

Human-Humanoid Walking Gait Recognition

Verena V. Hafner ^a, Ferry Bachmann ^b

Humboldt-Universität zu Berlin, Institut für Informatik, Berlin, Germany

^ahafner@informatik.hu-berlin.de, ^bbachmann@informatik.hu-berlin.de

Abstract—This paper addresses a typical Humanoid problem, that of walking gaits. Creating a stable walking gait on a humanoid robot is a problem that is addressed by many roboticists worldwide. We want to focus on a similar issue that is often overlooked: recognition of different walking behaviours in humans and humanoid robots. We introduce a method that is able to recognise persons as well as their current walking behaviour based on accelerometer data. This is an important skill and prerequisite for all human-Humanoid imitation and interaction tasks.

I. INTRODUCTION

Humanoid robots have been fascinating us for many decades or even centuries. Popular examples of Humanoid robots in the past in film or fiction are Frankenstein’s monster, Rabbi Loew’s Golem, or the robots in ‘I, Robot’. Examples for real Humanoids of the last years are Honda’s Asimo¹ [13], Sony’s QRIO, and recently Aldebaran’s Nao. The particular interest in Humanoids is based on their similarity in looks and movement to humans. Similar morphology also implies similar types of sensors and actuators, though never exactly the same.

The main feature concerning motion and movement that distinguishes humans from other animals and Humanoids from other robots is the ability to walk upright. Walking upright gives a huge evolutionary advantage since it leaves the hands free for grasping, throwing, and tool use. From a developmental perspective, crawling and standing up are the prerequisites for this skill.

We therefore focus on the recognition of human walking gaits in this paper. We show a method that can automatically recognise the person and her specific walking gait from acceleration sensor data.

II. MOTIVATION

A. Aim of this Study

The future aim of this study is to provide a behaviour recognition method for Human-Humanoid interaction. A humanoid robot should be able to recognise and imitate the behaviour of humans and of fellow humanoid robots. The first step to reach this goal is to provide the ability to analyse its own behaviour and draw the connection between executed and recognised behaviour.

B. Decision on Sensors

There are different requirements for sensors depending on the scientific approach and application. For self-imitation and analysis and recognition of own behaviour, acceleration

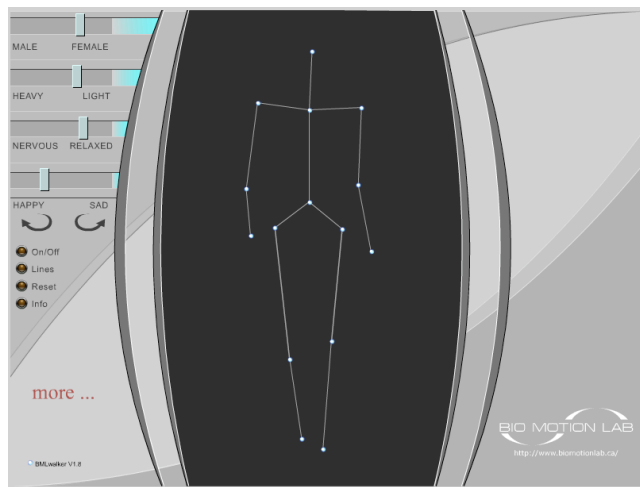


Fig. 1. Point-light walker animation demo from the Biomotionlab [16]. The figure has 15 marker positions at the main joints.

sensors at the joints are a sensible decision. Using raw proprioceptive data such as the set motor values of the joints would also be a possibility, however some of the dynamics inherent in acceleration data would be lost. Proprioceptive data in humans also contains dynamic information rather than static one due to the muscular properties [6].

Humans use vision as their main sensor for recognising the behaviour of other humans. Mapping visually observed movements of others to perceived own behaviour is an unsolved problem, called the body correspondence problem [1].

In the case of humanoid robots, either a solution to the correspondence problem has to be found, for example by simplifying the problem using markers and extracting acceleration data from visual data, or we could assume that all humans and humanoids are equipped with acceleration sensors providing their data to self and others. In this paper, we assume the second case and focus on behaviour recognition from acceleration sensor data.

III. RELATED WORK

A. Point-light Walkers

Experiments on point-light walkers [11], [16] have shown several interesting aspects of human behaviour recognition abilities. Not only could a specific person be recognised by a small set of moving point-light markers representing the joints (see figure 1), but also general features of a person such as weight, sex, and even mood could be detected. Experiments

¹Honda, History of Humanoids, <http://world.honda.com/ASIMO/history/>

on point-light-walkers by Casile and Giese [3] showed that the features are likely to be determined from optic flow and precise position information is not necessary.

B. Developmental and Evolutionary Background

Recognising people by their gait is an important skill that probably evolved very early in human history. It is important to distinguish friend and enemy, and for children to recognise their parents from a distance. Simion et al. [15] recently showed that even babies as young as two days clearly prefer biological motion over random motion or inverted biological motion. Before, it was believed that this preference only appears after about three months.

Humans are also very sensitive to motion cues, originating from the ability to quickly detect predators or prey. Human motion contains information about actions, intentions and emotions. We have evolved to quickly extract socially important features from observing human motion. Psychophysical experiments showed that only two gait cycles are needed to reliably recognise gender from motion without any shape cues [16].

C. Walking gaits in animals

A gait is a particular way or manner of moving on foot. Common human gaits are walk, run and crawl. Common horse gaits are walk, trot, canter and gallop. For each gait, there are several different variants. The gaits mentioned above are natural gaits, other more artificial or trained ways of movement could also be called a gait, e.g. a side-step. In medicine, observation of the gait can provide clues to a number of diagnoses including Parkinson disease and stroke.

The first scientific article on animal walking gaits has been written 350BC by Aristotle [2]. He observed and described different walking gaits of bipeds and quadrupeds and analyses why all animals have an even number of legs.

D. Robotic Walking Gaits

A biologically plausible way to generate robotic walking gaits are central pattern generators (CPG) [4]. Many robotics researchers have examined how to create stable biped, quadruped or polyped gaits in robots. Recent examples are looking closely at the interface between physiology and mechanics [9]. These studies have also shown, how using compliance and passive dynamics add to the understanding of biological walking gaits and reproduce them in a stable and energy-efficient way in robots.

IV. EXPERIMENTS AND RESULTS

Here we present the experimental setup used, explain our choice of hardware, and present the feature extraction and classification methods.

A. Experimental Setup

Nine 3-axis accelerometers were attached to each of four different persons performing five different walking patterns: walking, walking backwards, running, going up- and down-stairs. Data has been recorded for about 10 seconds four times

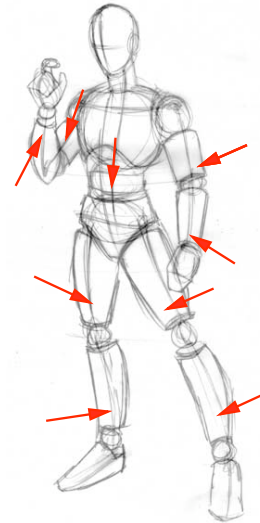


Fig. 2. Positions where the acceleration sensors are attached to the body (sketch taken from polykarbon.com).

for each person and gait with a rate of 100Hz (see figure 3), and three seconds of data have been extracted from the middle of the movement for further analysis. The positions where the nine sensors were attached to the persons are: lower and upper legs, lower and upper arms, lower back (see figure 2).

B. Choice of Hardware

As three-axis accelerometer sensors we used Nintendo Wiimotes. They are ideal for rapid prototyping since they have a wireless connection to a host computer using Bluetooth, and open source code for accessing the sensor data exists. They are also equipped with an infrared camera, a vibration sensor, and a loudspeaker. These are not used for the experiments presented here. An additional advantage of the Wiimote is its small price and its wide spread. A large number of experimental data sets could be easily collected by addressing the Wii community.

C. Feature Extraction Methods

We will present here two feature extraction methods. The first one is based on information distances, the second one is based on energy and applied in several variants.

For some of the experiments, we used the three axis for each acceleration sensor separately (27 in total), for some we transformed them to one sensor value each (9 in total).

In order to convert the x,y,z accelerometer information into one scalar each, the data was transformed by subtracting its mean, and the Euclidean distance X_t was calculated as the new compound sensor value:

$$X_t = \sqrt{x_t^2 + y_t^2 + z_t^2}$$

Subtracting the mean is necessary since the accelerometers also measure the earth gravity in addition to acceleration information by movement and are therefore not centered around zero.

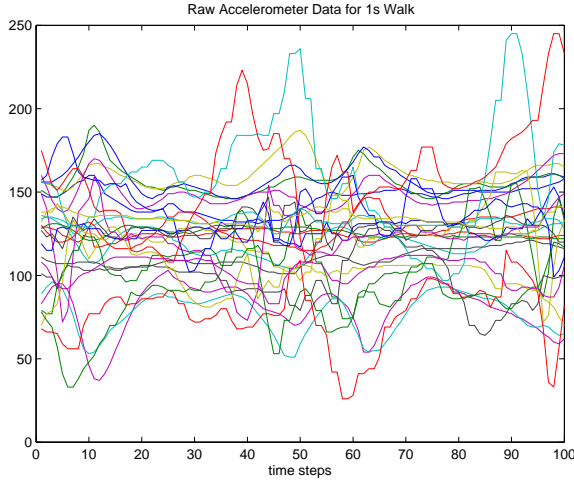


Fig. 3. Three-axis accelerometer data from nine Wiimotes during a walking behaviour of one second.

The classification results are mentioned in this section in order to explain some of the design decisions. These are based on an MLP method explained in more detail in the following section.

1) *Information Distance Method*: This method is based on information theory.

The information distance d between two sensory data streams X_i and X_j introduced by Crutchfield [5] is defined by the normalised sum of the conditional entropies:

$$d(X_j, X_i) = \frac{H(X_i|X_j) + H(X_j|X_i)}{H(X_i, X_j)}$$

A detailed description of the method as applied to sensor streams on robots can be found in Kaplan and Hafner [12]. The parameters used in our experiments are windows of 3 seconds, and 10 bins with adaptive binning and using the 9 compound acceleration sensors.

The data from nine acceleration sensors results in 36 features since the information distance for sensor pairs is symmetrical and zero for the same sensor.

The recognition rates for behaviours were 80% (random: 20%) and 61% for persons (random: 25%).

2) *Energy-based Method*: A more straightforward method which is not based on the relation of the sensor signals but on each sensor signal is to look at the energy of each acceleration sensor. If X is the signal, then $E = \sum_I X_i^2$ is the energy over a certain time period I . Ravi et al. [14] got surprisingly good results with using simple features - one of them being the energy - with just one accelerometer. Using only one accelerometer has an interesting application: It could be used with a cell phone [10]. We first applied this method to our data with windows of three seconds and all 27 sensors.

The recognition rates for behaviours were 98.75% (random: 20%) and 95.5% for persons (random: 25%). These

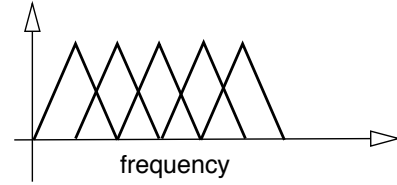


Fig. 4. Filterbank used for the experiments

are surprisingly high recognition rates, in particular for the persons. We assume that the fact that the Wiimotes can be slightly rotated depending on the volunteering person, the recognition is more based on the actual fixation than on specific motion features of the person.

We therefore decided to perform the same feature extraction method again, this time with the compound signal with nine sensor streams as explained above. This makes the single sensors rotation-invariant.

This time, the recognition rates for behaviours were 95% (random: 20%) and 61.25% for persons (random: 25%). The recognition rate for persons is surprisingly low compared to the previous experiment. Probably, many person-specific features contain information that is directional. The recognition rates here are about the same as the recognition rates from the information distance method.

The next method - while still using energy features - is inspired from methods in speech recognition. It is called Filterbank Analysis [8]. This method is specifically tuned to interesting and meaningful frequency parts. Several overlapping filters are applied to the frequency space of the signal. Each filter extracts a certain part of the frequency space for which the energy can be calculated as a scalar feature.

We first used the Fourier-transformed signal of three seconds and applied five linear triangular filters on it which were all half-overlapping (see figure 4). The filters ranged from 0 to 24Hz.

This time, the recognition rates for behaviours were 95% (random: 20%) and 66.25% for persons (random: 25%). These are already better results than with just looking at the energy.

This could still be improved. Since we are only looking at the amplitudes and not at the phases, a Fourier transform might not be ideal. The final experiment for the feature extraction therefore uses a discrete cosine transform (DCT) which is better suited since it only works with the amplitude values.

The recognition rates here were for behaviours 96.25% (random: 20%) and 83.75% for persons (random: 25%).

TABLE I

RECOGNITION RATES FOR PERSONS AND BEHAVIOURS USING DIFFERENT FEATURE EXTRACTION METHODS AND AN MLP.

Method	rec. rate behaviours	rec. rate persons
Information distance 9	80	61
Energy 27	98.75	95.5
Energy 9	95	61.25
Energy 9 + dft filterbanks	95	66.25
Energy 9 + dct filterbanks	96.25	83.75

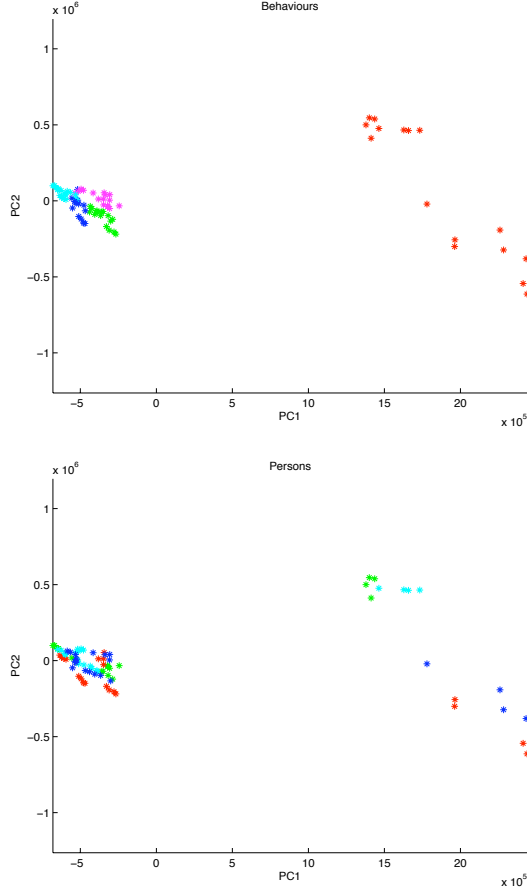


Fig. 5. Energy 27 experiments. Data projected onto the first two principal components (covering 97% of the variance). a) behaviours marked (red: run, green: walk, cyan: backwards walk, blue: going upstairs, magenta: going downstairs), b) persons marked.

D. Categorisation Learning

For the categorisation, we used the data analysis tool WEKA [7] with a Multi-Layer Perceptron (MLP) and backpropagation. The MLP was three-layered with half the number of input- plus output nodes in the hidden layer. We used 10-fold-cross-validation for all the experiments. Surely, the recognition rates could be slightly improved by using different classification methods, e.g. support vector machines, but this was not the aim of our study.

For visualisation purposes, we also had a look at the feature

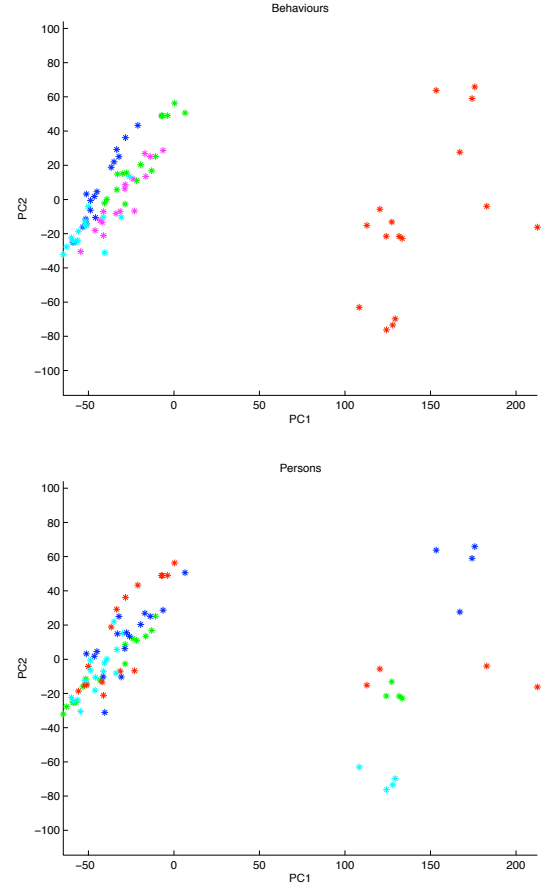


Fig. 6. Energy 9 + DCT experiments. Data projected onto the first two principal components (covering 80% of the variance). a) behaviours marked (red: run, green: walk, cyan: backwards walk, blue: going upstairs, magenta: going downstairs), b) persons marked.

data projected onto the first two principal components. The feature data used for the PCA plots is that of the energy with 27 features and that of the experiment with filterbanks and discrete cosine transform. The first two principal components covered 97% respectively 80% of the variance. In figure 5, all experimental runs are plotted in that way for the energy data with 27 features. As can be seen on top, the different behaviours can be easily separated. The running gait differed the most from the other gaits and also differed most between persons. The reason is probably that it contains more energy than the other behaviours. The DCT filterbank experiment led to similar PCA plots (see figure 6), however covering only 80% of the variance in the first two principal components, and the person data is not as easily separable. The different persons are also visibly separated, but mainly within the running gait.

V. DISCUSSION

We showed an experimental setup for behaviour recognition based on accelerometer data, which is motivated from a behavioural and robotics point of view. Different methods for feature extraction from these data were presented and



Fig. 7. Nao robot from Aldebaran.

compared (see table I for a summary of the recognition rates). The most promising method, energy analysis based on discrete cosine transformed linear filter banks is a modified method from speech recognition that has been applied to acceleration data for the first time in our paper. High recognition rates and fast calculation are the main advantages. The information distance method has the advantage that only correlations between sensors are considered, and not single sensors, but this method is time-consuming. The recognition rates with all methods are higher for the recognition of a behaviour than for the recognition of a person. This is an intuitive result, and persons can be more easily recognised once the behaviour is determined. All the experiments were based on only three seconds of data. The recognition rates will be higher the more data is available.

VI. OUTLOOK

These encouraging first results for recognising human walking behaviour based on acceleration sensor data can now be applied to interesting Human-Humanoid interaction problems. We are currently performing first movement tests with our two humanoid Nao robots from Aldebaran (see figure 7). The goal of these experiments is that Nao would be able to recognise and imitate the walking behaviour of itself and other humanoid robots as well as humans.

Concerning the behaviour recognition based on features from the acceleration sensors, we will collect more data to be able to better generalise. The data collection will be possibly done by distributing our source code to the Wii community as mentioned in the beginning. Future experiments will explore whether it is possible to recognise gender and other features as in the point-light walker experiments.

ACKNOWLEDGMENT

Thanks go to Sawa Fuke from Osaka University for her work on the Darwiin remote driver, and to the Nao team Humboldt who already tested out walking behaviours on the Nao robot for the Robocup. We also thank the volunteers for the walking experiments.

REFERENCES

- [1] Aris Alissandrakis, Chrystopher L. Nehaniv and Kerstin Dautenhahn (2007), Solving the Correspondence Problem in Robotic Imitation across Embodiments: Synchrony, Perception, and Culture in Artifacts. In C. L. Nehaniv and K. Dautenhahn (Eds.), *Imitation and Social Learning in Robots, Humans and Animals: Behavioural, Social and Communicative Dimensions*, Cambridge University Press, pp. 249-273
- [2] Aristotle (350BC), *On the Gait of Animals*, Translated by A. S. L. Farquharson 2007
<http://etext.library.adelaide.edu.au/a/aristotle/gait/>
- [3] Casile, A., Giese, M. A. (2005), Critical features for the recognition of biological motion. *Journal of Vision*, Vol. 5, pp. 348-360
- [4] Cruse, H. (2002). The functional sense of "central oscillations" in walking. *Biol. Cybernetics* 86, pp. 271-280
- [5] Crutchfield, J.P. (1990), Information and its metric. In Lam, L., Morris, H.C., eds.: *Nonlinear Structures in Physical Systems - Pattern Formation, Chaos, and Waves*. Springer Verlag, pp. 119-130
- [6] Fitzpatrick, R. and McCloskey, D. I. (1994), Proprioceptive, visual and vestibular thresholds for the perception of sway during standing in humans. *J Physiol.*, 478 (Pt 1): pp. 173-186
- [7] Frank, E., Hall, M.A., Holmes, G., Kirkby, R., Pfahringer, B., Witten, I.H. and Trigg, L. (2005), Weka - a machine learning workbench for data mining. In Oded Maimon and Lior Rokach, editors, *The Data Mining and Knowledge Discovery Handbook*, pp. 1305-1314. Springer
- [8] Hao Huang; Jie Zhu (2006). Minimum Phoneme Error based Filter Bank Analysis for Speech Recognition. 2006 IEEE International Conference on Multimedia and Expo, pp. 1081-1084
- [9] Iida, F., Rummel, J., and Seyfarth, A. (2008). Bipedal walking and running with spring-like biarticular muscles, *Journal of Biomechanics*, Vol. 41, pp. 656-667
- [10] Iso, T. and Yamazaki, K. (2006), Gait analyzer based on a cell phone with a single three-axis accelerometer. *Mobile HCI 2006*, pp. 141-144
- [11] Johansson, G. (1977), Studies on visual perception of locomotion, *Perception* 6(4), pp. 365-376
- [12] Kaplan, F. and Hafner, V. V. (2006), Information-theoretic framework for unsupervised activity classification, *Advanced Robotics*, 20:10, pp. 1087-1103
- [13] McNicol, T. (2004), These bots were made for walking: Honda's bold new bots are headed into your home, *Japan Inc.* 52 (Feb 2004), p. 10
- [14] Ravi, N., Dandekar, N., Mysore, P., and Littman, M. L. (2005). Activity recognition from accelerometer data. In *Proceedings of the Seventeenth Conference on Innovative Applications of Artificial Intelligence (IAAI)*, pp. 1541-1546. AAAI Press
- [15] Simion, F., Regolin, L., Bulf, H. (2008), A predisposition for biological motion in the newborn baby. *Proceedings of the National Academy of Sciences PNAS*.2008; 105: pp. 809-813
- [16] Troje, N. F. (2002), Decomposing biological motion: A framework for analysis and synthesis of human gait patterns, *Journal of Vision*, 2(5):2, pp. 371-387
<http://journalofvision.org/2/5/2/>, doi:10.1167/2.5.2.