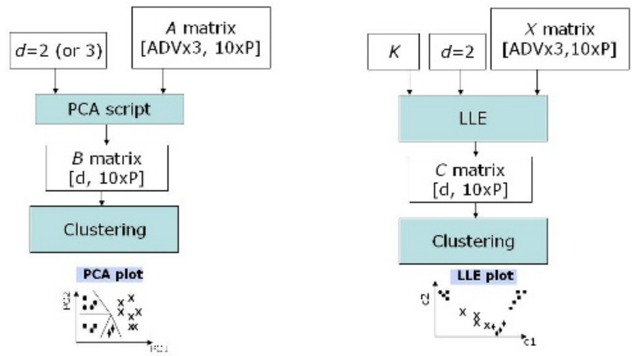

Fig. 6. Data processed by PCA (leftside) and LLE (rightside) scripts

range of values but do break down as K becomes too small or large (i.e. out of the proposed range).

The LLE script provides as output a matrix C, which has $P \times 10$ columns and d rows. The same technique used for PCA is applied to C, therefore different symbols are associated with different people. The LLE plots shown in Figure 5 c and d are again examples relating to gestures 1 and 2. Each symbol represents a gesture mapped in the 2-dimensional space ($d = 2$) obtained after dimensionality reduction performed by LLE.

As already mentioned, the analysis was limited by the different duration of gestures among people, so that it was impossible to compare all the people with a unique matrix without affecting the original data (either truncating or padding the columns of the matrix. The second gesture was more difficult to cluster, again, with both techniques. But the general result is satisfying, since this gesture is really simple and easily repeatable, thus easy to fake. Gesture 3 obtained very good separation. Gesture 4 has lower scores because during the sample acquisition process, we observed that people often performed the gesture incorrectly, thereby explaining the high variability in the intra-personal data collected. To reduce variability in such cases a possible solution is to augment the number of training samples.

An attempt was made to relate physical attributes of the participants (height, length of limb, etc.) to the results obtained from PCA and LLE. We tried to understand whether physical similarities could explain the overlapping clusters, but without success. People corresponding to overlapped results are different in weight, height and gender. This fact probably confirms that physical characteristics cannot explain the way we move, if they are not integrated with behavioral information.

4.1 Classification: KNN

In order to provide a quantitative evaluation of the qualitative results from PCA and LLE presented in the previous section, we applied a supervised technique called k-nearest neighbor, briefly described below. To allow for direct comparison of the following analysis with the preceding graphical qualitative analysis, we further decided to evaluate data in a bi or tri dimensional space. In future work we will explore the possibility of using higher dimensions to obtain better scores.

Evaluation through k-nearest-neighbor classifier. The k-nearest neighbor classifier (kNN) labels an unknown object or point (e.g. a gesture sample) with the label of the majority of the k nearest neighbors [20]. A neighbor is deemed nearest if it has the smallest distance, in the Euclidean sense, in feature space, which is in our case the space obtained after applying PCA or LLE. For $k = 1$, this is the label of its closest neighbor in the learning set. Thus, using a training matrix containing a set of samples for which the label is known, it is possible to classify each new sample, with an unknown label, into one of the groups in the training matrix, calculating its distance in the training space.

The discrimination function implemented by this classifier will in general be an irregular, piece-wise linear function since it is influenced by each object available in the learning set. A disadvantage of this method is its large computing power requirement, since for classifying an object its distance w.r.t. all the objects in the learning set has to be calculated.

Workflow. We applied the k-nearest neighbor classifier with $k = 1$ both on the results coming from the PCA analysis and the LLE technique with different values of K. The following are the steps in both cases:

1. Load a matrix having gestures as columns (subgroup of ten column for each participant).
2. Perform dimensionality reduction with PCA or LLE. In the case where PCA is used, the output matrix is the set of gestures projected in the plane defined by the first three Principal Components. In the case of LLE the output matrix is the set of gestures projected in a two-dimensional space after having been processed with the LLE algorithm with a given value of K (ranging between 5 and 30).
3. Apply the kNN classifier. The algorithm requires as input a reference matrix, containing the training samples for which the classification is known, and one or more samples for which the classification is unknown. Each unclassified sample is assigned to a given group using the nearest neighbor method. In our case the different groups correspond to the different participants in the study. The training matrix is derived from the matrix obtained after feature extraction at step 2 extracting one column. This column, corresponding to a gesture, becomes the sample to be classified with the kNN method. If the gesture is 'near', in the meaning of kNN, to the other nine gestures performed by the same person, it will be classified as belonging to the right group. One by one, all columns are extracted from the matrix obtained at step 2, each becoming the sample to classify.
4. The result of the processing is registered in a vector, which is compared with another vector containing the expected results to obtain the percentage of matching. Thus the percentage says how well samples, processed with PCA or LLE, are grouped in clusters and therefore how well gestures performed by a given person are separated from the same gestures performed by another.

Table 1. kNN scores for first and second set of experiments

NofPeople	% PCA	Best % LLE	gest
3	0.8333	1	1
3	0.8667	0.9333	1
4	1	0.95	1
3	0.9667	0.8	2
3	0.9667	1	2
4	0.8	0.7	2
3	1	0.8	3
3	0.9333	0.9667	3
4	0.8667	0.875	3
3	1	1	4
3	0.7333	0.7333	4
4	0.6333	0.675	4

a) I set of tests

NofPeople	% PCA	Best % LLE	gesture
5	0.84	0.74	2
5	0.8	0.88	2
5	0.98	0.96	2
3	0.9667	0.9667	1
4	0.9750	0.975	1
2	1	1	1
3	0.9667	0.9667	1
5	0.98	0.96	2
5	0.98	...	2
4	0.9750	0.975	2

b) II set of tests

Score and comments. Tables 1.a and 1.b summarize the results coming from the application of KNN. The first column indicates the number of users which is related to the number of gestures stored in the matrix processed (10 multiplied by the number of users). The percentage of matching indicated in the second column of the table refers to the data described in three-dimensional vectors, that is gestures re-mapped along the first three principal components. The third column refers to the best score for data processed with LLE, using different values for K ranging from 10 to 30. The last column indicates the gesture to which analysis is referred. The percentages both for LLE and PCA are in the majority of cases satisfying, but the number of people in each group is low.

The Table 1.a is also organized in groups of three rows. The information is organized as in Table 1.b. Each row describes the percentage of correct classifications after applying LLE or PCA and kNN techniques. The groups of three rows are dedicated to each kind of gesture; the difference among them is related to the data processed. The first row refers to the first ten gestures collected from people, the second row refers to the second ten gestures collected asking people to imitate a target gesture, the third row refers to the same data as in the second row but adding also the ten gestures from the person being copied. It can be seen that the PCA technique provides better scores compared to LLE. In general very good results are obtained except for gesture 4. Note that the plot shown in Figures 5 a and b refers to a 2D space (defined by the first two principal components), while the vectors to which the kNN has been applied are mapped in the 3D space defined by the first three principal components. This improves in some cases the separation of data.

5 Conclusion

The present work investigated the feasibility of using gestures as biometrics, asking whether it is possible to distinguish someone by the way she/he performs gestures. The study was restricted to four hand gestures performed by holding a box with motion sensors embedded in it and collecting acceleration data along three orthogonal axes. Gestures can be considered as behavioral biometrics, and we therefore expected a less defined separation of data between individuals than is found in the case of physical biometrics. In this respect our results shows that percentage of matching is high even with a simple linear cluster method like PCA. Use of a non linear method does not yield significantly better results and has the additional shortcoming of adding the value of a tuning parameter. However, both PCA and LLE lead to a high percentage of matches, around 95% or higher in the majority of cases, thus both techniques can be considered valuable for our purpose. Nevertheless, many improvements to these procedures are possible, e.g. integrating in the acquisition device other kinds of sensors (gyroscopes, bend sensors) into the data acquisition apparatus, analyzing data in higher dimensional spaces and applying more powerful techniques for data pre-processing and feature extraction.

In conclusion, this study shows that for small groups of people (e.g in families or members of small work groups), it is possible to distinguish one person from another by the way gestures are performed. Thus, a system using this result can potentially be useful in Ambient Intelligence applications for context-aware services and application profiling. Moreover, this biometric technique in combination with other biometrics can enforce security policy to log-in to distributed systems or to control access to restricted areas and protected physical environments.

References

1. I. Oakley, J. Angesleva, S. Hughes, S. O'Modhrain: Tilt and Fill:Scrolling with Vibrotactile Display. Proc. of EuroHaptics 2003 (2003)
2. `http://www.samsung.com/AboutSAMSUNG/index.htm`
3. Pirhonen, A., S.A. Brewster, and C. Holguin.: Gestural and Audio Metaphors as a Means of Control for Mobile Devices. ACM CHI'02 Minneapolis (2002)
4. K. Hinckley, J. Pierce, M. Sinclair, E. Horvitz: Sensing Techniques for Mobile Interaction. ACM WIST (2000).
5. Harrison, B., et al.: Squeeze Me, Hold Me, Tilt Me! An Exploration of Manipulative User Interfaces. CHI'98 (1998)
6. S.Strachan, R.Murray-Smith, I.Oakley, J.Angesleva: Dynamic Primitives for Gestural Interaction. Mobile HCI (2004)
7. Kobayashi T., Sugiyama K.: Hand Image Recognition for Code Numbers. ICITA2002 (2002)
8. Patel, S.N., Pierce, J. and Abowd, G.D.: A Gesture-based Authentication Scheme for Untrusted Public Terminals. Proc. of UIST 2004 (2004)
9. `http://www.cnse.caltech.edu/Research01/optics.shtml`
10. D. Gupta: Computer Gesture Recognition: Using the Constellation Method. Caltech Undergraduate Research Journal (2001)
11. R. Collins, R. Gross, J.o Shi: Silhouette-based Human Identification from Body Shape and Gait. 5th Intl Conference on Automatic Face and Gesture Recognition (2002)
12. M.P. Murray:Gait as a total pattern of movement. American journal of Physical medicine Vol.46 (1967) 290-333

13. G. Johansson: Visual perception of biological motion and a model for its analysism. Perception and Psychophysics (1973)
14. J.K. Perng, B. Fisher, S. Hollar, K. S. J. Pister: Acceleration Sensing Glove (ASG). IEEE Symposium on Wearable Computers (1999)178–180
15. Strachan, S. Murray-Smith,R.:Muscle Tremor as an Input Mechanism. Proc. of UIST 2004 (2004) Santa Fe
16. http://www.biometricsinfo.org/biometrics.htm
17. J.B. Tenenbaum, V. de Silva, and J.C. Langford: A global geometric framework for nonlinear dimensionality reduction. Science vol. 290 (200) 2319-2323
18. I.T. Jolliffe: Principal Component Analysis. Springer-Verlag New York (1986)
19. http://www.cs.toronto.edu/r̃oweis/lle/
20. L. Devroye, L. GyorfiA: Probabilistic Theory of Pattern Recognition. Springer (1996)
21. S.T. Roweis and L.K. Saul.: Nonlinear Dimensionality Reduction by Locally Linear Embedding. Science. Vol. 290 (2000) 2323-2326
22. L.Saul and S.Roweis. Think Globally: Fit Locally: Unsupervised Learning of Nonlinear Manifolds. Tech. Report MS CIS-02-18. University of Pennsylvania (2002)
23. L.K. Saul and S.T. Roweis. Think Globally: Fit Locally: Unsupervised Learning of Low Dimensional Manifolds. J. Mach. Learn. Res. Vol. 4 (2003) 119-155
24. Hughes, S. Oakley, I and O'Modhrain, S.: MESH: Supporting Mobile Multi-modal Interfaces. Proc. of ACM UIST'04 (2004)
25. V. Boultreau, N. Vincent, R. Sabourin, H. Emptoz: Handwriting and Signature : One or Two Personality Identifiers?. International Conference on Pattern Recognition (ICPR'98)(1998)63-84.
26. W.Bang, W. Chang, K. Kang, E.C. Potanin, A.D. Kim: Self-contained spatial input device for wearable computers. Proc. of Seventh IEEE International Symposium on Wearable Computers (2003) 26- 34
27. S. Pankanti, R. Bolle and A.K. Jain: Special issue of IEEE Computer on Biometrics. Vol.33, I. 2 (2003)