UNIVERSITY OF CALGARY

Application of Pattern Recognition Methods in Biomechanics

by

Bjoern Michael Eskofier

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF DOCTOR OF PHILOSOPHY

FACULTY OF KINESIOLOGY

CALGARY, ALBERTA

APRIL, 2010

Your file Votre référence
ISBN: 978-0-494-62143-1
Our file Notre référence
ISBN: 978-0-494-62143-1

Abstract

Biomechanical studies often attempt to identify differences between groups. Several scientific methods are available for identifying such differences. Traditional methods often focus on the analysis of single variables and do not take into account high-dimensional dependencies. Moreover, the analysis procedures are often biased by the expectations of the researcher. Pattern recognition based methods provide data driven analysis often conducted simultaneously in multiple dimensions. Such algorithms have recently been applied for biomechanical analysis tasks. However, the use of pattern recognition algorithms is still not well understood in the biomechanical community. Therefore, the contribution of this thesis was to add further understanding of tools from pattern recognition to biomechanical tasks of group differentiation.

Two main application scenarios were addressed. In the first part of the thesis, questions of human gait classification were examined. Existing studies with respect to this task had two main shortcomings. First, the features used for classification were often specific to the input measurements, derived from specific time points and thus not directly transferable to different tasks. Second, frequently only information from single variables was analyzed and high-dimensional dependencies neglected. Therefore, techniques for running and walking gait pattern classification were developed that overcame these shortcomings. They employed generic features that used a more complete representation of the available information compared to traditional methods. Moreover, high-dimensional dependencies were accounted for. Several group classification tasks were successfully solved using the developed methodology. The techniques are general and applicable to different group classification tasks without adaptation.

In the second part of the thesis, the implementation of pattern recognition algorithms on embedded systems was considered. Such systems allow, for instance, the application of pattern recognition systems outside the lab for sports biomechanics as well as for many other domains. General considerations for the implementation of pattern recognition algorithms on this specific hardware environment were still missing in the literature. A general methodology for embedded classification was therefore developed. The ability of this approach to produce acceptable results in sports biomechanics related classification tasks was shown. Furthermore, the applicability of embedded solutions for data collection in sports classification studies was demonstrated.

Preface

The chapters three through seven of this thesis are based on the following manuscripts submitted to and/or published in the following peer-reviewed journals.

1. ESKOFIER, B. M., KRAUS, M., WOROBETS, J. T., STEFANYSHYN, D. J. & NIGG, B. M. (submitted). Pattern classification of kinematic and kinetic running data to distinguish gender, shod/barefoot and injury groups with feature ranking. *Computer Methods in Biomechanics and Biomedical Engineering*.
2. ESKOFIER, B. M., FEDEROLF, P., KUGLER, P. & NIGG, B. M. (submitted). Marker-based classification of young-elderly gait pattern differences via direct PCA feature extraction and SVMs. *Journal of Biomechanics*.
3. ESKOFIER, B. M., OLESON, M., DIBENEDETTO, C. & HORNEGGER, J. 2009. Embedded surface classification in digital sports. *Pattern Recognition Letters*, 30, 1448-1456.
4. ESKOFIER, B. M., WAGNER, M., MUNSON, I. & OLESON, M. (submitted). Embedded classification of speed and inclination during running. *International Journal of Computer Science in Sports*.
5. ESKOFIER, B. M., HARTMANN, E., KÜHNER, P., GRIFFIN, J., SCHLARB, H., SCHMITT, M. & HORNEGGER, J. 2008. Real time surveying and monitoring of athletes using mobile phones and GPS. *International Journal of Computer Science in Sports*, 7, 18-27.

The thesis related work has been presented at several conferences (cf. References). This dissertation is based on a collection of manuscripts, and therefore has some redundancy in the introduction and methods sections of some of the manuscripts. Whenever custom made software programs are mentioned in this thesis, they were mainly implemented by the author of this thesis.

Acknowledgements

I would like to express my thanks to the following persons:

Dr. Benno Nigg for taking me as a PhD candidate in the first place and for letting me learn a lot more than just biomechanics.

Dr. Joachim Hornegger for serving on my committee and for all the support and guidance during my PhD studies.

Dr. Vincent von Tscharner for serving on my committee and for allowing me to find out about a bodacious way of scientific thinking. His wife Michaela and him for helping me out on numerous occasions during my time in Calgary.

Dr. Christian Jacob for serving on my candidacy and on my defense committee.

Dr. Arnold Baca for being the external examiner on my defense committee.

Dr. Peter Federolf for serving on my candidacy committee and for his support of my research.

Ursula Heinz for her administrative assistance here in Calgary.

My many fellow HPLers for sharing knowledge and ideas about their projects, and for the good times in and around Calgary on hiking tours and skiing trips.

All the people at adidas, especially Berti, Roland, Heiko, Ute, Mark, Ian and Janet for their time, project ideas, and many good times in- and outside the office. Adidas and the adidas innovation team provided the financial support for my research.

My brothers, Sven and Dominic, two great siblings who help me keep a good perspective on life outside of science.

Anett, for all the proofreading, for keeping me sane, and for many wonderful moments in my life.

My parents, Siegi and Bernd, for their encouragement, support and for all the lessons about what is really important in life. One day, I want to be as good a teacher and as good a creator as you are. This thesis is dedicated to you.

Table of Contents

List of Tables

List of Figures

Epigraph

Alle Tage rauscht die Fülle der Welt an uns vorbei;
alle Tage blühen Blumen, strahlt das Licht, lacht die Freude.
Manchmal trinken wir uns dankbar daran satt,
manchmal sind wir müde und verdrießlich und mögen nichts davon wissen;
immer aber umgibt uns ein Überfluss des Schönen.
Das ist das Herrliche an jeder Freude,
dass sie unverdient kommt und niemals käuflich ist;
sie ist frei und ein Gottesgeschenk für jedermann,
wie der wehende Duft der Lindenblüte.

Hermann Hesse

Imagination is more important than knowledge.

Albert Einstein

CHAPTER 1 INTRODUCTION

1.1 Overview

Biomechanical studies often attempt to identify differences between groups such as male/female, elderly/young or injured/not injured. A commonly applied strategy to describe group differences is to collect data for variables that are assumed to contain relevant information. However, single variables seldom allow group differentiation and high-dimensional dependencies are often difficult to identify using traditional analysis methods. Pattern recognition methods offer themselves as a tool for solving such biomechanical problems. These methods have the potential of identifying groups and of revealing relevant variables in multiple dimensions. Pattern recognition algorithms have recently been applied for isolated biomechanical studies. The use of such data mining algorithms, however, is still not well understood in the biomechanical community. Therefore, the purpose of this thesis was to apply selected pattern recognition methods to biomechanical questions of group differentiation.

1.2 Background on pattern recognition

Pattern recognition is a scientific discipline with the goal to classify objects into a number of categories or classes (Theodoridis and Koutroumbas, 2009). Pattern recognition has a strong applied aspect, with contributions towards many facets of daily life. Early research in pattern recognition had applications in speech and optical character recognition (Mori et al., 1992). Patents on optical character recognition go back to the 1920's (Tauschek, 1929), and today character recognition systems operate in every mail sorting plant. Also the more modern form of mail (namely email) is sorted into 'ham' and 'spam' (Koprinska et al., 2007) using pattern recognition methodology. Further applications of pattern recognition applications include speech recognition and face and object detection (Furui, 2004; Pontil and Verri, 1998; Viola and Jones, 2004). Despite the fact that the analysis methods from pattern recognition have such long tradition, they have only recently started to be applied in research concerned with locomotion, sports and, in general, with biomechanics. The value of such methods lies in the fact that they have an unbiased mode of operation and the capability to handle a large quantity of collected data. For example, a system for the classification of human feet has been developed (Grimmer et al., 2009) based on measured foot surfaces. The data set consisted of more than 11000 measured feet and each foot consisted of about 50000 individual points. The developed pattern recognition data mining needed no prior assumptions and could be applied efficiently for automatic group differentiation despite this high amount of data.

The classical pattern recognition approach (e.g. Duda et al., 2001; Niemann, 2003) uses data from arbitrary sensors as input into a classification system (Fig. 1.1). For biomechanical studies, these data represent typical measurements as they are collected in biomechanical studies, e.g. kinematic and kinetic data from a motion analysis system (Nigg et al., 2007). The measurements are subjected (if needed) to a preprocessing step in order to enhance the signal properties. The subsequent feature extraction provides a feature vector for subsequent classification. This vector describes the input measurements in feature space. For supervised classification, labeled feature vectors are presented to a classifier for training (Fig. 1.1). The vectors used for classifier training constitute the training set. The mentioned labels assign a feature vector to one of several possible classes. One example for a class label is the gender of a person that is measured, with two possible classes in this case. The classifier then 'learns' the class specific properties and establishes a decision rule. In the working phase, the trained classifier uses this decision rule and automatically assigns a feature vector to a class. Different classifiers using diverse strategies for learning can be used, including for example, Support Vector Machines (Vapnik, 1998) and Linear Discriminant Analysis classifiers (Fisher, 1936).

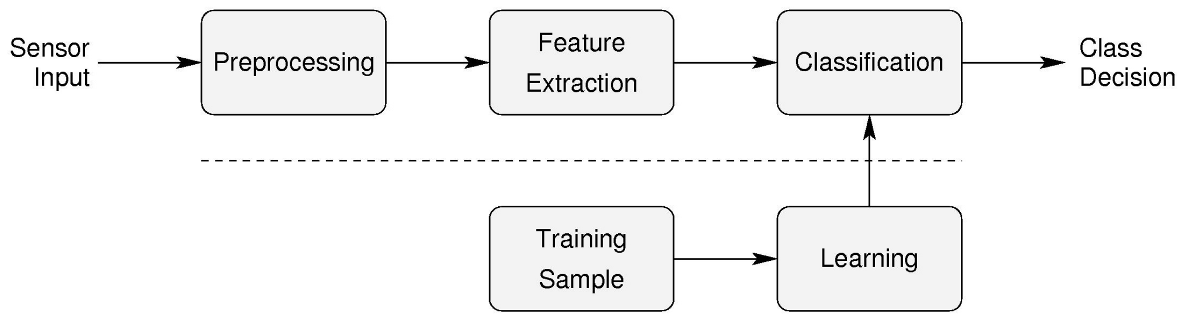

Fig. 1.1. Flow chart of a pattern recognition system (Niemann, 2003).

However, there are few theoretical guidelines to determine what preprocessing, which features and which classifier are required to optimally solve a given classification problem (Duda et al., 2001). Research is therefore required whenever pattern recognition is applied to new problem domains, for example to biomechanics.

1.3 Applying pattern recognition in biomechanics

Today's methods for quantifying variables of interest in locomotion studies are rather sophisticated. For example, the moments and forces that the human body is exposed to during a specific movement task can be estimated using motion analysis data combined with an inverse dynamics approach (Nigg et al., 2007). Furthermore, movement characteristics can be quantified using mobile, computationally powerful embedded systems. These ‘ubiquitous computing’ systems can provide monitoring, direct feedback and interventions by interacting with equipment (Baca et al., 2009).

In scientific studies related to such questions, a considerable amount of data is generated. This demands for automatic, computerized analysis methods that have the ability to deal with large data sets and that use algorithms that are powerful and effective. The objective, data-driven methods from pattern recognition can contribute to these projects by offering useful tools for such analysis procedures.

In this thesis, pattern recognition methods have been applied in two different application domains. The first application was group classification in biomechanics and the second was classification of data on embedded systems.

1.3.1 Group classification in biomechanics

Biomechanical research questions are often related to the quantification of specific characteristics of groups. Various studies have been conducted to distinguish between groups with respect to gender (von Tscharner and Goepfert, 2003), age (DeVita and Hortobagyi, 2000), footwear (Nigg et al., 2003) and identification of factors important for the development of injuries (DiBenedetto et al., 2005; Messier and Pittala, 1988; Stefanyshyn et al., 2006; Taunton et al., 2002). These studies quantified kinematics, kinetics, soft tissue vibrations and/or electromyography (EMG). The classification was often done by comparing mean and standard deviation of discrete variables (e.g. impact peak force, maximal foot eversion, EMG root mean square, etc.). Recently, new methods have been proposed and successfully applied for such classifications based on pattern recognition (e.g. Begg and Kamruzzaman, 2005; Janssen et al., 2008; Schöllhorn et al., 2002; Wu et al., 2007). However, several research questions were left unanswered by these published studies.

1.3.1.1 Research question one

A shortcoming of many of the recently proposed methods was that the features used for classification were specific to the input measurements. One example publication used angles and forces at specific time points of the human gait phase (Wu et al., 2007). This approach is characterized by the fact that information was lost when the ‘analog’ locomotion cycle of walking or running was reduced to ‘discrete’ time points. The classification rate in every classification study depends on the information content of the features that are used (Niemann, 2003). Feature sets for biomechanical locomotion data were therefore missing that represented as much information as possible, that needed no additional information about specific time points and that could be calculated on the ‘analog’ time-dependent kinematic and kinetic data directly. Such so-called ‘generic’ features were developed for this thesis and the first research question addressed by this thesis was:

“Is acceptable group classification for gender and shod vs. barefoot possible using generic features for biomechanical data?”

A classification rate of 80% was arbitrarily set for testing acceptable group classification. This value was deemed to demonstrate a good classifyability of the underlying data. Gender and footwear (shod/barefoot) classes were selected for investigation because they are present in most biomechanical studies. The two hypotheses tested with this research question were:

<u>*Hypothesis H1*</u>

“Using generic features for biomechanical data, a class-wise mean classification rate of at least 80% is possible for gender classification.”

<u>*Hypothesis H2*</u>

“Using generic features for biomechanical data, a class-wise mean classification rate of at least 80% is possible for shod versus barefoot running classification.”

By showing that both hypotheses H1 and H2 are fulfilled, supporting evidence to the general usability of such a classification methodology for biomechanical group differentiation tasks was provided.

Research question one was presented in two publications (Eskofier et al., 2010a; Eskofier et al., 2010b) that form Chapter 3 and Chapter 4 of this thesis.

1.3.1.2 Research question two

A further disadvantage of published classification methods for biomechanical data was that information about what exactly characterized the differences between the groups identified was often not provided. For example, it has been shown that several emotional states (e.g. ‘angered’, ‘normal’) can be classified using gait data (Janssen et al., 2008). However, the publication was not able to characterize the differences within the original gait data with the chosen approach. A methodology that is ideally suited for biomechanical group classification tasks does not only obtain high recognition rates, but also identifies the differences in the data that allowed the classification. Such feature selection and ranking strategies for biomechanical data were investigated and research question two in this context was:

“Are feature selection and ranking methods for biomechanical data capable of pointing out discriminating characteristics of classes with acceptable classification rate, i.e. can they identify the measured variables containing the discriminating information?”

Research question two was also addressed in the publication (Eskofier et al., 2010b) in Chapter 3.

1.3.1.3 Research question three

A further shortcoming of existing methods for finding differences between groups in biomechanical data is that multi-dimensional dependencies of features are often not used for classification. In the classical analysis, a one-to-one comparison of features for both groups was frequently made (e.g. Stefanyshyn et al., 2006). However, the information necessary for classification is often multi-dimensional (Duda et al., 2001) and a data-driven combination of features for classification is desired to allow better discrimination of the groups under investigation. For that reason, it was needed to be demonstrated that combinations of features are beneficial for discrimination of the studied groups. Consequently, research question three was:

"Are pattern recognition methods able to show high-dimensional dependencies of classes on features that have previously not been revealed?"

The hypothesis that was deduced from this research question directly tested its validity.

Hypothesis H3

"Using pattern recognition methods, a better class-wise mean classification rate is obtainable by combination of multiple features compared to using selected individual features for classification."

Research question three was addressed in the publications (Eskofier et al., 2010a; Eskofier et al., 2010b) that form Chapter 3 and Chapter 4 of this thesis.

1.3.2 Classification on embedded systems

The second research task of this thesis was the application of classification methods to embedded systems. Embedded systems are lightweight and highly mobile systems, capable of measuring, storing and processing data. They are ideally suited for classification tasks in sports because they allow taking measurement and decision-making outside the laboratory. As a result, a growing number of embedded systems are applied in a variety of application scenarios in sports. Examples include systems for physiological measurement and processing tasks (Laukkanen and Virtanen, 1998) and embedded devices that sense locomotion information and utilize it to adapt sports gear (DiBenedetto et al., 2005).

Despite their advantages for classification of sports related data, embedded systems have not been in the focus of biomechanical research so far. Therefore, several basic points needed to be addressed in order to first obtain a generally applicable framework for embedded classification. This general framework was then applied to classification research tasks in sports biomechanics. To do this, several research questions needed answering.

1.3.2.1 Research question four

Embedded systems have restricted memory and processing capabilities. At the start of this project, it was not clear which of the often computationally demanding pattern recognition algorithms (Duda et al., 2001; Fukunaga, 1990; Niemann, 2003) were suitable for implementation on embedded systems. A selection and validation of appropriate algorithms was, therefore, important. There existed, however, neither a theoretical or practical result as to what method of selection is appropriate, nor how a validation in the context of embedded systems should take place. Therefore, a better understanding of all the steps necessary to adapt pattern recognition algorithms to embedded devices was needed. A general methodology for development of classification systems for embedded microprocessors was thus missing that took into account all important steps in the pattern recognition process (Fig. 1.1). For this, the types of features that were computable on embedded hardware needed to be discussed. Furthermore, a selection of classification algorithms for this hardware needed to be made. Thus, research question four was:

“Are the employed types of features and classification algorithms suited for obtaining acceptable classification rates on an embedded microprocessor?”

Two hypotheses were tested for this research question. A classification rate of 80% was set for testing them for the reason already explained in Section 1.3.1. Surface and speed classes were selected for investigation because their recognition was an important task in a sports example system (cf. Chapter 5).

Hypothesis H4

“Using the developed methods for embedded recognition, an on-system class-wise mean classification rate of at least 80% is possible for surface classification (2 classes).”

Hypothesis H5

“Using the developed methods for embedded recognition, an on-system class-wise mean classification rate of at least 80% is possible for speed classification (3 classes).”

By showing that both hypotheses H4 and H5 were fulfilled, supporting evidence to the general applicability of the developed embedded classification framework was collected.

Research question four was addressed in two publications (Eskofier et al., 2009a; Eskofier et al., 2010c), which are Chapter 5 and Chapter 6 of this thesis.

1.3.2.2 Research question five

A key point of the development of every pattern recognition method is the validation of the classification rate on the target system (Duda et al., 2001). Normally, extensive testing of the proposed methodology on a desktop computer is sufficient for this purpose. However, desktop computers and embedded microprocessors can exhibit important differences regarding their hardware capabilities. For example, floating point multiplication might not be available on the microprocessors and integer arithmetic must be used instead. It was therefore not clear whether the classification results on the desktop computers that were used for testing and those results on the embedded system were identical. Research question five was thus:

“Can the developed algorithms for microprocessors effectively be tested on the embedded hardware?”

The hypothesis that was tested for this research question directly tested its validity.

Hypothesis H6

“Testing the class-wise mean classification rate on the embedded device leads to the same result that was achieved during testing of the classifier on a desktop machine.”

Proving this hypothesis was necessary in order to obtain a classification framework for embedded systems that was not only accurate in theory, but also in practice.

Research question five was also addressed in the publication (Eskofier et al., 2009a) in Chapter 5.

1.3.2.3 Research question six

Another aspect where embedded systems can play an important role in sports classification studies is data collection. In pattern recognition, a maximum of data about the classification problem is desired to properly learn the class conditional statistical properties (Duda et al., 2001; Niemann, 2003). Therefore, the development of such a system for data collection that is general and applicable to a wide range of different studies was desirable in order to generate important input for the classification. Such a system needed to be capable of collecting high-quality data while still being unobtrusive, i.e. not hindering the athlete. Therefore, an embedded system implementation was developed for the purpose of athlete monitoring and data collection and the research question in this context was:

“Is this developed system capable of collecting high-quality data while still being unobtrusive, i.e. not hindering the athlete during their sports activity?”

Research question six was addressed in the publication (Eskofier et al., 2008a) in Chapter 7.

1.4 Purposes of the thesis

The first purpose of this thesis was to show the applicability of pattern recognition methods to biomechanical data, to define a set of generic features and to demonstrate that shortcomings depending on the classical approach can be overcome.

The second purpose of this thesis was to develop a general methodology for embedded classification and to demonstrate its capability to produce acceptable results in sports biomechanics related classification tasks.

CHAPTER 2 LITERATURE REVIEW

This chapter provides an overview of the relevant literature for this thesis. Background knowledge dealing with pattern recognition in general is discussed in the first section. Publications that have used pattern recognition methods together with biomechanical data are discussed in the second section. Publications that deal with classification on embedded systems are discussed in the third section.

2.1 Pattern recognition in general

Pattern recognition is a relatively young science and will play an important role in this thesis. Therefore, a review of the most important literature of pattern recognition will be presented. This review is based on the schematic flow chart presented earlier (Fig. 1.1). Therefore, publications addressing relevant aspects of sensing, preprocessing, feature extraction and classification are included.

2.1.1 Sensor input

Pattern recognition is a scientific discipline with the goal to facilitate decision making by machines (Theodoridis and Koutroumbas, 2009). In order to make a decision in any context, it is required that information is obtained. For automatic pattern recognition systems, this information is often provided by sensors (Duda et al., 2001). These sensors measure physical quantities and provide an analog or digital characteristic of the objects that have to be classified (Niemann, 2003). An example task in pattern recognition is the recognition of spoken words (e.g. Furui, 1986). For this purpose, the sensor input to the classification system consists of the acoustic signal that is recorded by a microphone and is subsequently digitized.

The design of sensors is not the focus of pattern recognition research (Duda et al., 2001). Sensor design is conducted only in very specific cases, for example when Braille signs (for visually impaired persons) need to be recognized with tactile sensors (Arabshahi and Jiang, 2005). Instead, existing sensor technology is most often employed for a given pattern recognition task. The selection of the appropriate sensor technology is in the hands of the researcher. No theoretical result exists yet that outlines what sensor is most appropriate for a specific pattern recognition problem (Theodoridis and Koutroumbas, 2009).

2.1.2 Preprocessing

After measurement, the sensor data is preprocessed in terms of a signal to signal transformation. The purpose of this step (Niemann, 2003) is to enhance the signal, to alleviate the subsequent analysis (reduction of input signal complexity) or to increase the classification performance (improvement of obtainable classification rate). Preprocessing is an important aspect of the pattern recognition flow chart, but it is not the focus of pattern recognition research. Therefore, only a brief overview of preprocessing is given.

Several examples for preprocessing techniques exist. Digital filters (Shenoi, 2006) can be used to extract specific frequency bands from a measured signal. Time normalization (for instance using Spline interpolation (De Boor, 1978)) is often applied to facilitate the computation of time independent features. Threshold operations can be employed to suppress selected ranges of values (Niemann, 2003). Segmentation of individual objects (e.g. words within an audio signal) is mandatory for pattern recognition systems to identify single occurrences of certain events within the data (Duda et al., 2001, pp. 9-10).

For the parameter selection and evaluation of a specific preprocessing approach, detailed rules are often available. When, for example, digital filters like general lowpass filters are employed, these filters need to dampen a defined frequency range. However, the decision which of the manifold of available preprocessing methods is appropriate for a given problem is in the hands of the researcher. It is not known yet how to choose preprocessing techniques independent of the complete pattern recognition system (Niemann, 2003). This is a direct deduction from the 'No Free Lunch Theorem' (Duda et al., 2001). They basically state that a universally applicable approach does not exist. In order to evaluate a specific implemented preprocessing method in the context of the pattern recognition system, heuristics are therefore often applied. For example, the classification result with and without the selected preprocessing method is evaluated and the selection of the appropriate algorithms is based on this evaluation. Another heuristic that is often employed is to subjectively compare the quality of a signal before and after preprocessing by visual or auditory inspection (Niemann, 2003).

The differentiation between preprocessing and the subsequent feature extraction is often not straightforward. In some publications and textbooks, preprocessing is seen as part of feature extraction and vice versa (Niemann, 2003).

2.1.3 Feature extraction

The goal of feature extraction is to provide a representation of the measured object characteristics that alleviates subsequent decision making. This goal can be achieved by representing objects of the same category by feature values that are similar, and objects of different categories by feature values that are distinct (Duda et al., 2001). The representation of a spoken word, for instance, by the originally measured time series is not optimal in order to distinguish male and female speakers. For such a classification task, the fundamental frequency of the spoken word has frequently been employed (Childers and Wu, 1991).

The generation of appropriate features "is of paramount importance in any pattern recognition task" (Theodoridis and Koutroumbas, 2009, p. 323). One major concern of this thesis is therefore appropriate feature extraction strategies for biomechanical group and embedded data classification tasks. For this purpose, generic as well as manual feature extraction methods can be employed. As an optional step of feature extraction, a reduction of the feature set is often beneficial for classification (Dash and Liu, 1997). Publications related to the generic features, manual features and feature reduction are discussed below.

2.1.3.1 Generic features

Generic (heuristic) features are defined as features that are not adapted to the sensor input. This approach does not rely on specific properties of the signal that are required to be known a priori (Zhang and Rockett, 2009). Such features can thus be applied to measurements even if the characteristics of the input changes, for example by using a different sensor.

The first group of generic features that are employed in the literature applies a transformation into the frequency domain on the sensor input. Examples include the discrete Fourier transform (Proakis and Manolakis, 1992) and related frequency transformation algorithms, e.g. the discrete cosine transform (Ahmed et al., 1974). Features derived from the fast implementation of the discrete Fourier transform, the fast Fourier transform, were used for fingerprint recognition (Willis and Myers, 2001), for instance. The discrete cosine transform was, for example, employed to efficiently retrieve images from a large database (Fan and Wang, 2002). The same discrete cosine transform algorithm was also applied to recognize faces in images (Hafed and Levine, 2001).

These frequency domain approaches for feature generation do not take into account the temporal aspect of a signal. If this aspect should be considered, the Wavelet transformation (Aldroubi and Unser, 1996; Daubechies, 1992) can be applied. This transformation performs a simultaneous decomposition of the signal in time and frequency. The resulting coefficients can directly be used as generic features for classification. The Wavelet feature extraction technique was, for instance, applied to the classification and segmentation of textures of images (Unser, 1995). In a different example study (Mallet et al., 1997), an adaptive implementation of the Wavelet transformation was used for feature extraction.

A different class of generic features for classification is directly computed from the time domain signal without transformation. The arguably simplest of these features are mean and variance of the input measurement (Duda et al., 2001). These features were, for instance, used for the classification of audio signals with respect to genre or content (Wold et al., 1996; Tzanetakis and Cook, 2002). More advanced directly computed features are derived from a polynomial regression of the given signal up to a predefined order. These polynomial coefficient features were, for example, employed for face verification (Sanderson and Paliwal, 2002) and for speech recognition (Furui, 1986).

The Principal Component Analysis, also called Karhunen-Loève transformation (Karhunen, 1947; Loève, 1977), can also be applied for generic feature extraction. This method transforms a multi-dimensional signal according to its directions of highest variance. These directions of highest variance often contain the information that is necessary for classification. Efficient implementations of this technique for large dimensionality and small sample sizes exist (Fukunaga, 1990). Principal Component features were employed in a variety of classification tasks (e.g. Bicciato et al., 2003; Draper et al., 2003; Hubert and Engelen, 2004).

2.1.3.2 Manually designed features

Manually designed (analytic) features are defined as features that represent specific characteristics of the sensor input. For example, individual heart beats measured in electrocardiogram signals exhibit a detailed structure that is characteristic of the electrical activity of the heart (Fig. 2.1). The information contained in this specific heart beat structure can be closely represented by designed features. These could, for instance, be computed by measuring timing between R-peaks or the integral of the T-wave (Fig. 2.1). Such features that are specific to electrocardiogram signals were used for classification of pathologic heart beats for example in (Chazal et al., 2004).

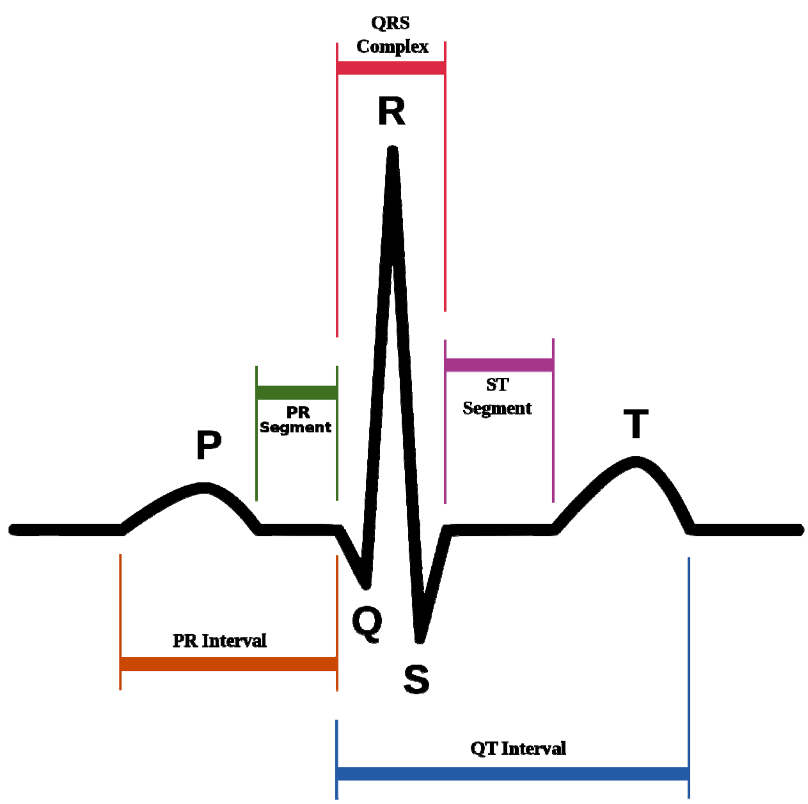

Fig. 2.1. Illustration of an example for manually designed features. Depicted is a typical individual heart beat in an asymptomatic electrocardiogram signal. Also shown are possible design features, for example specific intervals. These features can be used for the purpose of classification. Image from Wikimedia Commons, created by Anthony Atkielski, printed with permission.

One shortcoming of manually designed features is that they can not be directly transferred to other classification tasks. For instance, the straightforward application of the above-mentioned electrocardiogram features to signals with different structure, e.g. audio signals, is not possible. This is in contrast to generic features (cf. Section 2.1.3.1), that can be applied to various signals with only minor modifications. Furthermore, the composition and implementation of manually designed features takes more time than for generic features (Theodoridis and Koutroumbas, 2009). However, the advantage of specifically designed features is that they often represent the information content of the underlying measurements better than generic features (Duda et al., 2001). Therefore, a better classification system performance can frequently be achieved using a manual design approach.

The number of possible approaches for manual feature design is large. This is due to the variety of conceivable input signals to a classification system and the circumstance that manual features are specific for the actual signal they are used for. Therefore, comprehensive reviews are not even conducted in dedicated textbooks (Theodoridis and Koutroumbas, 2009, p. 411). Nevertheless, a selection of additional examples for manual feature design is presented.

Simple numeration features were shown to be applicable for the classification of proteins (Chen and Kurgan, 2007). The features that were used in this study count the occurrences of certain helix, strand and coil structures of the proteins that they classify. In a different study (Varela et al., 2006), benign versus malignant masses in mammographic mass lesions were classified by designing features that represented the degree of sharpness and microlobulation of the mammographic mass margins. The computation of features that measure specific distances and areas was also conducted in classification tasks (DeKruger and Hunt, 1994; Oh and Suen, 1998). Many of these studies that applied manual feature design reported successful classification results. All of them required a profound knowledge about the question studied and the measured signal characteristics.

2.1.3.3 Feature reduction

Feature reduction is an optional step following feature computation. In this step, the dimensionality of the feature representation is reduced. It is, for example, a common approach to create a large set of candidate features where individual features contain some redundancy (Kudo and Sklansky, 2000). This redundancy is deliberate, with the goal to represent the information in the original signal as completely as possible. Nevertheless, this overrepresentation sometimes degrades the performance of a classification system. It is therefore beneficial for classification to reduce the set of features.

Applying feature reduction in a pattern recognition system has several additional advantages. First, the complexity of the classification system can be reduced (Niemann, 2003). Second, the generalization properties of the classification system (i.e. the ability to classify samples that are not in the set that is used for training) can be improved by feature reduction (Theodoridis and Koutroumbas, 2009). Third, classification is based on learning category properties from a representative sample. The number of required samples grows exponentially with the dimensionality of the feature representation (Duda et al., 2001). This is also referred to as 'curse of dimensionality' in the literature (Theodoridis and Koutroumbas, 2009). One goal of feature reduction is therefore to obtain a feature representation with lower dimension compared to the original feature space.

Feature reduction can be performed with two different approaches. The first is feature reduction by selecting a subset of the best individual feature values from a set of candidate features. The second is direct dimensionality reduction within feature space, for example by linear combination of features.

Feature reduction by subset selection

Several feature reduction methods exist for selecting a subset of features from a larger candidate set. These feature reduction methods are called feature selection.

The criterion for the selection of a new set of features is most often the resulting classification rate.

One possible subset selection algorithm is ‘exhaustive search’. It tests all possible combinations of features and guarantees finding the best subset with respect to the evaluated criterion. Exhaustive search may be computationally prohibitive when the dimensionality of the feature representation is high (Narendra and Fukunaga, 1977).

A computationally more efficient strategy is used by the ‘beam search algorithm’ (Bisiani, 1987). This algorithm starts with evaluating all combinations of a given number of features. All these combinations are evaluated with the given performance criterion, and only those that perform best are promoted to the next iteration. In this subsequent iteration, the surviving subsets are combined with the remaining features, and are again evaluated.

An additional approach is feature reduction by ‘dynamic programming’ (Niemann, 2003). This algorithm searches through the feature space in multiple iterations. In each iteration, one single feature is added to the feature subset that gives the highest improvement in classification rate for the worst class pair. For this algorithm, a distance criterion has to be utilized that is monotone and separable. For this purpose, the Mahalanobis distance criterion (Mahalanobis, 1936), a type of generalized Euclidean distance, is frequently employed.

A further method is selection by classification algorithms with the inherent ability to identify the most relevant features for a classification task. The AdaBoost classifier (Freund and Schapire, 1997) can be used for this purpose. Feature selection using AdaBoost is for example demonstrated by Viola and Jones (2004). Genetic algorithms with a neural network classifier can also be used for feature selection (Jaremko et al., 2002). A comprehensive review of other feature selection methods can be found in (Dash and Liu, 1997).

Feature reduction by dimensionality reduction

In addition to methods that select feature subsets from a base set, direct dimensionality reduction methods can also be employed to reduce the complexity of the classification task. In principle, these methods can be used directly on the signals as well; however, this subsection will concentrate on features.

The Principal Component Analysis can be applied for this purpose (Duda et al., 2001). The Principal Component Analysis linearly transforms the input space in such a way that each axis is ranked according to the variance that it contributes to the original dataset. By projecting the features on the axes of highest variance, the noise that is inherent in the features is reduced and the classification rate is typically increased (Theodoridis and Koutroumbas, 2009). Furthermore, the resulting feature components are mutually independent and normally distributed. Feature reduction by Principal Component Analysis was for example applied for machine defect classification (Malhi and Gao, 2004).

The Principal Component Analysis can also be performed using a kernel function in order to investigate nonlinear dependencies within the original data (Schölkopf et al., 1998), which has the advantage that such nonlinear dependencies can be incorporated and used for the analysis. This so-called kernel Principal Component Analysis also found applications in pattern recognition (e.g. Stamkopoulos et al., 1998; Wu et al., 2007).

Another approach for dimensionality reduction is the Linear Discriminant Analysis (Duda et al., 2001). In contrast to Principal Component Analysis, the features are not projected onto the axes of highest variance, but on the axes of highest discriminability. These axes are those directions in the multi-dimensional feature space that represent the highest difference between classes. This requires features that are labeled. Linear Discriminant Analysis for feature reduction was for example employed for face recognition in images (Yang et al., 2005).

2.1.4 Classification algorithms

The last aspect of the schematic flow chart presented in Fig. 1.1 is classification. Classification algorithms operate in a training and a working phase (Theodoridis and Koutroumbas, 2009). In the training phase, the specific characteristics of the classes are learned from a representative sample. For this purpose, the objects are labeled with the information to which class they belong and the classification algorithms are provided with this information. Given this information, a decision criterion is established. This criterion can be represented by a decision boundary in the feature space. In the working phase, the decision criterion, or decision boundary, is employed to new sensor input. The newly measured objects are consequently assigned to the class or category that they most likely belong to.

As of yet, there exists no theoretical result, which of the numerous classification algorithms that are published is applicable in a specific pattern recognition task (Duda et al., 2001). An approach that is often followed is to compare different classification algorithms with respect to the classification rate that they achieve (Niemann, 2003). In this thesis, several classification algorithms were employed in such a way. The relevant literature for these classifiers is described in the following.

2.1.4.1 Linear Discriminant Analysis

The Linear Discriminant Analysis classifier is based on Fisher’s (1936) work on discriminant methods. It is a transformation that aims at minimizing the variability within a class, and maximizing the distance between classes. This algorithm can also be employed for feature reduction as was discussed earlier. When Linear Discriminant Analysis is applied for binary classification, the feature space is effectively projected onto a single axis. On this single axis, a threshold is applied for differentiation. The linear decision boundary that results in the feature space can be easily implemented for the working classification system and uses only minor computational resources. More complex decision boundaries can not be represented. Applications of Linear Discriminant Analysis can be found in a variety of fields, including face recognition (Lu et al., 2003) and document classification (Ye and Li, 2005).

2.1.4.2 Boosting

The goal of boosting is to increase the accuracy of a learning algorithm using the idea of combining multiple simple classifiers to form a strong ensemble (Schapire et al., 1998). One of the most important boosting algorithms is AdaBoost (Freund and Schapire, 1996). In the related publication, a specific algorithm for combining the weak classifiers and establishing the strong ensemble classifier is presented. One advantage of AdaBoost is that it generalizes very well to new samples. However, it has also been shown that its performance is affected by noise within the data (Rätsch et al., 2001). Numerous variations of this method exist with varying strengths and weaknesses (e.g. Li and Zhang, 2004; Rätsch et al., 2001; Schapire and Singer, 1999). AdaBoost has been applied to face detection (Viola and Jones, 2004) and for fingerprint recognition (Liu, 2010).

2.1.4.3 Support Vector Machine

Support Vector Machines often operate by first transforming the features into a high-dimensional space (Vapnik, 1998; Burges, 1998). This transformation can be computed quite efficiently by different kernel functions (Schölkopf and Smola, 2002). A linear decision boundary with maximum margin is then established in the resultant space. Extensions of the classifier to multiple classes are proposed (Hsu and Lin, 2002; Lee et al., 2004). Support Vector Machines obtain high classification rates in many pattern recognition tasks (Sapankevych and Sankar, 2009). Numerous applications of this classifier exist. They include image classification (Chapelle et al., 1999), email categorization (Drucker et al., 1999) and 3D object recognition tasks (Pontil and Verri, 1998).

2.1.4.4 Naïve Bayes

The Naïve Bayes classifier makes use of the assumption that all features are mutually independent (Theodoridis and Koutroumbas, 2009). This assumption is often not justified, but it allows a straightforward estimation of the classifier parameters from the sample set during training. The number of classifier parameters is considerably reduced by this approach compared to classifiers that estimate more complete representations of the data distribution. The ‘curse of dimensionality’ (Duda et al., 2001) is thereby mitigated, i.e. the classifier allows working in high-dimensional feature spaces directly. The Naïve Bayes classifier has been proven to perform well in many classification tasks (Langley et al., 1992; Domingos and Pazzani, 1997).

2.1.4.5 Nearest Neighbor

An early publication in classification theory by Cover and Hart (1967) forms the basis for the Nearest Neighbor classifier. This classifier performs categorization of newly measured feature vectors based on the neighboring feature vectors that are already labeled. It can be proven that this classifier has a low achievable error rate that is, given certain requirements, between the Bayes error rate and twice the Bayes error rate (Cover and Hart, 1967). The Bayes error rate is the lowest achievable error rate of any classifier (Niemann, 2003). It also places considerable demands to computer memory and processing speed. Despite being one of the first published classification algorithms, the Nearest Neighbor classifier is still applied in recent publications (e.g. Lee, 1991; Blanzieri and Melgani, 2008). The Nearest Neighbor classifier operates very well when the data exhibits many subclusters.

2.1.4.6 Neural Networks

Neural Networks are built to simulate neuron interaction in the human brain (Specht, 1990). The neurons are implemented by multiple single nodes that are connected in multilayer nets (Duda et al., 2001). Each node has an input and an output. A feature value that is input into the node is subjected to a specified nonlinear function, e.g. a sigmoid function. Weights specify the contribution of individual nodes to the classification result. These weights can be adjusted using different learning strategies (Hagan and Menhaj, 1994). The networks are often hard to interpret (Chau, 2001b). Neural networks are frequently applied and several survey articles cover them (e.g. Baxt, 1995; Chua and Yang, 1988; Hunt et al., 1992).

2.1.5 Summary

There are several possible classification systems that can be applied for pattern recognition. The selection of the appropriate method depends on the question at hand and the boundary conditions. In order to solve a given group classification task, a careful analysis of sensor input, preprocessing, feature extraction and classifier selection has to be conducted. The different proposed algorithms for these steps have varying strengths and weaknesses. Choices with respect to the complete system have to be made in order to come up with a solution for a given task.

2.1.6 Further reading

Various textbooks for further reading on pattern recognition exist. Amongst them are classics like Fukunaga’s (1990) introduction into the topic, and Niemann’s (2003) book (in German). Among the most cited is the book by Duda and coworkers (2001), which contains many application examples. One of the most state-of-the-art books is the recently edited work by Theodoridis and Koutroumbas (2009).

2.2 Biomechanical group classification

To provide the basis for the biomechanical group classification work done in this thesis, a review of the important publications for this work will be presented. In the first section, general applications of pattern recognition to biomechanics that are not related to group classification are described. Studies that are relevant for group classification are presented in the second section.

2.2.1 General applications of pattern recognition

General applications of pattern recognition to biomechanics that are not specifically related to group classification are described in this section. Previous studies that explicitly target feature extraction methods are described first. Then, publications relating to the application of pattern classification algorithms to biomechanics are presented.

2.2.1.1 Feature extraction methods

Several feature extraction methods that are specific to biomechanics have been published. A novel approach for feature extraction from EMG measurements (Von Tscharner and Herzog, 2007) was introduced by von Tscharner (2000). The author used the property of the Wavelet transformation to give a representation of a signal both in frequency and in time. Thus, the features were capable of resolving events within the EMG signal in time, frequency and intensity. By averaging those features over multiple experiments, functional aspects of muscle activation could be determined. These so-called Wavelet features were applied in subsequent classification studies (von Tscharner, 2009; von Tscharner and Goepfert, 2003).

An important idea in these studies was to use multi muscle Wavelet patterns for subsequent classification (von Tscharner and Goepfert, 2003). These multi muscle patterns combined the information from several muscle groups. This approach allowed, for instance, employing timing relations between these muscle groups for classification.

A Wavelet feature extraction method was also used in a different study (Nyan et al., 2006). The authors classified gait patterns from three classes: ascending stairs, descending stairs and level walking. A recognition rate of more than 97% was reported. As basis for the classification, the authors used data from accelerometers that were strapped to the shoulder of their subjects. Employing accelerometer measurements for such a task requires careful action by the researchers (Nigg and Boyer, 2007). Especially the tightness of the strapping of the accelerometer has been shown to affect the measured signal amplitude. It has therefore yet to be demonstrated whether the information from this type of sensor is reliable enough to conduct more complex group classification tasks.

Results were published using the kernel Principal Component Analysis for feature reduction in a study related to gait classification (Wu et al., 2007). The feature reduction was applied to kinematic variables that were computed at specific time points of the gait cycle (e.g. heel-strike, toe-off). The computation of these kinematic variables can lead to an error amplification (e.g. when skin movement affects the marker positions) and requires additional assumptions (e.g. about the joint axis directions). Furthermore, a substantial amount of information is lost by using only specific time points of the measured time series. It has already been stated in the literature that the incorporation of time dependent patterns yields valuable information for gait analysis (Chau, 2001a).

2.2.1.2 Pattern recognition algorithms

One pattern recognition concept that is often used in biomechanical modeling is Artificial Neural Networks. An example is the prediction of dynamic muscle force from EMG measurements (Von Tscharner and Herzog, 2007). This prediction task was performed in a study (Liu et al., 1999) using such an Artificial Neural Network. In this study, a model that predicted the force produced by the cat soleus muscle from EMG measurements was presented. The results for intrasession as well as intersubject prediction showed cross-correlation values of the measured and the predicted signal of over 0.9. In study with a similar approach, an Artificial Neural Network was employed in order to simulate lumbar muscle response to static moment loads (Nussbaum et al., 1997). A further study (Sepulveda et al., 1993) showed that a Neural Network can be used to model the relationship between muscle activity and lower-limb dynamics of human gait. Neural Network classifiers were also applied for the recognition of the progression of scoliosis (Jaremko et al., 2002). A comprehensive overview of further applications of Artificial Neural Networks in the area of clinical biomechanics in general can be found in (Schöllhorn, 2004).

Other classification algorithms have also been applied in biomechanical studies. For example, a Support Vector Machine was employed to derive a diagnostic tool based on shoulder strength data (Silver et al., 2006). The goal of this study was to develop a representative shoulder strength score using a regression type of analysis. The resulting shoulder strength value was deemed essential for the post-operative evaluation of the shoulder function. The authors showed that their developed Support Vector Machine based score was a measure that seems to be promising for future applications for summarizing shoulder strength data.

2.2.2 Group classification in biomechanics

Applications of pattern recognition to group classification in biomechanics are described in this section. Studies that are not related to walking gait classification are described first. Then, publications related to the classification of walking gait are presented.

2.2.2.1 Studies not related to walking gait classification

Gender classification was performed in a classification study (von Tscharner and Goepfert, 2003) based on EMG data. The authors applied the Wavelet feature extraction method (cf. Section 2.2.1.1) that was developed earlier (Von Tscharner, 2000) to the EMG data. The Wavelet transformation that was employed was specifically adapted to be physiologically meaningful with respect to the EMG measurements. In this classification study (von Tscharner and Goepfert, 2003), the authors were able to linearly classify the gender groups with a statistically significant classification rate of 95%. The same Wavelet feature extraction technique was recently used to distinguish EMG patterns of runners during prolonged running (Stirling et al., 2009). The authors of this study showed that muscle intensity patterns from a non-fatigued (early) and fatigued (late) running phase could be classified using a Support Vector Machine with a statistically significant classification rate of 92.9%. It appears, therefore, that the physiologically adapted Wavelet transformation method is well suited for group classification based on EMG measurements.

EMG measurements were also used for the classification of injury groups (Christodoulou and Pattichis, 1999). The authors of this study applied features derived from motor unit action potential signals that needed to be identified in the EMG. Artificial Neural Networks and Discriminant Analysis were used to classify neuromuscular disorders with a classification rate of 97.6%. This technique is specific to the measured signal and not generally applicable to other types of input data.

Another study performing injury group classification has also been published (Umapathy and Krishnan, 2006) that was based on vibroarthrographic signals. Vibroarthrographic and vibromyographic measurements are a simple, non-invasive way of assessing the mechanical signal in the low frequency range that is generated by the moving joint or contracting muscle, respectively (Nigg and Boyer, 2007). The purpose of the aforementioned study was the classification of knee joint disorders using features that were specific to these vibroarthrographic measurements. The classification was based on two groups consisting of 51 normal and 38 abnormal vibroarthrographic signals of the knee joint. A classification rate of 80% was reported using a Linear Discriminant Analysis type classifier. Since this injury classification method is based on vibroarthrography, it is specific to pathologies and classification tasks that are related to joint disorders. The method is therefore not directly adaptable to more general movement analysis tasks that are not explicitly related to joints.

2.2.2.2 Studies related to walking gait classification

Many research studies focused on walking gait classification. The research can be divided in two different groups based on whether motion capture data was used for the classification.

Gait classification studies not based on motion capture data

Numerous gait classification studies were published that analyzed data that was not originating from 3D motion capturing systems. An early application of pattern recognition to the classification of normal and four pathological gait patterns was presented (Bekey et al., 1977), which used features from EMG signals measured from six muscles of the foot and ankle joint complex. For classification, a Linear

Discriminant Analysis was performed. The work reported successful classification results for 15 of the 19 cases that were investigated.

A different study (Zhang et al., 2005) presented a method for characterizing human locomotion outside the laboratory. In this study, a portable device was developed that measured timing variables describing the ground contact. The authors demonstrated that they could classify different locomotion types (level walking, running, ascending/descending stairs) with high classification rates of more than 97% for each activity.

A further study (Aminian et al., 1993) investigated the classification of body accelerations during human walking. The features computed from these body accelerations were presented to a Kohonen Neural Network classifier (Kohonen, 1990). Such Kohonen Neural Networks have been described as “the most prevalent non-traditional methodology for gait data analysis in the last 10 years” (Chau, 2001b, p. 102). The authors of the aforementioned study (Aminian et al., 1993) used this technique to successfully distinguish level and uphill walking. However, it has been suggested that an analysis based on accelerometer measurements is often difficult (Nigg and Boyer, 2007).

Another input to a gait classification system was proposed in (Begg et al., 2005). The authors employed minimum foot clearance measurements with the purpose of an age related gait classification. Minimum foot clearance is defined as the minimum height that the foot lifts off the ground during the stride cycle. Successful classification with rates of 90% was reported. This result was different than earlier findings that showed no age differences in level walking (Chen et al., 1991). It has yet to be shown whether minimum foot clearance can be used reliably to classify complex changes in human gait.

A different system based on measured pressure patterns below the foot was presented in (Barton and Lees, 1995). For this study, foot pressure patterns (Nigg,

2007) were recorded from 18 subjects during normal walking. The purpose of the study was the classification of foot pathologies into three output categories: healthy feet, pes cavus (‘high arch’) and hallux valgus (‘bunion’). The authors used a complex neural network with two hidden layers to categorize the measured maximum pressure patterns. The network inputs consisted of 1316 measured pressure values. Classification rates in the range from 77% to 100% were reported. These results were based on a relatively small number of test and training samples when compared to the high number of pressure inputs. This mismatch can lead to an overadaptation of the Neural Network to the input pressure values (Duda et al., 2001). Overadaptation can cause a poor generalization performance of the network, i.e. the classification might fail when new patterns are presented to the classifier. Furthermore, is has been criticized (Chau, 2001b) that the use of a two hidden layer network was not well motivated by the authors. It is known that for learning most functional relationships, a single hidden layer is sufficient (Bishop, 1996).

Ground reaction force data were extensively used as input to gait classification systems. An example of a classification system based on ground reaction forces was presented in (Holzreiter and Kohle, 1993). The proposed system employed Fourier transform features and an Artificial Neural Network classifier. The method was successfully applied to the distinction of normal and pathological gait, where the pathologies consisted of subjects with calcaneus fractures (71 subjects) and artificial limbs (12 subjects). In a different study using ground reaction force measurements (Mezghani et al., 2008) a method was developed to distinguish between asymptomatic and osteoarthritis knee gait patterns. In this study, the authors extracted two different feature types from the force vector variations: the coefficients of a polynomial expansion and the coefficients of a Wavelet decomposition. The features were classified with the Nearest Neighbor classifier. Linear Discriminant Analysis classifiers were used (Bertani et al., 1999) on features calculated from 3D ground reaction force measurements. The purpose of this study was to apply pattern recognition as an aid in clinical decision-making to diagnose flat foot pathologies. Success rates of over 90% were reported by the authors, indicating that their method provided important information for the evaluation of flat foot pathologies. A different approach (Wu and Wang, 2008) was based on the vertical component of the ground reaction force data. The authors performed feature extraction using the complete temporal information by applying the Principal Component Analysis. Subsequently, a Support Vector Machine classifier with polynomial kernel was utilized to classify gait patterns 30 young and 30 elderly participants. A classification rate of 90% was reported. Similar methods that use the complete collected temporal information could also be beneficial when motion capture data is additionally used for the classification (Chau, 2001b).

Gait classification studies based on motion capture data

For several group classification tasks related to gait, additional contributions for classification can be identified by analyzing selected body kinematics. For example, several studies showed that the motion of shoulder and hip in the sagittal plain is an important feature for gender differentiation (Barclay et al., 1978; Cutting et al., 1978). It was furthermore demonstrated (Mather and Murdoch, 1994) that the extent of lateral body sway is another postural aspect that changes in the gender example. This information about postural changes in gait is collected in studies that employ 3D marker-based measurements of selected body kinematics for classification.

Several publications reported on utilizing the kinetic and kinematic data computed from 3D markers and force plate measurements. Schöllhorn and colleagues (2002) used these data and self-organizing Kohonen maps (Kohonen, 1990), a type of Neural Network for unsupervised learning, to distinguish different heel heights of females walking in dress shoes. This method was successful in identifying group and subject specific gait changes. A similar approach was investigated for a different classification task in (Janssen et al., 2008). In this study, the authors also analyzed kinetic and kinematic data with an Artificial Neural Network. The purpose of their study was the recognition of emotion. In two experiments, the authors first classified four emotional states (normal, happy, sad, angry) from gait patterns, and then analyzed effects on walking gait patterns when listening to three types of music (excitatory, calming, no music). In another study based on Neural Network classification (Barton and Lees, 1997) hip-knee joint angle diagrams were used as the basis for identification of gait patterns. The purpose of this study was the classification of three conditions: normal walking, simulated leg length difference and simulated leg weight difference. Eight healthy subjects walked on a motorized treadmill at constant speed and hip and knee angles were calculated from a set of four reflective markers. The reported classification rate for discriminating the three walking conditions was 83.3%. As they did in previous work (Barton and Lees, 1995), the authors employed a complex Neural Network with two hidden layers. Again, no justification for the second hidden layer was given.

A classification of patellofemoral pain syndrome injury groups based on kinetic and kinematic variables was performed in (Lai et al., 2009). The authors reported a classification rate of 88.9% for the classification of subjects that suffered from this syndrome and such that do not. A Support Vector Machine was used for the purpose of classification. In this study, the authors used features that were specific to the kinematic and ground reaction force measurements that they collected. Only certain time points within the measured time series were analyzed and temporal information was therefore discarded. Furthermore, the groups under consideration were based on retrospective assignment of the injury classes. This retrospective assignment is an approach that might not give optimal results (Stefanyshyn et al., 2006).

In a different study (Wu et al., 2006), a classification system that was also based on kinetic and kinematic data was published. The purpose of this study was age related gait classification. The authors used the computed kinematics and kinetics at specific time points of the human gait phase as basis for the classification. The authors applied kernel-based Principal Component Analysis (KPCA) to these data to extract relevant information from the highly correlated time-dependent gait variables. They reported that this procedure improved the generalization performance of the Support Vector Machine classifier that they employed. The authors obtained classification rates of 89.6%. In a similar study (Begg and Kamruzzaman, 2005), spatio-temporal features (e.g. double support time, stride length) were used in addition to kinetic and kinematic variables. The authors also investigated the application of a Support Vector Machine classifier to recognize young and elderly gait groups. In total, their feature set comprised 24 gait features and they obtained an overall classification rate of 91.7% when using them for the classification.

The aforementioned approaches have two major shortcomings. Firstly, the methods require the computation of kinetics, kinematics and sometimes spatio-temporal variables. These computations have several disadvantages, because the determination of these variables a) usually requires additional assumptions, for example, about the joint axes direction, which are difficult to validate; b) might lead to an error amplification, for example when skin movements affect the marker positions; c) is often time consuming. Secondly, only specific time points of the gait cycle are used to extract the feature vectors. However, a large amount of time dependent information that is contained in the measurements is possibly neglected (Chau, 2001a). Therefore, a method that is directly based on the measured 3D marker positions and incorporates the complete gait cycle information for classification might be better suited for gait classification. This approach is generally not used in biomechanics.

A method was developed (Troje, 2002), that was directly based on the measured 3D marker positions. The author used a Principal Component Analysis based feature extraction method. The method was applied in gender classification using a Linear Discriminant classifier. This method did not require the intermediate step of calculating kinematic features. However, the author also discarded temporal information and fine structure within the gait data by applying a simple modeling to the measured variables.

2.2.3 Summary

Several successful applications of pattern recognition to biomechanics were presented in this section. Among these, a considerable number of publications were related to gait based classification. Some studies that were not based on kinematics provided important cues for postural aspects of locomotion. Some studies related to gait classification were using features computed from specific time points of the collected kinetic and kinematic variables. Still missing in the literature is a method that uses the complete temporal information from motion capture data for classification. Ideally, this method should not be specific to the input measurements to be transferable to different classification tasks. Furthermore, it should be able to identify and visualize high-dimensional dependencies of the classification on the computed features. A first attempt to develop such algorithms was conducted in this thesis.

2.3 Embedded classification

To provide the basis for the embedded classification work done in this thesis, a review of the important literature for this work will be presented. In a first section, relevant literature for the employed measurement method is included. Literature that is related to a general framework for classification on an embedded microprocessor is presented in a second section. Studies that are related to the application of this framework in sports classification tasks are presented in a third section.

2.3.1 Measurement method

Hall sensors have been employed to collect the relevant data for embedded classification. Hall sensor measurements are based on the Hall effect, which “is one of the best known and earliest exploited solid-state sensor effects” (Popovic, 1989, p. 39). The effect has been discovered by Hall (1879). The sensor that measures this effect evaluates the potential difference that is created on both sides of a conductor when it is placed in a magnetic field. The potential difference is proportional to the magnetic field strength when the electric current flowing through the conductor is constant. Present-day Hall sensor applications are made possible by developments of semiconductor technology. In 1948, a germanium Hall device has been proposed as a magnetic sensor (Pearson, 1948). Such devices are small, lightweight and power-efficient and ideally suited for integration into embedded systems. Applications of this sensor technology can be found for example in field applications to identify metal objects (Ripka, 1994) and for contactless wear-free angular position measurements (Lozanova and Roumenin, 2010).

2.3.2 Embedded classification framework development

Only few classification studies exist that target embedded devices. None of them provides general considerations for implementing a pattern recognition system on microprocessors. Those studies that have been published represent explicit solutions for a given problem. An early example for such a study (Kerrick and Bovik, 1988) described a system for the recognition of hand printed English characters. For this purpose, an 8-bit microprocessor was employed. Another early study (Botros, 1988) investigated the differentiation of normal or abnormal human liver tissue by ultrasound. Both studies do not consider a general approach to microprocessor classification.

It was also shown that real-time gesture recognition is possible on a specific device, so-called smart cameras (Wolf et al., 2002). These smart cameras are a new generation of digital cameras that directly analyze the scene they are pointed at. Computationally powerful 32-bit microprocessors with 100 MHz clock frequency were employed.

Possibilities to naturally control upper extremity prostheses were investigated in (Englehart and Hudgins, 2003). For this purpose, four channels of myoelectric signals were measured and multiple classes of desired limb movement were discriminated. The classification was subject based and learned the muscle activation patterns for each desired class for each individual.

In another application of pattern recognition algorithms to microprocessor classification (Rahman et al., 2004) an electronic taste sensing system was described. This system was capable of discriminating liquid samples that did and did not contain Eurycoma longifolia (a type of herbal remedy). Specially fabricated screen-printed arrays of lipid-membrane sensors were employed for this task. A simple Artificial Neural Network was employed for the implementation on the microprocessor.

A different classification system (Hacker et al., 2006) used the video camera of a cell phone to identify whether or not the user is directly looking into the camera. For this purpose, the AdaBoost classifier (Freund and Schapire, 1997) and an established face detection algorithm (Viola and Jones, 2004) were applied. A further study (Benbasat and Paradiso, 2002) also reported about classification on a mobile communication device, a Palm III (Palm Inc., Sunnyvale CA, USA). In this publication, a gesture recognition application based on inertial sensor measurements was described. Simple motions were classified by the developed system, such as movements in a straight line, twists, etc. These simple motions were then combined in order to recognize composite gestures that could then be tied to output routines.

An example of a microprocessor classification system for the recognition of robot motion (Stewart and Wang, 2003) used feedback patterns produced by the robot's electromotor in the analog waveform. A Support Vector Machine was used for the purpose of data classification.

All of these studies consider the implementation of pattern recognition algorithms on the embedded system that they target, but they do not discuss a general methodology that could be used for classification on microprocessors. No feature selection strategy is applied and different classifiers are not compared.

One paper (Boni et al., 2005) was more oriented towards making classification feasible on general microprocessor hardware. However, the authors did not discuss the complete pattern recognition flow chart (Fig. 1.1). Their work aimed at an efficient implementation of one specific classifier, the Support Vector Machine, on an 8-bit microprocessor. This contribution is only one building block of a general methodology for embedded classification. If the Support Vector Machine classifier, or any other type of classifier, is selected for the implementation on the embedded device, efficient implementation strategies need to be considered.

A similar idea was followed in a different publication (Benbasat and Paradiso, 2004). The authors explored general design rules for techniques, which could reduce power consumption of the employed microprocessors. This was facilitated through real-time sensor selection and reduced storage requirements for the computed features. This publication is not a dedicated classification study, but the need for power consumption reduction is important for all autonomous embedded devices. Reduced power consumption should be observed in the considerations for a general embedded classification methodology.

2.3.3 Applications of the embedded classification framework

The embedded classification methodology that is proposed in this thesis has been employed in two running related tasks. These are the classification of the surface condition that a runner is on in Chapter 5, and the classification of the current running speed and inclination in Chapter 6. A compression measurement based on a Hall sensor has been made for this purpose. In this section, studies that are related to these classification tasks are reviewed.

Publications that use such a Hall sensor (cf. Section 2.3.1) for the given classification purposes do, to the knowledge of the author, not exist yet. Surface classification methods can only be found for research domains other than biomechanics and sports. For example, a land surface classification methodology based on satellite images was presented (Welch et al., 1992). The applied method was based on Discriminant Analysis and Neural Network classifiers and only classified large scale surface regions for meteorology. A method for large scale surface classification in a remote sensing application was also published (Schaale and Furrer, 1995). The authors employed spectrographic imagers mounted on an airplane for data acquisition and classification using a Kohonen Neural Network classifier (Kohonen, 1990). A further example was contributed in (Zhou et al., 2009). In this study, surface classification of objects was performed by utilizing a Laser scanner. The authors used a simple threshold classifier for their surface classification. None of these approaches can be transferred to a running surface classification task in sports because each of them targets a completely different application domain.

Running speed and surface inclination classification applications can be found in some biomechanics related publications. An example study (Aminian et al., 1995) proposed a method based on accelerometry for walking gait. The authors applied Neural Networks for the walking speed measurement task. This technique was later extended for outdoor running (Herren et al., 1999). Subsequently, this approach for speed classification was further modified (Song et al., 2007) in order to obtain more accurate results. All these approaches made use of accelerometer measurements. Accelerometers inherently give information about movement speed, but their application has been criticized in biomechanics related tasks (Nigg and Boyer, 2007). It has, for instance, been demonstrated that the tightness of the mounting of the accelerometer considerably influences the amplitude of the measured acceleration signal. Even more important in the context of embedded classification is that complex Neural Networks are needed for the speed and inclination classification. Complex Neural networks require a considerable memory (Williams and Zipser, 1989) and processing power (Orponen, 1994) and can not be efficiently implemented on a microprocessor.

2.3.4 Summary

The presented studies in Section 2.3.2 showed specific implementations of pattern recognition algorithms on embedded hardware. A general discussion of design criteria for an implementation of the complete pattern recognition procedure (Fig. 1.1) on embedded systems is still missing in the literature. Such a more general discussion of a methodology for embedded classification was attempted in this thesis.

Using this developed methodology, it was also attempted in this thesis to successfully solve embedded surface, speed and inclination classification tasks in sports. These tasks were based on Hall sensor measurements (cf. Section 2.3.1).

The presented studies in Section 2.3.3 showed that these embedded classification tasks were also addressed for the first time.

2.4 Concluding remarks

The review of the literature has shown that a considerable number of applications of pattern recognition methods to biomechanics exist. Most of these previous studies reported successful classification results, indicating the benefits of a data driven analysis that pattern recognition algorithms offer. However, several weaknesses of existing techniques were identified. Furthermore, a number of research questions that have not been answered yet in the literature were disclosed in Chapter 1. A careful selection and implementation of the different available pattern recognition techniques has to be made in order to successfully answer these research questions.

CHAPTER 3 PATTERN CLASSIFICATION OF KINEMATIC AND KINETIC RUNNING DATA TO DISTINGUISH GENDER, SHOD/BAREFOOT AND INJURY GROUPS WITH FEATURE RANKING

3.1 Introduction

Biomechanical studies have been conducted to use kinematics, kinetics, soft tissue vibrations and/or EMG data to distinguish between groups such as gender (von Tscharner and Goepfert, 2003), age (DeVita and Hortobagyi, 2000), footwear (Nigg et al., 2003) and to identify injury mechanism characteristics (Stefanyshyn et al., 2006). The classification is often done by comparing means and standard deviations of discrete variables (e.g. peak impact force, maximal foot eversion, etc.). Recently new methods have been proposed and applied for such classifications based on pattern recognition methods. For example, it has been shown (Janssen et al., 2008) that Artificial Neural Networks can identify emotional state from human gait data. Furthermore, Support Vector Machine classifiers have been used (Begg and Kamruzzaman, 2005) to differentiate young and elderly gait patterns. These and other published classification methods (Schöllhorn et al., 2002; Wu et al., 2007) were successful in identifying groups using biomechanical data. However, the results depended obviously on the analyzed features, and those features were very specific to the input measurements and the specific studies. Furthermore, the applied methods did not always provide information about what exactly characterized the differences between the groups identified.

Therefore, the purpose of this paper was to develop a pattern classification approach for typical tasks of biomechanical group classification that was not specific to the input measurements and additionally provided information about the group differences. This general approach was applied to three different groups without adaptation. In particular, groups composed of gender, shod/barefoot running and injured/non-injured subjects were considered. While the first two groups were primarily tested for proof of concept, the last group had clinical relevance. The injury that was examined was the patellofemoral pain syndrome, a condition affecting up to a quarter of all persons active in sporting activities (Devereaux and Lachmann, 1984; Malek and Mangine, 1981). For the purpose of classification, generic features were used that were not in any way specific to the collected measurements. These measurements consisted of kinetic and kinematic data collected during a longitudinal running injury study (Stefanyshyn et al., 2006).

3.2 Methods

The employed classification methods (Fig. 3.1) followed a classical pattern recognition approach (Duda et al., 2001). Dynamic biomechanical measurements (Section 3.2.1) were first subjected to preprocessing to enhance the signal properties. The output of the subsequent feature extraction step (Section 3.2.2), the feature vector $\boldsymbol{c} \in \Re^{N_c}$, described the input in $N_{\boldsymbol{c}}$ dimensional feature space.

All groups under investigation were differentiated using AdaBoost (Freund and Schapire, 1997), a supervised classifier that has the inherent ability to select features according to their importance for the classification task (Section 3.2.3). Thus, it was possible to identify which features contributed most to the differentiation of the groups under investigation.

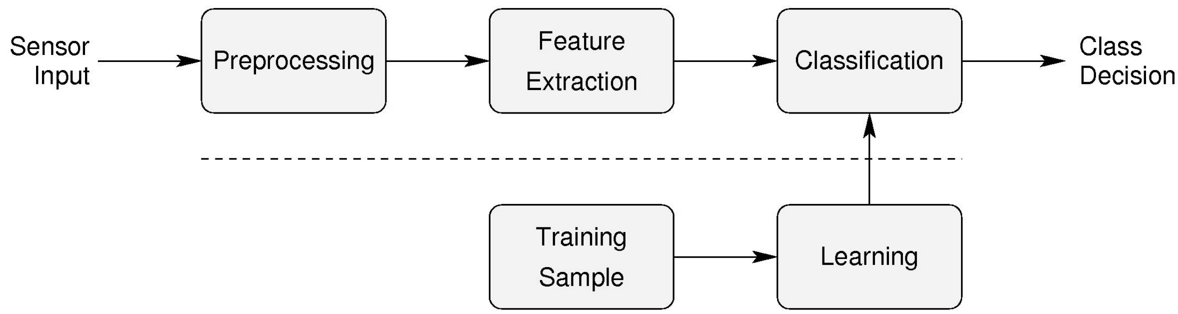

Fig. 3.1. Overview of a pattern recognition system.

For AdaBoost training, a labeled sample S_{train} was used (Fig. 3.1 below the dotted line). The label assigned a feature vector $\boldsymbol{c}$ to a class $\omega_\kappa \in \Omega$, where

Ω	set of N_Ω possible classes
$\kappa = 1,\ldots,N_\Omega$	class number.

One example class label was the runner gender, with $N_\Omega = 2$ in that particular case. The class specific statistical properties were learned during training and decision criteria were established. In order to test the generalization performance of the AdaBoost classifier, cross-validation was performed (Section 3.2.4).

3.2.1 Data for experimental evaluation

The process of collecting the experimental data for the prospective study has been described earlier (Lun et al., 2004; Stefanyshyn et al., 2006). The study was reviewed and approved by the University of Calgary Conjoint Health Research Ethics Board. Shortly, at the beginning of the running season, 153 recreational runners (71 women, 82 men) were recruited, written consent was obtained and anthropometric and biomechanical measurements were conducted. During the following six months, any running related injuries were recorded and a running journal was kept.

Of the initial 153 subjects, N_r = 80 (40 women, 40 men) were suitable for analysis (Tab. 3.1). 73 subjects dropped out of the study because they stopped running (4), work (2), incomplete running journal (16), injury at time of recruitment (2), loss to follow up (49).

Tab. 3.1. Characteristics of the 80 runners who completed the prospective running injury study. Values are presented as mean (standard deviation).

	n	Age [y]	Mileage [km]	Experience [y]	Height [m]	Mass [kg]
Women	40	36.0 (8.8)	33.7 (16.9)	6.4 (6.5)	1.68 (0.08)	66.2 (10.0)
Men	40	41.1 (8.7)	36.3 (16.8)	11.4 (9.2)	1.79 (0.07)	85.7 (13.6)
Mean		38.5 (9.1)	35.0 (16.8)	8.9 (8.3)	1.74 (0.10)	75.9 (15.4)
Max		64.0	100.0	40.0	1.93	123.5
Min		22.0	6.0	0.5	1.55	47.6

3.2.1.1 Dynamic biomechanic measurements

All runners had data collected on both legs. The measurements included a shod condition, with the runners wearing their own running shoes and a barefoot condition. Between two to ten complete dynamic datasets were collected for each subject. In total, 496 datasets were used for experimental evaluation.

Three leg segments (upper leg, lower leg, foot) were prepared using a total of nine reflective markers. The 3D marker positions were collected using four electronically shuttered, high-speed video cameras (NAC MOS-TV, V-14B, Japan) equipped with 12.5 mm – 75 mm zoom lenses (Cosmicar, Japan) and a VP310 video processor (Motion Analysis Corp, Santa Rosa, California). 3D force data were collected using a force platform (Kistler AG, Winterthur, Switzerland) mounted flush with the floor in the center of a 30 m runway. Running speed was controlled at v = 4.0 ± 0.2 m/s using two photocells, 1.9 m apart, at shoulder height. Markers were tracked for 50 ms before and after force plate contact. Kinematic and kinetic data were imported into Kintrak 4.0 (Motion Analysis Corp) for analysis. Joint attitude and angular motions were determined using a 3D joint coordinate system implemented in Kintrak 4.0.

A total of N_{dyn} = 51 variables were collected for each leg of the athlete (Tab. 3.2). Each of them consisted of N_f = 101 normalized time frames from touch-down on the force platform to toe-off. These variables are referred to as $d_i[k]$, with $i = 1,\ldots,N_{dyn}$ denoting the specific measurement (Tab. 3.2).

Tab. 3.2. Dynamic variables that were acquired during data collection.

Number *i*	Measured variable
1-3	Ground reaction force; vertical, medial-lateral, anterior-posterior
4,5	Center of pressure location; medial-lateral, anterior-posterior
6	Free moment on the force plate
7-10	Ankle flexion-extension angle, velocity, moment, power
11-14	Ankle inversion-eversion angle, velocity, moment, power
15-18	Ankle abduction-adduction angle, velocity, moment, power
19-22	Knee flexion-extension angle, velocity, moment, power
23-26	Knee abduction-adduction angle, velocity, moment, power
27-30	Knee internal-external rotation angle, velocity, moment, power
31	Hip flexion-extension moment
32	Hip abduction-adduction moment
33	Hip internal-external rotation moment
34,35	Foot sagittal segment plane angle, angular velocity
36,37	Foot frontal segment plane angle, angular velocity
38,39	Foot transverse segment plane angle, angular velocity
40,41	Shank sagittal segment plane angle, angular velocity
42,43	Shank frontal segment plane angle, angular velocity
44,45	Shank transverse segment plane angle, angular velocity
46,47	Thigh sagittal segment plane angle, angular velocity
48,49	Thigh frontal segment plane angle, angular velocity
50,51	Thigh transverse segment plane angle, angular velocity

3.2.1.2 Injury information

After completing all measurements, the athletes were observed on a monthly basis for six months during their usual training routine from April to September (Lun et al., 2004). For all runners participating in the study it was documented if any injury attributed to running occurred. An injury was defined as any musculoskeletal symptom of the lower limb that required a reduction or stoppage of normal training. A weekly drop-in injury clinic was available to subjects for evaluation of injuries. The injuries were assessed by two experienced sports medicine doctors at the University of Calgary Sport Medicine Centre.

3.2.1.3 Class labels

The data were analyzed with class labels based on gender, on the shod/barefoot condition, and on whether the runners developed a specific injury type (patellofemoral pain syndrome, PFPS). For each experiment, the dynamic dataset of a single step cycle was labeled according to the membership to a certain class. For gender classification, 244 datasets were from females and 252 from males. For the shod and barefoot experiments, 217 and 279 sets were assigned to each class, respectively.

The injury group consisted of runners that suffered from patellofemoral pain syndrome during the six month study period, compared to matched uninjured runners. This specific injury was selected for this project, because (a) this was the most frequent injury in the six months prospective study and (b) it is a very common injury in runners, affecting up to a quarter of all persons active in sporting activities (Devereaux and Lachmann, 1984; Malek and Mangine, 1981). For this experiment, six patients that were diagnosed with patellofemoral pain syndrome by the clinicians were matched with respect to mass, gender, mileage and running experience (Tab. 3.3) with six subjects who remained injury free throughout the study (Stefanyshyn et al., 2006). These characteristics have been proposed to be associated with injury and/or to have an influence on resultant joint moments (Moisio et al., 2003; Van Mechelen, 1992). In this study, 28 data sets from the injury (PFPS) group were classified against 27 data sets from the matched asymptotic runners (ASYM group).

Tab. 3.3. Comparison of the characteristics of the six injured patients and the six asymptotic patients who were used as matched controls.

Patient	Injured side	Gender	Mileage [km]	Experience [y]	Mass [kg]
Injured #1	Left	Female	12.0	1.0	65.5
Match #1			15.0	0.7	59.0
Injured #2	Left	Female	40.0	16.0	59.3
Match #2			35.0	14.0	63.5
Injured #3	Right	Female	15.0	1.0	76.6
Match #3			15.0	2.0	76.2
Injured #4	Left	Male	55.0	20.0	82.5
Match #4			60.0	15.0	75.2
Injured #5	Right	Male	40.0	1.5	79.1
Match #5			40.0	4.0	84.0
Injured #6	Right	Male	30.0	1.5	76.0
Match #6			30.0	4.0	74.1
Mean					
Injured			32.0	6.8	73.2
Non-injured			32.5	6.6	72.0

3.2.2 Feature extraction

Generic features for classification were calculated in order to be independent from specific characteristics of the original measurements. Generic in this context means that the features were not restricted to the calculation of key variables like angles or forces at specific time points. Rather, their calculation considers the complete temporal information of the measurements. Consequently, the set of features was chosen so that it represented an arbitrary measurement as completely as possible. Therefore, the features can be straightforwardly applied in different group classification tasks. The chosen features have already proven to perform well in other pattern recognition tasks (Fan and Wang, 2002; Furui, 1986; Hafed and Levine, 2001).

3.2.2.1 Basic features

Basic features for time-dependent dynamic measurements $d_i[k]$ were mean

$$c_{i,\mu} = \mu_i = \frac{1}{N_f}\sum_{k=1}^{N_f} d_i[k] \tag{3.1}$$

and variance

$$c_{i,\sigma} = \sigma_i^2 = \frac{1}{N_f - 1}\sum_{k=1}^{N_f} (d_i[k] - \mu_i)^2 \tag{3.2}$$

for each measurement curve with

N_f	number of discrete time frames,
k	discrete measurement time point, $k = 1,\ldots,N_f$,
i	specific measurement.

Further basic features were derived from temporal positions and from the actual and absolute maxima and minima of the curves (Tab. 3.4).

Tab. 3.4. Basic extrema features that were generated. Min, Max, Abs and Idx abbreviate Minimum, Maximum, Absolute Value and Index of, respectively. The variable i that is used to distinguish the different features corresponds to the number assigned to the measurements (Tab. 3.2).

Feature		Formula
Min	$c_{i,min}$	$\min\ D_i[k]$
Max	$c_{i,max}$	$\max D_i[k]$
Abs Min	$c_{i,absmin}$	$\min\ \lvert D_i[k]\rvert$
Abs Max	$c_{i,absmax}$	$\max \lvert D_i[k]\rvert$
Idx Min	$c_{i,minidx}$	$\arg\min_k\ D_i[k]$
Idx Max	$c_{i,maxidx}$	$\arg\max_k D_i[k]$
Idx Abs Min	$c_{i,absminidx}$	$\arg\min_k\ \lvert D_i[k]\rvert$
Idx Abs Max	$c_{i,absmaxidx}$	$\arg\max_k \lvert D_i[k]\rvert$

3.2.2.2 Transformation features

Transformation features project the original function into a different space (e.g. frequency space). In this work, discrete cosine transform (DCT, (Ahmed et al., 1974)) features were used for the purpose of incorporating the information contained in the frequency components of the measurements. These features have been successfully used for 2D image classification (Fan and Wang, 2002), and face recognition (Hafed and Levine, 2001). For the discrete cosine transform, the original function values $d_i[k]$ were linearly transformed into real numbers $D_i[f]$, $f = 1,\ldots,N_f$, that represented the measurement function in the frequency domain. The transformation formula was

$$\mathrm{DCT}(d_i[\mathrm{k}]) = D_i[f] = \sum_{k=1}^{N_f}\left(d_i[k]\cos\left[\frac{\pi}{N_f}\left((k-1)+\frac{1}{2}\right)(f-1)\right]\right) \tag{3.3}$$

where

f discrete frequency.

The values were used as features $\boldsymbol{c}_{i,dct}$ by setting

$$c_{i,dct,f} = \mathrm{DCT}(d_i[\mathrm{k}]). \tag{3.4}$$

In this example, the high-frequency components contributed little information. Consequently, only the first ($f = 1,\ldots,N_{dctcut}$, $N_{dctcut} < N_f$) coefficients were considered. The parameter N_{dctcut} was set to 30 for this study, because no activity was observed in components with higher frequency.

3.2.2.3 Regression features

Cosine transform features described the measurement frequency characteristics, but did not sufficiently characterize general trends. Therefore, polynomial regression (PR) features were calculated. Regression features have successfully been used in speech recognition (Furui, 1986), but also in gait classification tasks based on ground reaction force data (Mezghani et al., 2008). Depending on the degree N_p of the chosen polynomial, signal properties such as gradient or curvature are adequately described. For this purpose, each measurement $d_i[k]$ was approximated in a least squares sense by a polynomial function

$$p(x) = \sum_{m=0}^{N_p} r[m]x^m \tag{3.5}$$

where

N_p number of polynomial coefficients,

$r[m]$ polynomial coefficients, $m = 0,\ldots,N_p$.

The coefficients were used as features by setting

$$c_{i,pr,m} = r_i[m]. \tag{3.6}$$

The polynomial degree N_p was set to three for this study, which corresponds to a cubic fit of the original measurements. Experimental results showed no improvement when calculating fits of higher degree.

3.2.3 Pattern classification using AdaBoost

AdaBoost (“adaptive boosting”, (Freund and Schapire, 1997)), is a meta algorithm for supervised learning that utilizes other learning algorithms (i.e. classifiers). The idea of AdaBoost is to use a linear combination (ensemble) of multiple simple decision rules (weak classifiers), which combined create a more accurate complex decision rule (a strong classifier). If each classifier predicts the correct class with an accuracy $\alpha_{simple} > 0.5$, then the ensemble classification accuracy approaches $\alpha_{complex} \rightarrow 1$ as the number of simple classifiers increases (Boland, 1989).

The AdaBoost classifier was trained in several iterations, adding one simple classifier to the ensemble in each of these iterations. The iteration number for all experiments was set to 20 to prevent overadaptation of the classifier to the input data. During training, each feature vector $\boldsymbol{c}_t,\ t = 1,\ldots,N_t$, where

N_t number of samples in the training set S_{train},

was assigned a weight w_t. These weights (Freund and Schapire, 1997) were initially equally distributed with

$$w_t = \frac{1}{N_t} \quad \forall \quad t = 1\ldots N_t. \tag{3.7}$$

As weak classifier for AdaBoost a so-called decision stump (Viola and Jones, 2004) was used. This is a simple classifier that uses a threshold value in only one of the N_c feature space dimensions. Using the decision stump weak classifier, AdaBoost performed implicit feature selection. By counting how often each feature contributed to the final decision process, a measure of the ability of a certain feature to separate the trained classes was derived. As final hypothesis for classification, a weighted vote of all classifiers in the ensemble was cast.

3.2.4 Cross-validation

In order to test the classifier for generalization performance and to prevent overadaptation to the training samples, cross-validation (Duda et al., 2001) was employed. The available data were partitioned into a fixed number N_{CV} of sets. Then the selected classifier was trained using $N_{CV} - 1$ sets. Following training, the remaining set S_{test} was used for classifier testing. By iterating this process until each of the N_{CV} sets was used as test set once, generalization performance was tested because the information in the test set had always been unknown to the classifier. For each cross-validation iteration $v = 1,\ldots,N_{CV}$, it was recorded which features were selected for classification, and the test classification rates $\alpha_{test,v}$ were also stored. The final classification rate α_{test} was calculated as the mean over all v iterations.

The samples of one person were naturally more similar among each other than samples of different persons. Thus, the data were partitioned into $N_r = 80$ disjoint sets for cross-validation so that the samples from one athlete were always in the same set, i.e. $N_{CV} = N_r$.

3.2.5 Statistics

Classification rates were deemed significant if the null hypothesis that classification was random could be rejected using a binomial test with significance level $\alpha = 0.01$. A one-tailed t test was used for statements about the statistical significance of differences of single features between groups with significance level $\alpha = 0.01$.

3.3 Results

3.3.1 Gender classification

For gender, the mean classification rate was 84.7%. This is significantly different from random ($p < 0.001$). Two basic features were present in each cross-validation iteration (Fig. 3.2), the variance c_σ of the hip flexion-extension moment and the variance c_σ of the vertical ground reaction force. The top 30 ranked features included 19 basic and 11 discrete cosine transform features. The first polynomial regression feature appeared at rank 33.

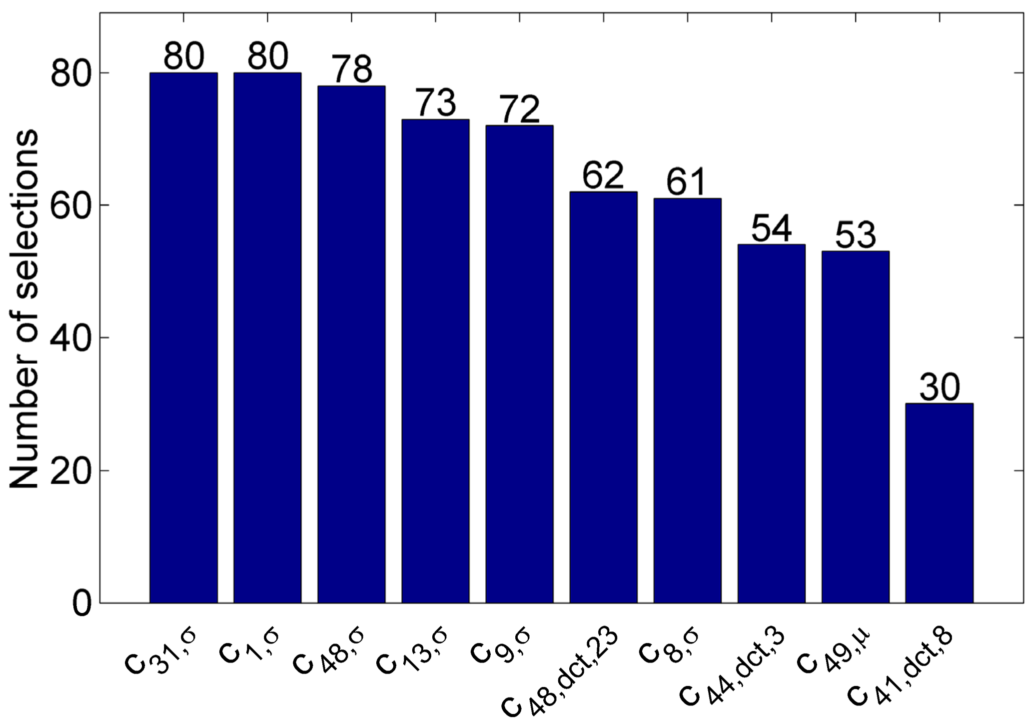

Fig. 3.2. The ten features that were most often selected for gender classification. Numbers above each bar represent how often each feature was selected. The respective feature identifiers are defined in Eq. (3.1), (3.2), (3.4) and (3.6).

3.3.2 Shod/barefoot classification

For shod/barefoot, the mean classification rate was 98.3%, with 8 of 496 sets misclassified. This is significantly different from random ($p < 0.001$). Two regression features were present in each cross-validation iteration (Fig. 3.3), the quadratic polynomial component $c_{pr,2}$ of the foot sagittal plane angle and the linear polynomial component $c_{pr,1}$ of the shank sagittal plane angle. The top 30 ranked features included 16 cosine transform, 10 polynomial regression and 4 basic features.

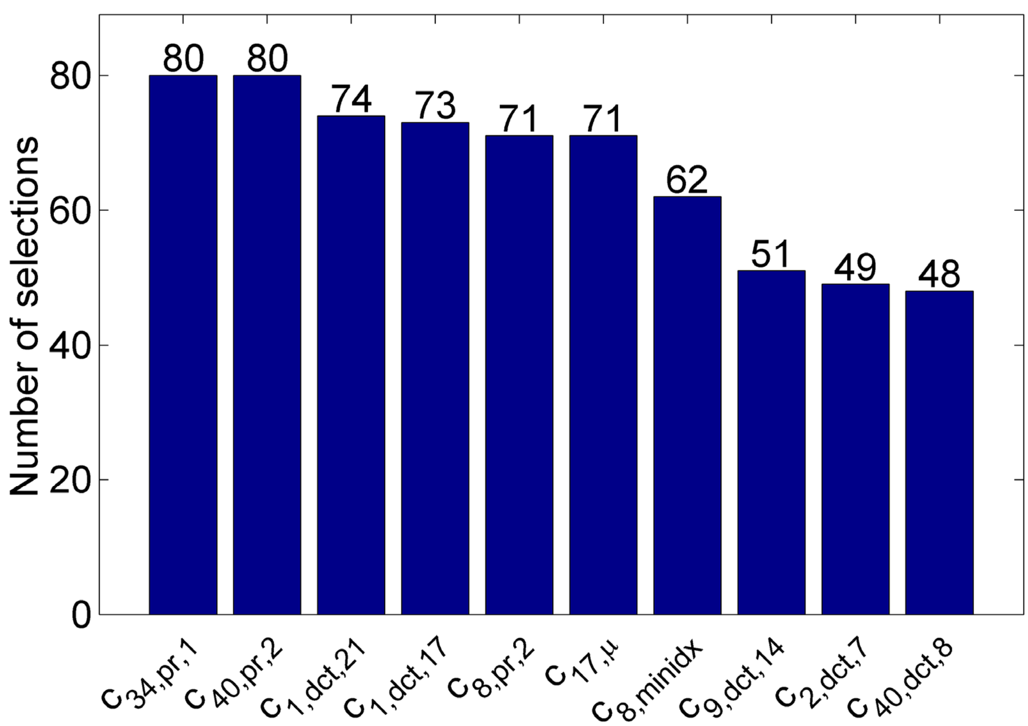

Fig. 3.3. The ten features that were most often selected for shod/barefoot classification. Numbers above each bar represent how often each feature was selected. The respective feature identifiers are defined in Eq. (3.1), (3.2), (3.4) and (3.6).

3.3.3 Injury/non-injury classification

For injury/non-injury, the mean classification rate was 100%. Thus, every single dataset and in effect every runner was assigned to the correct group. This is significantly different from random ($p < 0.001$). One single feature was selected in each cross-validation iteration, the mean c_μ of the hip abduction moment. This basic feature allowed classification without further combination with other features.

The mean hip abduction moments were significantly higher ($p < 0.001$) for all six PFPS group patients compared to those of the six matched ASYM group patients (Fig. 3.4). This difference was also visible in the originally measured mean hip abduction moments (Fig. 3.5).

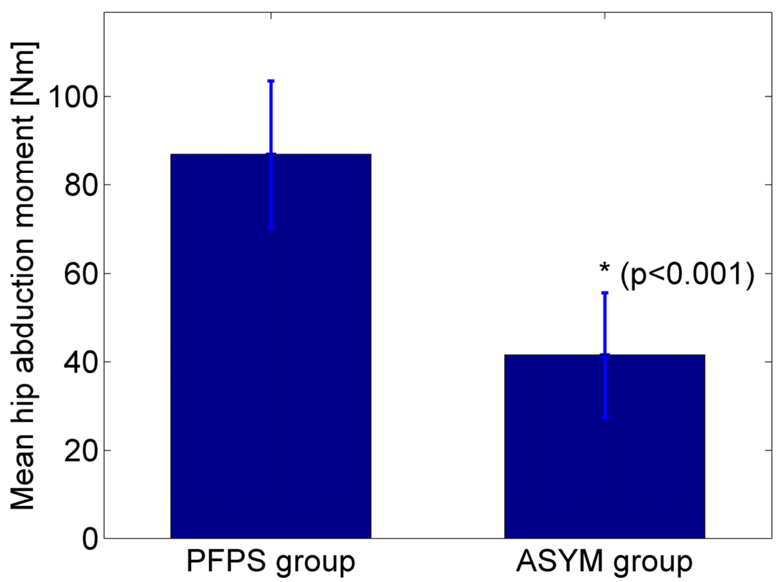

Fig. 3.4. Mean (standard deviation) hip abduction moments for the six patients who developed patellofemoral pain syndrome (PFPS group) and the six asymptomatic matched controls (ASYM group).

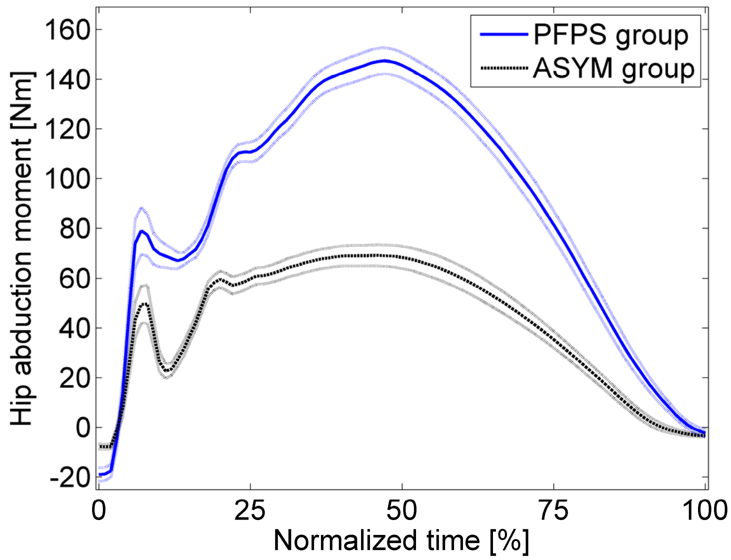

Fig. 3.5. Mean (standard error) resultant hip abduction moments for the six patients who developed patellofemoral pain (PFPS group) and the six asymptomatic matched controls (ASYM group).

3.4 Discussion

One purpose of this paper was to show that the applied methods are capable of pointing to the characteristics that discriminate groups. For the case of injury/non-injury classification, one single feature was found that distinguished runners from the injury group and the asymptomatic group with 100% accuracy. Because it is only a single feature, traditional methods should provide the same result. However, the method presented in this paper are not relying on assumptions that may be subjectively biased but allow an objective, data-driven evaluation of the input variables. Stefanyshyn and colleagues (2006) had identified knee angular impulse as a predictor of patellofemoral pain syndrome in an earlier study on the same data using traditional methods. However, this variable only discriminated the runners on a one-to-one matched basis, and is individually true only for five of the six matched pairs. The result suggested by the presented algorithm shows that using the mean hip abduction moment as a discriminating predictor, all individuals can be classified simultaneously and correctly without the need to heed the matching.

Because prospective data were only available for six runners that developed patellofemoral pain syndrome, it remains to be evaluated whether this single indicator is also sufficient for a larger group of patients. However, even if a single variable is not sufficient for larger groups, the presented algorithms would be capable of finding those variables that are additionally needed for discrimination. Furthermore, the results are highly significant even for the relatively small group and indicate that increased hip abduction moments should be deemed risk factors that play a role in the development of patellofemoral pain in runners. Muscular deficits in the hip abductors have already been presumed to be a major factor in the development of knee injuries in runners (James, 1995; Fredericson et al., 2000). Footwear and running style can influence hip abduction moments, and the appropriate manipulation of these variables may play a preventive role for patients who are predisposed to patellofemoral pain.

The ability to discriminate classes has also been shown for gender and shod/barefoot classification. Although those tasks lack the clinical relevance of the injury classification, a different strength of the methodology was shown. While traditional methods often evaluate individual discrete variables, the presented algorithms are able to evaluate what combination of features is needed to discriminate classes. Those features that were most often selected contain the important indicators for the difference between the classes. For gender classification, two variables (hip flexion-extension moment and vertical ground reaction force) were always selected (Fig. 3.2). The hip flexion-extension moment has already been shown to be different for gender by Kerrigan and colleagues (1998) for walking. The importance of the vertical ground reaction force for the gender differentiation can be explained by the higher weight of the male study participants (Tab. 3.1). For shod/barefoot classification, the variables that were most often selected for classification (Fig. 3.3) were the foot and shank sagittal plane angles. These variables have already been reported to be significantly different for shod/barefoot running earlier (De Wit et al., 2000).

Highly significant classification rates could be achieved in all experiments. This shows that the generic features computed for the dynamic data contained the information content necessary for accurate group discrimination. While the features did not directly represent key time points and associated variables like angles and moments, they were computed for every measurement without modification. Thus, they can be used for a wide range of biomechanical measurements.

The ability of the presented methodology to rank features showed that typically a combination of the different feature types leads to correct classification. Although, for example, regression features are not mandatory for gender classification, they add important discrimination information for shod/barefoot classification. Consequently, it is a good idea to compute all the proposed features. This is particularly true because the inherent feature selection of AdaBoost will reject features for classification if they do not contribute indicators critical to the classification.

A further aspect of the presented methodology is that the applied algorithms are not restricted to a two-class problem. Freund and Schapire (1997) have already shown that AdaBoost is capable of discriminating multiple classes. Thus, the most important effects of for instance wearing different running shoes could be identified using biomechanical measurements.

One limitation of the methodology is the required parameter setting for the extracted features. However, these parameters can be set by simple experimental evaluation once the methodology is implemented. Firstly, the number of necessary

DCT features needs to be determined. These can be set by observing the amount of energy that is contained in higher frequency components. Secondly, the required polynomial fit order has to be identified. Experimental evaluation of the classification rates using different orders straightforwardly reveals the optimal setting.

A further limitation of the methodology is that the exact mechanisms responsible for the discrimination were not revealed by the different classification tasks. Nevertheless, the methodology pinpointed those variables that were most relative to a certain differentiation. From this starting point, more detailed experiments may be conducted in order to unveil the relationship between those variables and the specific classification task.

3.5 Summary

A classification approach using generic features and AdaBoost was shown that provides an effective tool for identifying variables that allow subject or patient group discrimination. Besides high classification rates for gender and shod/barefoot running, the results also suggested that the mean hip abduction moment may be a very important indicator connected to the development of patellofemoral pain syndrome in runners.

3.6 Acknowledgements

The authors thank Drs. Willem Meeuwisse and Victor Lun of the Calgary Sport Medicine Centre for the injury assessment during the original data collection. Likewise, we thank Florian Hoenig (University of Erlangen) for many helpful discussions.

CHAPTER 4 CLASSIFICATION AND VISUALIZATION OF YOUNG-ELDERLY GAIT PATTERN DIFFERENCES VIA DIRECT PCA FEATURE EXTRACTION AND SVMS

4.1 Introduction

The automated recognition of gait patterns is of importance because of its potential applications in medical diagnostics, e.g. for the identification of at-risk gait in the elderly. In such clinical gait analyses, 3D positions of markers attached to the human body are typically measured to determine joint angles and range of motion.

Previous studies used pattern classification methods to differentiate gait patterns of young-elderly groups based on such kinematic variables (Wu et al., 2006) or the combination of kinematic and spatio-temporal variables (Wu et al., 2007) with classification rates of 89.6% and 91%, respectively. While these classification rates indicate the ability of pattern classification to differentiate the group gait patterns, a possible loss of information may have been introduced by the methods that were applied to the data. First, the calculation of the kinematic variables required to combine marker information. Therefore, the amount of spatial information was reduced. Second, only specific time points of the gait cycle (e.g. touch-down, toe-off) were considered in the evaluation. Thus, a substantial part of the available temporal information was discarded (Chau, 2001a).

It is suggested that more information is available for group differentiation if the 3D marker trajectories, which represent the complete available temporal information, are used for feature computation and classification. Such a method for the classification of gait has not been presented in the literature. It is postulated that higher group classification rates will be obtained when using such an approach.

Therefore, the purposes of this study were (a) to present a method for classifying gait pattern group differences using a more complete representation of the spatial and temporal information of the individual markers and (b) to compare the classification rates to previous studies using conventional classification features.

4.2 Methods

4.2.1 Data preparation

4.2.1.1 Collected data

Kinematic data was collected from 48 healthy female subjects (Tab. 4.1). The age of 24 subjects was between 55 and 70 years (elderly group), the age of the other 24 subjects was between 21 and 30 years (young group). All subjects gave informed written consent according to the guidelines of the University of Calgary's Conjoint Health Research Ethics Board, which approved the study.

All subjects were equipped with 37 reflective markers that were placed on head, trunk, arms, hands, legs and feet consistent with Vicon's (Oxford Metrics, Oxford, UK) Plug-In-Gait model (e.g. Orendurff et al., 2006; Buckley et al., 2009).

Before data collection, all subjects walked for five minutes on a treadmill in order to warm up and to select a comfortable walking speed. During data collection, the subjects walked for 80 seconds on the same treadmill at the self selected speed (Tab. 4.1). The walking speed remained constant for each subject throughout the data collection.

Tab. 4.1. Characteristics of the 48 subjects that were used for analysis purposes. Values in the first two rows are presented as mean (standard deviation). Values in the last two rows indicate the ranges of the different parameters.

	n	Age [y]	Height [m]	Mass [kg]	Treadmill speed [m/s]
Elderly group	24	59.9 (4.5)	1.61 (0.05)	68.8 (10.9)	1.24 (0.27)
Young group	24	25.3 (2.4)	1.66 (0.07)	67.2 (13.0)	1.53 (0.17)
Ranges					
Elderly group		[55.0;70.0]	[1.50;1.75]	[51.1;89.4]	[0.76;1.67]
Young group		[21.0;30.0]	[1.52;1.78]	[50.5;101.0]	[1.07;1.79]

Marker positions were recorded at 240 frames/second using a system of eight synchronized digital infrared high-speed cameras (Eagle and Hawk, Motion Analysis Corp., Santa Rosa, CA, USA). The coordinate system defined by the calibration of the camera system had the subjects walk along the y-axis (anterior-posterior direction). The x-axis was aligned with the medial-lateral, the z-axis with the vertical direction.

Of the 37 originally collected markers, 28 were at anatomical landmarks (Tab. 4.2). These markers were selected for the current study. The remaining nine markers were placed on arbitrary positions of specific segments and were not used for classification purposes.

The trajectories of individual markers were reconstructed using the software Eva Real-Time (EVaRT, Motion Analysis Corp., USA). The data was not filtered. Gaps shorter than 0.1 s in individual marker trajectories were reconstructed using either cubic interpolation or by determining the position of the missing marker from adjacent markers. Longer gaps were not present in the data. All subsequent analyses were performed using a custom MATLAB (version 7.6.0.324, The MathWorks Inc., Natick, MA, USA) software.

Tab. 4.2. Identifiers of the 28 markers that were used for classification purposes. The identifiers were in accordance with Vicon's plug-in-gait marker set.

Marker nr.	Marker identifier	Marker position
1	'LTOE'	Left toe
2	'LANK'	Left ankle
3	'LHEE'	Left heel
4	'LKNE'	Left knee
5	'RTOE'	Right toe
6	'RANK'	Right ankle
7	'RHEE'	Right heel
8	'RKNE'	Right knee
9	'RASIS'	Right anterior superior iliac spine
10	'LASIS'	Left anterior superior iliac spine
11	'RPSI'	Right posterior superior iliac spine
12	'LPSI'	Left posterior superior iliac spine
13	'STRN'	Sternum
14	'CLAV'	Clavicle
15	'C7'	7^{th} cervical vertebrae
16	'T10'	10^{th} thoracic vertebrae
17	'RSHO'	Right shoulder
18	'RELB'	Right elbow
19	'RWRA'	Right wrist thumb side
20	'RWRB'	Right wrist pinkie side
21	'LSHO'	Left shoulder
22	'LELB'	Left elbow
23	'LWRA'	Left wrist thumb side
24	'LWRB'	Left wrist pinkie side
25	'LFHEAD'	Left front head
26	'RFHEAD'	Right front head
27	'LBHEAD'	Left back head
28	'RBHEAD'	Right back head

4.2.1.2 Gait cycle extraction

For each of the 48 subjects, ten consecutive gait cycles were extracted from the collected walking gait data. For this purpose, gait phases without artifacts in the gait were used. Artifacts were defined as any measurements that deviated from a subject's automated (unconscious) gait pattern, for example when the subjects scratched themselves or moved their head in an unusual way.

The beginning of a new gait cycle was arbitrarily defined as the point in the time sequence when the left heel marker reached its lowest z-axis position. This point allowed unambiguous splitting of gait cycles.

The ten extracted gait cycles were then prepared for classification in three steps. First, they were individually time normalized. For this purpose, a normalization to 101 time steps from 0% to 100% was performed using a cubic spline interpolation (De Boor, 1978). Second, anthropometric differences of the subjects were eliminated by calculating the mean position of each of the 84 marker time sequences (28 markers in three axes each) for each subject and by subtracting them from the respective time sequence. Third, the classification was designated to use a mean gait cycle representation from each subject. For that reason, the mean of the ten consecutive gait cycles that were extracted was computed.

The 84 marker sequences were then concatenated into one movement pattern vector $\boldsymbol{m}_i$ per subject $i = 1,\ldots,48$, which was of dimension 8484 (84 marker sequences times 101 time steps).

The movement pattern vectors $\boldsymbol{m}_i$ were visualized (Fig. 4.1), where the mean movement pattern vector $\bar{\boldsymbol{m}}$ of all subjects is illustrated for four positions of the gait cycle. In each time point the marker positions were visualized in the x-z-plane (sagittal plane) on the left and the y-z-plane (frontal plane) on the right.

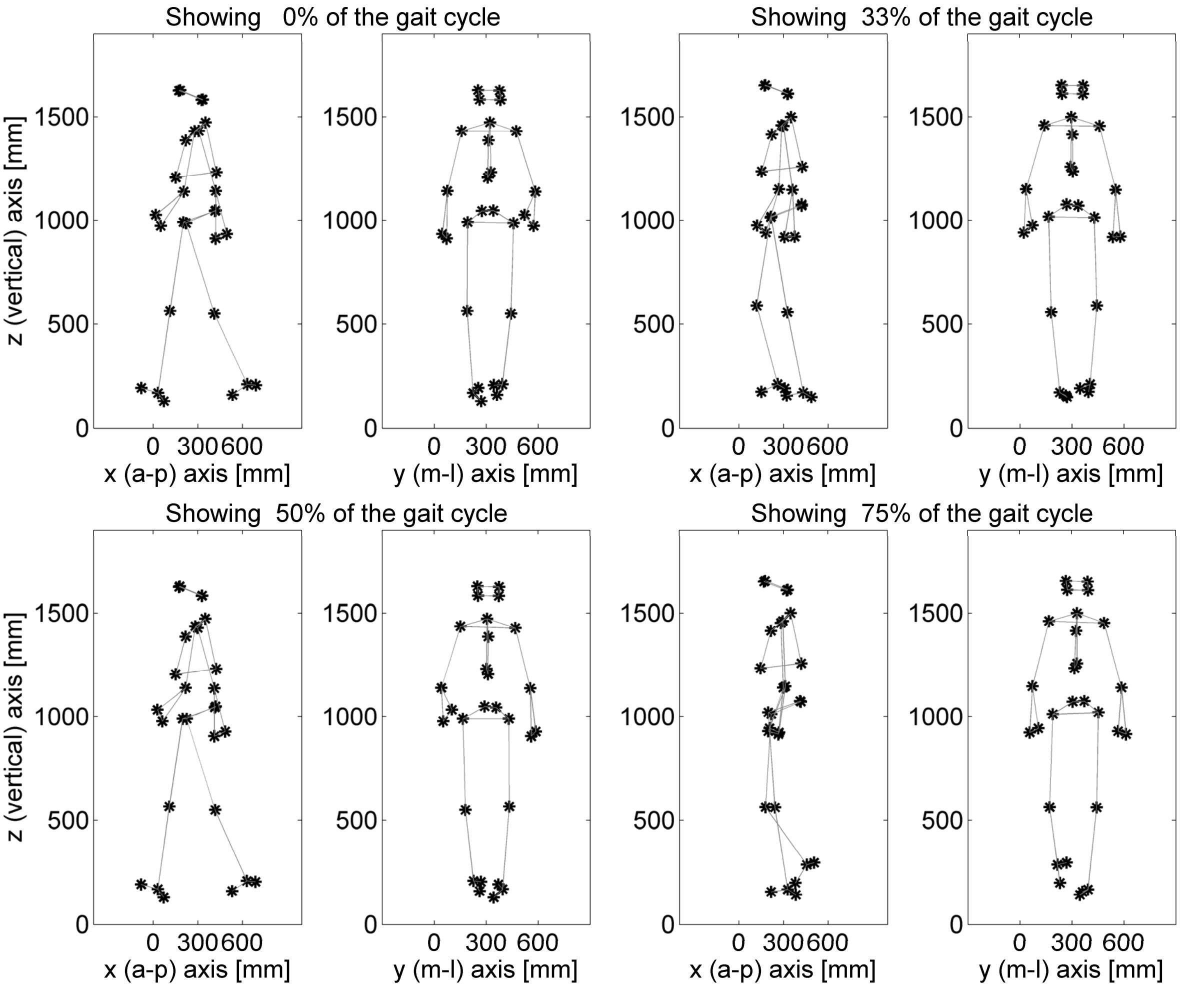

Fig. 4.1. Mean marker position of all subjects at different time points of the gait cycle. For each time point, the sagittal plane is shown on the left and the frontal plane on the right. Each star represents the position of one of the 28 markers that were used for classification.

4.2.2 Group classification algorithm

4.2.2.1 Feature extraction

The purpose of feature extraction was to retain as much of the spatial and temporal information of the movement patterns as possible. Therefore, a direct feature extraction from the movement patterns by principal component analysis (PCA, Fukunaga, 1990) was performed. The important characteristic of the PCA was that it conducted a transformation of the marker movement space that still incorporated all available information. Furthermore, the PCA feature representation is known to be suitable for classification (Theodoridis and Koutroumbas, 2009).

To perform the PCA, the movement pattern vectors $\boldsymbol{m}_i$ from all subjects were arranged in the data matrix **M**. As the number of individual samples (48) was smaller than the number of dimensions (8484), the small sample size PCA algorithm (Fukunaga, 1990) was used. In this algorithm the features for each pattern vector were computed by first removing the mean and then projecting it onto the eigenvectors of the sample correlation matrix $\mathbf{M}^t\mathbf{M}$. As a result of the PCA algorithm, the movement pattern of each subject could be represented as a linear combination of up to 48 principal movements $\boldsymbol{x}_i^d$, where

$d = 1,\ldots,48$ number of principal movement

$i = 1,\ldots,48$ number of runner.

The principal movements $\boldsymbol{x}_i^d$ were ordered according to the magnitude of the eigenvalues of the PCA, which corresponded to the amount of gait variability contained in the movement of all subjects. Therefore, the first few principal movements corresponded to the largest overall gait variability, and described the main variations in the movement over time.

The principal movements $\boldsymbol{x}_i^d$ were directly used as features for classification. For the evaluation of the classifier, an increasing number $d = 1,\ldots,48$ of principal movements $\boldsymbol{x}_i^d$ were used to evaluate the classification performance. This is a standard procedure for classification (Theodoridis and Koutroumbas, 2009), as every PCA component adds additional information for group classification according to overall data variability.

4.2.2.2 Group classification

For group differentiation, a support vector machine (SVM) classifier (Vapnik, 1998) was used. The SVM classifier (e.g. Begg et al., 2005; Wu and Wang, 2008) has, to the knowledge of the authors, previously not been applied to principal movement patterns of gait. Important characteristics of the SVM classifier for this project were (a) that it typically obtained high classification rates (Sapankevych and Sankar, 2009), (b) that the SVM implementation that was used (C-SVM, Chang and Lin, 2001) only possessed one free parameter that had to be set and (c) that the application of SVM with linear kernel allowed further analysis of the group differences with respect to spatial and temporal information of individual marker movement.

To obtain high group classification rates, the SVM had to find an optimal decision hyperplane (Fig. 4.2) that separated the principal movements of subjects from different groups with a maximal margin, i.e. the distance of the hyperplane to any principal movement was as large as possible.

The parameter that needed to be set for SVM classification was the cost parameter *C* (Schölkopf and Smola, 2002). It determined the tradeoff between the classification performance on the training set and the generalization ability, i.e. the ability of the classifier to correctly classify new samples. Since no general rule for setting the *C*-parameter existed (Chang and Lin, 2001; Vapnik, 1998), the classification results when using different settings for *C* were experimentally evaluated (e.g. Begg and Kamruzzaman, 2005). For the evaluation, a logarithmic range ($C = 10^n$, n = -3,-2.5,-2,…,3) was employed.

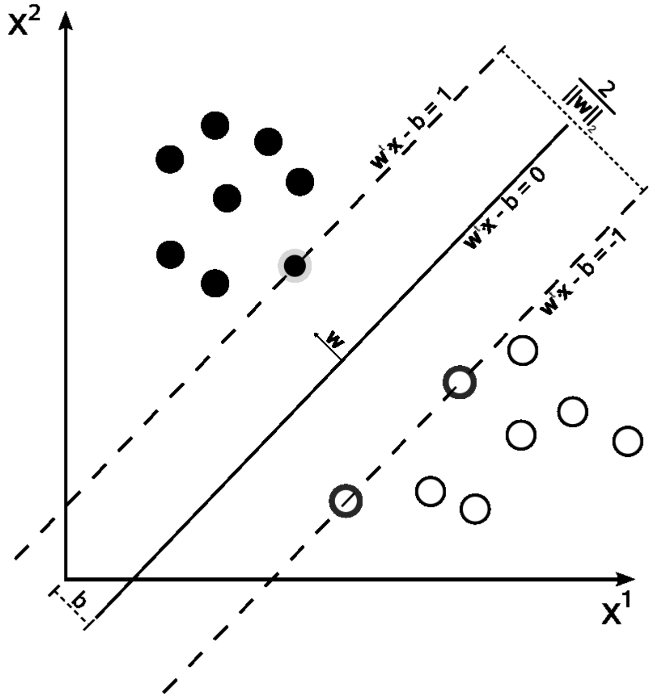

Fig. 4.2. An SVM example group classification for a two class problem. Class one is represented by white circles and labels -1 and class two by black circles and labels +1. The class representatives are shown for two feature dimensions x^1 and x^2. The decision hyperplane is represented by the normal vector $\boldsymbol{w}$ and the distance to the origin b.

The SVM operated by first subjecting the principal movements $\boldsymbol{x}_i^d$ to an implicit mapping to a higher dimensional space (Vapnik, 1998). For this purpose, different kernel functions (Schölkopf and Smola, 2002) were available. In the present study a linear kernel was chosen because it allowed functional analysis of the contribution of individual markers to group differences. The group difference visualization that was used for this purpose also allowed investigating what spatial and what temporal information was needed for group differentiation.

With the employed linear kernel, the hyperplane that separated the groups was parameterized by its normal vector $\boldsymbol{w}$ and by its distance to the origin b (Fig. 4.2). The vector $\boldsymbol{w}$ pointed in the direction of difference between the two groups on either side of the decision boundary (Fig. 4.2). Its length was defined by the mean distance to the individual group cluster centroids. As the vector $\boldsymbol{w}$ was an element of the principal movement space, it could be projected back onto the original marker movement space for further analysis of the group differences. For this purpose, a linear combination of the PCA eigenvectors using the components of $\boldsymbol{w}$ as weights was computed. This back projection of the vector $\boldsymbol{w}$ was called difference marker movement vector $\boldsymbol{m_w}$. It represented the spatial and temporal contribution of each individual marker movement to group differentiation. To show these individual contributions, the difference marker movement vector $\boldsymbol{m_w}$ was added (elderly group, labels +1) and subtracted (young group, labels -1) from the mean movement $\bar{\boldsymbol{m}}$ of all subjects that was shown in Fig. 4.1.

For evaluation of the classification rate, a leave-one-subject-out cross-validation was conducted with all trials from one subject being removed for classifier training. Then, the left out trials were classified and tested for correctness. This was repeated until each subject was left out once. The number of correctly classified trials divided by the total number of trials then gave the classification rate.

4.3 Results

The maximum classification rate of 95.8% was reached when using d = 36,…,39 principal movements (Fig. 4.3). In this case, two subjects that belonged to the elderly class were incorrectly classified during cross-validation.

The setting of the *C*-parameter did not affect the maximum classification rate or its locations. The mean classification rates over all d = 1,…,48 cross-validation runs for different *C*-parameters varied slightly (Tab. 4.3). The minimum and maximum mean classification rate were 79.1% for $C = 1000$ and 80.5% for $C = 0.1$, respectively.

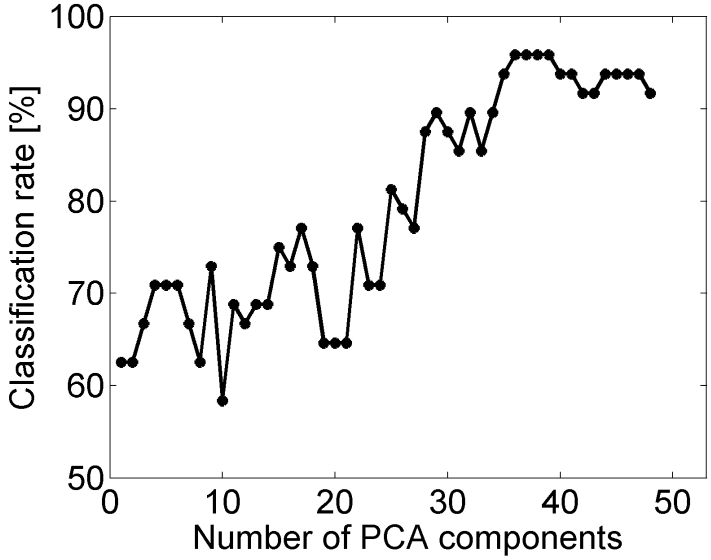

Fig. 4.3. Classification rate in percent when using d = 1,…,48 principal movement pattern features and C = 0.1.

Tab. 4.3. Mean classification rates for different C-parameter settings. The C-parameter was evaluated over a logarithmic range ($C = 10^n$, n = -3,…,3). The parameter n is given within the table. The mean classification rates were computed over all d = 1,…,48 cross-validation runs.

Parameter *n*	Mean classification rate [%]
-3.0	80.0
-2.5	80.0
-2.0	80.0
-1.5	80.1
-1.0	80.5
-0.5	79.7
0	79.5
0.5	79.2
1.0	79.1
1.5	79.3
2.0	79.2
2.5	79.2
3.0	79.1

The computation of the contribution of individual markers to group differentiation in position and time was performed at the point of maximum classification rate using 36 principal movement features (Fig. 4.3). The resulting difference information is illustrated for four time points of the gait cycle (Fig. 4.4). A video of the difference information for all time points of the gait cycle can be found at http://tinyurl.com/group-diff.

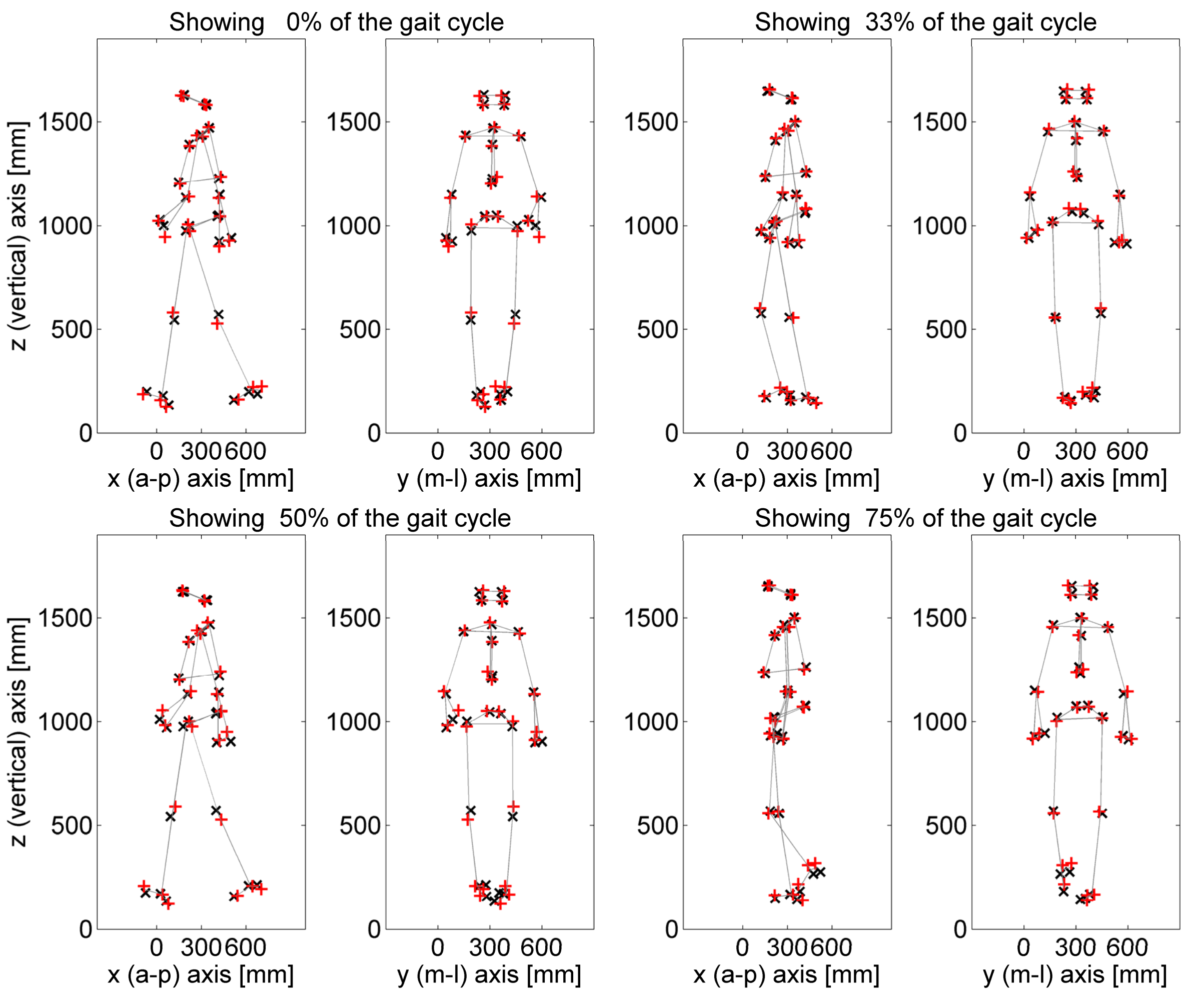

Fig. 4.4. Visualization of the contributions of individual markers to group differentiability at different time points of the gait cycle. The marker location differences in relation to the overall mean movement of all subjects (Fig. 4.1) is shown for the elderly group (black crosses) and the young group (red plusses).

4.4 Discussion

The results demonstrated that the proposed method was capable of obtaining higher classification rates compared to previous studies that differentiated young-elderly gait based on 3D marker data. The previous studies reported classification rates of 89.6% based on kinematic variables (Wu et al., 2006) and of 91.0% based on the combination of kinematic and spatio-temporal variables (Wu et al., 2007). Based on the proposed approach in the current paper, a classification rate of 95.8% (Fig. 4.3) could be obtained. This increased classification power is mainly attributed to the more complete availability of spatial and temporal information that was achieved by direct feature computation via PCA from the 3D marker data.

A direct PCA analysis of group marker information for gait classification had, to the knowledge of the authors, previously not been attempted. PCA was, however, applied to the analysis of the movement of individuals. Previous results showed that the gait patterns of individual subjects could be efficiently modeled using four principal components (Troje, 2002). In this light the number of principal movement patterns needed for a sufficient classification (36-39) in the present study may seem high. However, the current study did not focus on principal components of the movement of individual subjects, but on the movement differences between groups of subjects. Human gait comprises high inter-individual variance (Sadeghi et al., 2000). As opposed to the modeling of the movement of individuals, a higher number of principal movement patterns were therefore needed in the current study to efficiently generalize and classify the differences between groups.

Each of the individual principal movement patterns represented a combination of spatial and temporal information about the movement of the individuals. Both aspects contributed to group differentiation (Fig. 4.4). The illustration showed the differences in individual marker positions that led to classification. One example of the combination of the spatial and the temporal aspect to group differentiation could be seen by examining the knee markers. During the swing phase (Fig. 4.4, 33% and 75% gait cycle), the knee markers of both groups did not exhibit a large difference. However, during the double support phase (Fig. 4.4, 0% and 50% gait cycle), both knee markers exhibited a large difference. Thus, both the spatial and the temporal aspect of the movement were important for classification. The authors are not aware of a similar technique that allowed the analysis of individual marker contribution to group differentiation by taking both the position and time aspect into account.

For the age related example, previous findings could be reproduced by this analysis of individual marker contribution to group differentiation. Notable differences visible in Fig. 4.4 were the foot clearance and the stride length of young and elderly. In the representations of the swing phase (Fig. 4.4, 33% and 75% gait cycle), it could be seen that the young group had a higher position of the swinging foot than the elderly group. This observation of higher foot clearance in the young group was consistent with previous results (Begg et al., 2005). In the representations of the double support phase (Fig. 4.4, 0% and 50% gait cycle), it could be seen that the markers representing the feet were farther outwards for the young group than for the elderly group. This increase of stride width in gait of young subjects has also previously been reported, e.g. (Blanke and Hageman, 1989). Other previously reported positional and temporal differences in young-elderly gait could be reproduced as well. These were, for instance, an increased range of motion in the arm movement (Elble et al., 1991) and in the flexion-extension of the foot (Nigg et al., 1994). In the same manner, differences in individual body part movements could be observed by examining individual markers over time. In principle, the group difference representation in the original marker space also allowed further functional analyses by calculating kinematic variables.

Limitations of the proposed methodology existed in the necessity for a time normalization of the gait cycles due to the equal sample length requirements for the PCA. Information about the different walking speeds (Tab. 4.1) was therefore lost. Future work could include this information as an additional feature for classification.

As an additional algorithmic limitation, the *C*-parameter of the SVM had to be set correctly. However, this setting could easily be determined by an experimental parameter search. Even with a non-optimal *C*-parameter, the algorithm converged with high classification rate (Tab. 4.3).

When compared to more traditional approaches that used discrete kinematic variables at specific time points of the gait cycle for group differentiation, some further limitations existed. Given that the selection of variables and time points for the discrete approach was appropriate and functional, direct functional conclusions could be drawn about group differences. In the approach for the current paper, these conclusions were not as straightforward to draw, but required careful additional analysis of the reasons that led to group differentiation.

However, not to perform the classification on discrete functional variables at specific time points of the gait cycle also had several advantages. First, the selection of functional features for the classification procedures is usually based on prior knowledge of the researchers. The method presented here extracts the features based on a mathematical algorithm. The classification result therefore identifies group differences independently of prior knowledge. Second, the calculation of kinematic and kinetic variables typically requires assumptions (e.g. about the direction of joint axes), which are often difficult to validate. Third, the computation of these variables might lead to error amplification (e.g. when marker positions are affected by skin movement). Fourth, the process is often time consuming. Last, the incorporation of a more complete representation of the temporal information prevents discarding a substantial part of the available time dependent information (Chau, 2001a).

Since the presented methodology made no special assumptions, it could be applied for group classification tasks to any study involving marker measurements. Examples include analysis of pathological gait differences that are due to injuries, for medical pre-diagnosis of gait diseases and for evaluation of the outcome of treatment and rehabilitation.

4.5 Summary

The current study proposed a method for group classification that directly extracted spatial and temporal information from the 3D marker trajectories collected during human gait. Thus, this method did not require prior knowledge or assumptions, which are required when biomechanical features such as joint angles are determined in additional post processing steps. The classification using the SVM classifier yielded better group classification rates for a young-elderly group example than those reported in previous studies. The group discriminator could be visualized, which allowed identification of functional differences between the groups.

4.6 Acknowledgements

The authors thank Vincent von Tscharner for his time for many helpful discussions and Karelia Tecante for her help in data collection.

CHAPTER 5 EMBEDDED SURFACE CLASSIFICATION IN DIGITAL SPORTS

5.1 Introduction

The ability to perform accurate classification in real-time is a key factor for many applications. This is not only true when computationally powerful hardware is used. It is most often crucial in the restricted hardware environment of the power-efficient, highly mobile microprocessors used in embedded systems. Consider, for example, portable devices performing image classification or speech recognition. As (Hacker et al., 2006) showed, classification of the focus of attention of a user in interaction with a portable digital assistant (PDA) is possible with classification rates of up to 93%. The authors avail the signal of a built-in video camera and information from the speech signal to discern whether the user is trying to interact with the device or not.

The most important question is which of the complex algorithms known in pattern recognition can be used and implemented in the context of the restricted memory capacity and computational power of the employed microprocessors. Special considerations have to be made in order to adapt those algorithms to the specific hardware and classification task at hand. A lot of areas of engineering can benefit from the possibility of accurate classification in this restricted environment. Examples include, but are not limited to, automotive solutions, communications, industrial automation, speech recognition and medical care. In each of these fields, cheap and therefore mass producible systems that are highly portable can open up completely new ranges of applications.

We present an approach to accurate classification on a microcontroller that uses an example from the field of digital sports. It is quite common in sports related fields to use single-dimensional statistical analysis and model building techniques even for large, multi-dimensional datasets. (Lun et al., 2004) for example examined the relation between biomechanical variables of runners and the risk of running specific injuries. To facilitate this, they defined several injury classes and tried to identify significantly different parameters between those. While their work offers a lot of new insights, the underlying database is very complex so that important higher-dimensional coherences might not have been revealed. In the most recent years, pattern analysis concepts find their way into the field of digital sports, too. One application of pattern analysis methodology was shown in (von Tscharner and Goepfert, 2003). In this paper it is reported that electromyograph signals of muscle activity can be represented in pattern space using wavelet analysis. The authors furthermore demonstrate that the different activity patterns of males and females can be classified with a precision of more than 95%.

Our approach to guarantee accurate classification on the embedded system is to perform as much analysis as possible on computationally powerful PCs. This allows us to efficiently compare a lot of different approaches and select the one that is best suited for the classification task. Thereby, we keep the hardware restrictions in mind during every step of the pattern recognition chain. We identify the classifier that is best suited for the implementation on the specific microcontroller that is used, performing only the final verifications on the embedded hardware. For this presentation, we focus on the application of these concepts on the adidas_1 running shoe, which is the first shoe ever that features an embedded system. This shoe is built to adapt to various running conditions like the prevailing surface situation. A precise classification of these conditions is of course mandatory to guarantee this functionality. To facilitate this, the step signal of the runner is continually measured and processed by the embedded microcontroller. A detailed description of the adidas_1, its functionality and embedded system hardware can be found in Sections 5.2.1 and 5.2.2, in (DiBenedetto et al., 2004) and in (DiBenedetto et al., 2005). In the rest of Section 5.2 we describe the analysis methods that lead to accurate, real-time surface classification, including:

- the preprocessing steps that are a prerequisite to later obtain features that can be reliably computed (Section 5.2.4);
- the choice of discriminative features, which dependably represent the signal information while still being efficiently calculable (Section 5.2.6);
- a detailed description of the examined classifiers (Section 5.2.7).

The choice and parameterization of each of these factors of the classification chain is essential for ensuring optimal results. In Section 5.3 we present the conducted experiments and their results. In our summary in Section 5.4 we will show that, while the specific question of surface classification is solved, the employed methods are general in nature so that they can contribute to a wide area of applications featuring embedded systems. The presented example for a classification system has recently been implemented in the current version of the adidas_1 running shoe, which is commercially available. It is significantly contributing to the shoe's functionality and thereby offering runners an ideal adaptation during each phase of their run.

Signal classification techniques have long been successfully applied to radio signals (Schmidt, 1986), images (Haralick et al., 1973) and speech signals (Furui, 2004). Recently, examples of classification systems implemented on embedded systems have also been published, for example in (Englehart and Hudgins, 2003; Wolf et al., 2002). Pattern recognition algorithms are lately also applied in sports related problems (Assfalg et al., 2002; von Tscharner and Goepfert, 2003). However, to our knowledge, we are the first group to use these established techniques in order to classify a step signal on an embedded system in the context of sports.

5.2 Materials and methods

5.2.1 The adidas_1 running shoe

The adidas_1 is a running shoe, which possesses a built-in 8-bit microcontroller, a sensor for heel compression measurement and a motor for cushioning adaptation. This shoe is designed for avid runners, and is constantly adjusting itself to the running situation. In this presentation, we will focus on the classification of the surface that the athlete is running on. While other parameters are also important, it was the first goal of the ongoing research to develop an algorithm that is well suited for surface classification from the sensor signal alone. The general demand to establish constant cushioning when a change of running surface takes place and all other running conditions remain constant is to have

- a soft shoe on hard surfaces (e.g. asphalt, concrete) and
- a hard shoe on soft surfaces (e.g. grass, trail).

The shoe therefore provides higher cushioning on hard surfaces, and we aim at a good classification of any hard surface that the runner is on. Similarly, if we detect a change of surface, i.e. if the surface is not hard anymore, we adapt the shoe to this softer surface condition, making it harder and stiffer in order to prevent injuries that are attributed to a lack of control. We therefore decided not to use a continuous scale from the hardest possible surface to the softest, but rather to employ a hard decision threshold either for the hard or soft surface condition. This can thus also be regarded as a decision for either a ‘control’ or a ‘cushioning’ condition and a corresponding complete adaptation of the shoe. This automatic adaptation ideally takes into account the athlete's weight, speed, fatigue level and furthermore the current surface condition, elevation profile and shoe condition.

To facilitate this adaptation, the shoe features a cushioning element, whose ability to give way in vertical direction (in this chapter defined as z-axis) can be regulated by a motor-driven cable system. The cushioning element is depicted in Fig. 5.1. The regulating cable is visible in the z-axis X-ray image of the adidas_1 in Fig. 5.2. It is running from the motor through the middle of the cushioning element to its opposite end and is fixated there. The motor shown in Fig. 5.1 can adjust the attenuation setting by turning a screw, which lengthens or shortens the cable. When the cable is shortened, the cushioning element is tensed and compresses very little when external forces are applied. When the cable is longer, it allows the cushioning element to compress further by giving it more room to expand in the x-axis direction (forward-backward direction), effectively making the shoe softer. Changes to the softness setting are gradual. The attenuation setting from one extreme to the other is made in 15 increments. A decision for the current surface is made after every fourth step, taking the three preceding steps and the actual step into account by a majority vote. In the case of a tie, no adaptation is made. This is done to maintain the cushioning adaptation mechanism in the case that the runner takes only one or two single steps on a different surface and to save battery power. Thus, to go from the softest setting to the hardest and vice versa, 60 steps of the runner are required. We did not opt for an instantaneous change from one extreme to another once a definite surface change is detected, in order to once again save battery power. A complete change of the cushioning setting from one extreme to another is quite energy consuming. It is more economical to change the setting in small increments. This saves a lot of battery power if the runner only changes surface for a small number of steps, e.g. when running over a small stretch of grass while being mainly on a hard sidewalk surface. Using this approach, we can ensure that the battery (see also next Section 5.2.2) holds for the complete life-time of a running shoe, which is about 100 h. For more details on the shoe design the reader is referred to (DiBenedetto et al., 2004) and (DiBenedetto et al., 2005).

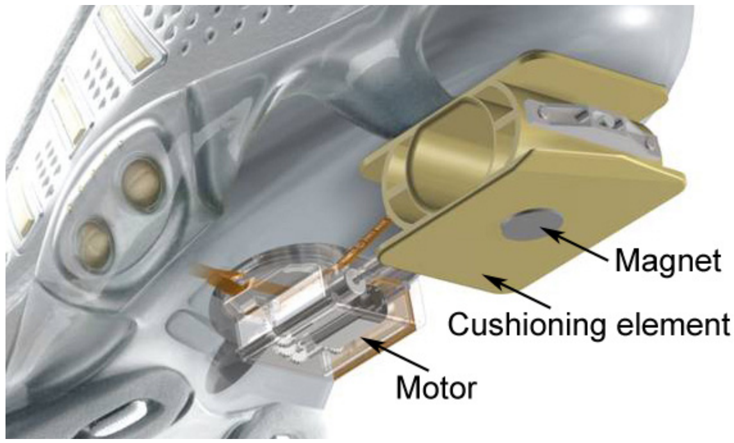

Fig. 5.1. A view of the adidas_1 shoe, depicting the cushioning element and motor unit. The indicated magnet induces a magnetic field for compression measurement.

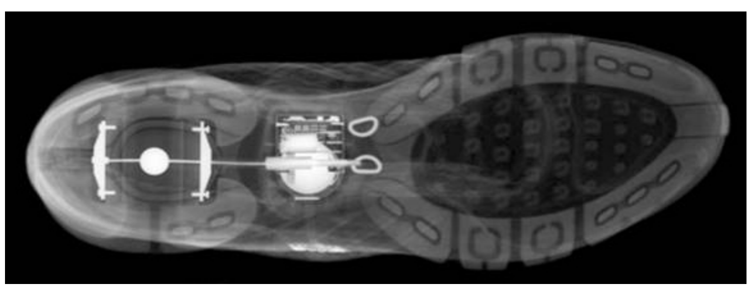

Fig. 5.2. X-ray image of the adidas_1 in z-axis direction. Motor unit, regulating cable and the magnet are clearly visible.

5.2.2 Embedded system hardware

The compression measurement of the adidas_1 shoe is made by a Hall sensor that is mounted at the top of the cushioning element. It detects the magnetic field strength induced by a small magnet, see Fig. 5.1, and can be sampled with a rate f_s of up to 1 kHz. The sensor- magnet distance d_m can then be computed from the magnetic field strength with an accuracy of ±0.1 mm. A decision whether the attenuation of the shoe has to be adapted is made based on the measured sensor data, see Section 5.2.3.

The sensor-magnet distance is sampled by the built-in microprocessor that is mounted on a flexible circuit board on the motor element. Currently, a Cypress Semiconductor Corporation controller CY8C21634 is used. However, the methodology that is presented below does not make any special requirements to the employed Microprocessor. The CY8C21634 possesses a clock speed f_{clock} of up to 24 MHz, 512 Bytes of SRAM and 8 kByte flash program store. Additional on-chip system resources include internal oscillators, control and communication interfaces and other highly configurable I/O options. The controller is designed with a standard Harvard architecture and focuses on low power consumption. The whole system is powered by a small 3 V coin cell, which is replaceable and lasts for the normal life-time of a shoe. The CY8C21634 and similar microprocessors are employed in a wide range of embedded applications. Examples include automotive solutions and consumer products like handhelds and digital cameras.

5.2.3 Sensor data

In order to get the data needed for the analysis, there is a special prototype system equipped with a data collection interface. The data from the magnet sensor is stored with a 256 kByte EEPROM array and is evaluated offline in a later stage. An example running signal is depicted in Fig. 5.3 with the sensor-magnet distance d_m plotted against time t. During the time where the shoe is in the air, the measured signal consists mainly of noise. In contrast, the heel (de-)compression phases of four steps can be distinguished. This measured signal is the basis for the surface classification experiments in Section 5.3.

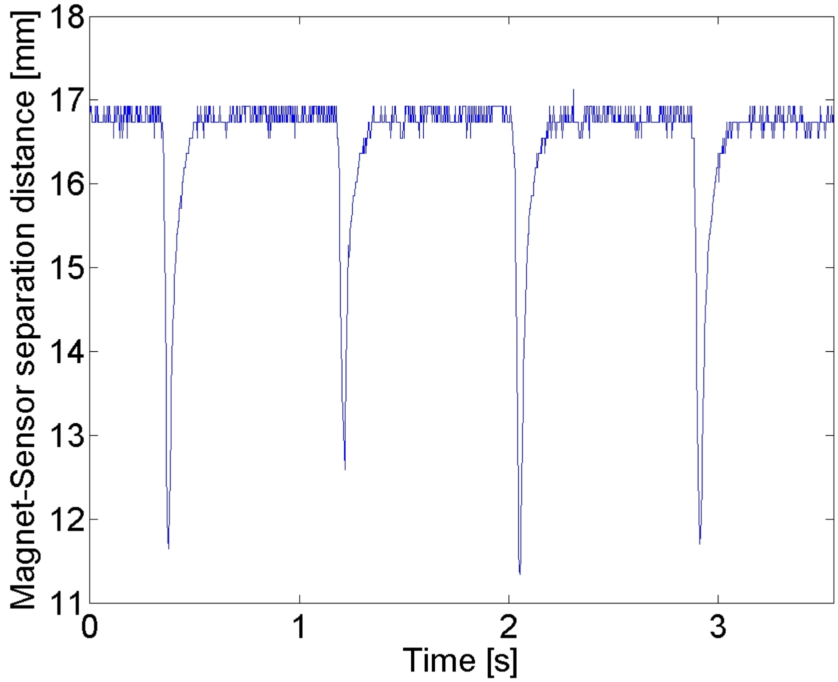

Fig. 5.3. Data example with 4 step maxima shown.

Because the signal from the Hall sensor consists mostly of noise while the foot is in the air, no relevant information for the cushioning adaptation can be gained. Therefore the sensor system and microcontroller are powered down for 120 ms after registering a compression maximum. Energy consumption during this period is very low, thus the system is saving battery power again. This phase is short enough to ensure that no step is missed when running normally.

5.2.4 Preprocessing

First of all, we have to extract the specific events that need to be classified in a reliable way. In this context, the events correspond to individual steps, which have to be found in the signal. All our features that we will present in Section 5.2.6 are based on exact identification of these single steps. To facilitate this identification, we first establish a baseline sensor-magnet distance value $d_{m,base}$. This value corresponds to the sensor-magnet distance when the shoe is in the air between steps. It can be reasonably assumed that it is the most frequently occurring value in the data. Next, all sample values that belong to a compressed state are detected. Initial experiments substantiated that compressed states occur when the sample values are below a sensor-magnet distance threshold $d_{m,thres} = d_{m,base} - 1.5 \cdot \sigma_{data}$, where σ_{data} is the overall standard deviation of a dataset.

We define the start and end of the compression phase as those points in the compression states where the distance from $d_{m,base}$ drops below three sample units, which corresponds to 0.7 mm. By using this approach, all steps could be identified in the datasets. This was confirmed by manually extracting 449 steps in 6 datasets and comparing the manual and automatic approaches. The results were completely identical. This result proved that our step detection algorithm provides reliable input for feature computation.

The presented algorithm is rather straightforward, which is a main design criterion for all processing steps for the microcontroller implementation. However, in this case no tradeoff had to be made between complexity and accurateness.

5.2.5 Labeling

In order to learn the necessary parameters for class separation, we needed information about the class membership of the samples. We therefore implemented a graphical user interface for data labeling. The interface is general and can be used for many different labeling tasks. Each event that has to be later classified is assigned to one of the classes manually because we believe this approach to be superior to an automatic labeling. Manual labeling was quite efficiently possible because we could batch label sequences of steps. This is due to the fact that a lot of consecutive steps are made on the same surface when running, i.e. the surface does not change at each step. We were therefore able to label from a start step to an end step of a sequence, assigning all intermediate steps to the same class. Due to the design of our data collection (see Section 5.3.1), these sequences were easily identifiable. This is because we knew what surface the runners were running on. In consequence, we could store the different surface data in separate files. Thus, data labeling was consistently and efficiently possible. The user interface with a sequence of steps labeled as belonging to the soft surface class is depicted in Fig. 5.4. A labeling interface like this can easily be programmed for a wide range of classification tasks and does not decrease the generality of the approach.

5.2.6 Feature computation

A set of features that can be used for microprocessor classification has to fulfill two main design criteria. It has to represent the sensor input information consistently while being computationally cheap as it has to be computed on the controller. The choice of features is critical, and "has to be performed for each specific problem to decide which feature of which type one should use" (Ohanian and Dubes, 1992). We therefore manually selected a feature set that is very specific to our task of running surface classification. The selected features should contain the information of the step signal as good as possible. We accordingly computed them such that the important properties of steps are well represented. Our experiments indicated that these features were sufficient because we noticed no improvement using any other imaginable feature. The features are listed in Tab. 5.1, where SD abbreviates standard deviation.

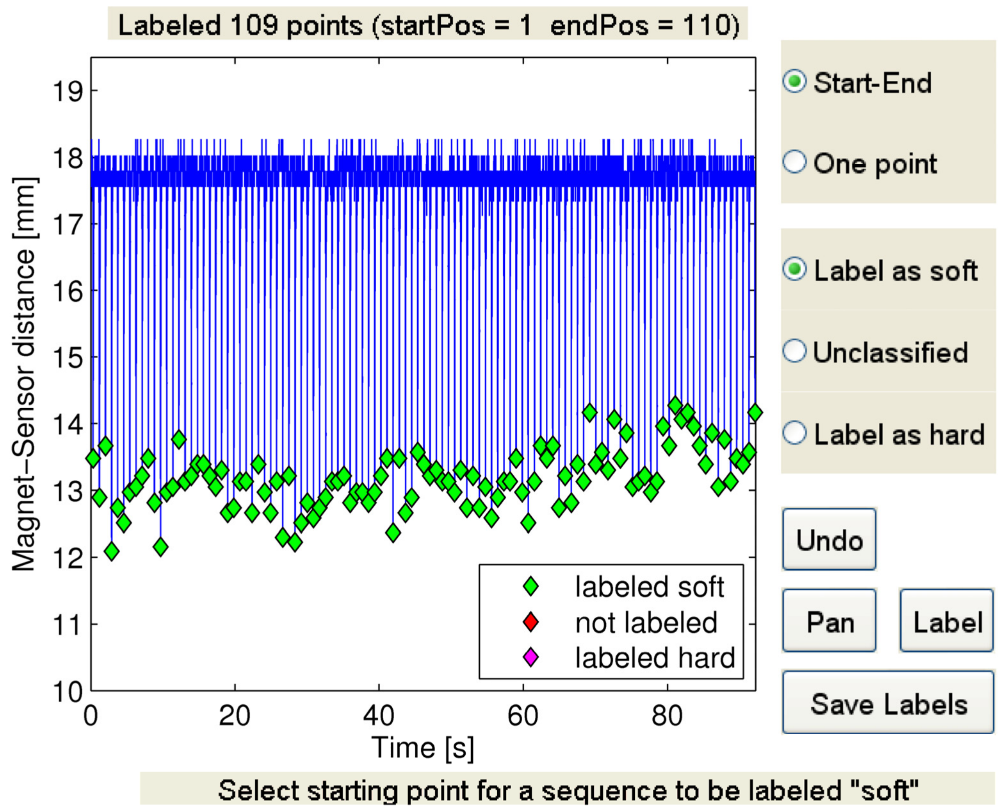

Fig. 5.4. The graphical user interface used for data labeling.

Features 1-11 are calculated on one step alone, with the exception of feature 3, which is computed on two consecutive steps. Features 12-19 are computed on the N preceding steps. The standard deviations $\sigma_{N,f}$, where $f = 11, \ldots, 19$, are computed as an unbiased estimator (Fukunaga, 1990) according to

$$\sigma_N = \left(\frac{1}{N-1} \sum_{k=1}^{N} (x_k - \overline{x})^2 \right)^{\frac{1}{2}}. \tag{5.1}$$

The unbiased estimator of the standard deviation has been chosen because $\sigma_{N,f}$ is computed for a sample drawn from a larger population in our case. Fig. 5.5 additionally illustrates features 1-10.

The obvious redundancy contained in the extracted features is volitional. It was a goal from the start to use only a subset of the given features to reduce complexity further, thereby using only features with small or no mutual dependence. We will explain our choice for the feature subset selection algorithm in Section 5.2.8.

Tab. 5.1. Overview of the features used for classification. Step compression and decompression refer to the respective phases where the shoe gets compressed and decompressed during heel strike.

Feature number	Feature description
1	Step compression first order least-squares fit
2	Step decompression first order least-squares fit
3	Time between step compression maxima points
4	Time from step compression maximum to step end
5	Time from step start to step compression maximum
6	Step curve area approximation by Trapezoid method
7	Time from step start to step end
8	Step mean value
9	Step median value
10	Step compression maximum value
11	SD of the values contained in one step
12	SD of the step minima (feature 10)
13	SD of the step means (feature 8)
14	SD of the step standard deviation (feature 11)
15	SD of the step duration (feature 7)
16	SD of the step area (feature 6)
17	SD of the time between steps (feature 3)
18	SD of the time to peak (feature 5)
19	SD of the time from peak (feature 4)

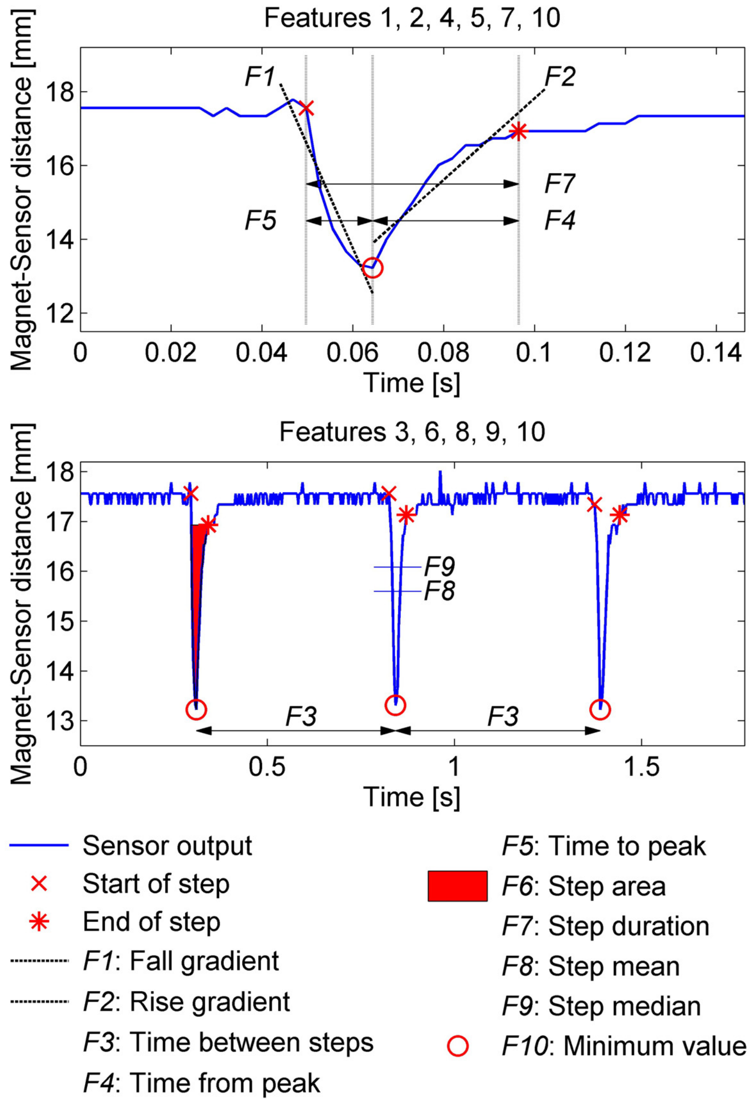

Fig. 5.5. Depiction of the step signal classification features 1-10.

Every single feature can be computed in real-time on the employed microprocessor (Section 5.2.2). In order to substantiate our other claim that the features dependably represent the signal information, we used them in a different classification task. In (Eskofier et al., 2008b), we report on work on running fatigue classification where we successfully applied the same step signal feature set. We also performed experiments that showed that our feature set even outperformed a computationally more demanding feature set successfully applied in biosignal classification. While our set could achieve classification rates of 73.9% in a comparative experiment, only 67.4% were achieved using the more complex feature set.

5.2.7 Classifiers

For our intended goal of embedded system classification we focused on classifiers that could be implemented in computationally efficient manner. Once again, this decision was motivated by the hardware limitations presented by the embedded system described in Section 5.2.2. Our approach is to experimentally compare a set of classifiers. (Duda et al., 2001) state that if one algorithm is outperforming another one in a particular situation, then this is a consequence of its fit to the particular pattern recognition problem, and not of the general superiority of the algorithm. We therefore want to find a fit to the problem by evaluating each particular classifier's performance. Of course there exists a large set of available classifiers, which we can not all individually test, so we reasonably selected such classifiers that are quite well known and widely discussed in the literature. The existence of reference implementations to compare the classifier performance experimentally further motivated our selection. Our choices included:

- Naive Bayes (NB), which is described and experimentally evaluated for example in (Duda and Hart, 1973; Langley et al., 1992).
- Neural Networks (NNet) with one hidden layer and varying number of hidden nodes (Duda et al., 2001; Specht, 1990).
- Nearest Neighbor (NNeigh) classifiers with different number of neighbors *k*, which has seen a lot of applications (Cover and Hart, 1967; Lee, 1991).
- Support Vector Machines (SVM) using kernels of low complexity (Duda et al., 2001; Vapnik, 1998).

- AdaBoost.M1 (Freund and Schapire, 1996) with decision stumps (Schapire et al., 1998) as weak classifiers and varying number of iterations N_{it}.
- Rule-based (Duda et al., 2001) approaches like PART (Frank and Witten, 1998).
- Linear Discriminant Analysis (LDA), see (Fisher, 1936) and (Duda et al., 2001).

In order to test these classifiers, we could efficiently use the WEKA toolbox, (Witten and Frank, 2005). This toolbox allowed us to compare all different approaches on powerful PC hardware in order to identify the algorithm that is best suited for the microcontroller implementation. Our experiments (see Section 5.3.2) proved that in our case LDA classification yielded comparable classification rates to other, more complex approaches while being the only approach meeting the real-time criterion. Because of the restricted hardware environment, we therefore decided to train a computationally cheap linear polynomial classifier using LDA. While the theory for other approaches will be omitted here and can be found in the according references, we will give a brief overview of Linear Discriminant theory for the sake of self-sufficiency. LDA classification uses the statistical properties of features, and furthermore provides rather simple linear decision surfaces even in high-dimensional spaces. To accomplish this, LDA transforms the feature space in a way that

- intraclass variation is minimized, i.e. the features of the same class are as densely packed as possible and
- interclass variation is maximized, i.e. distinct classes are as far apart from each other as possible.

For optimal classification, each point $\boldsymbol{x}$ in the d-dimensional feature space gets assigned to the class ω_i, where i denotes the class index, so that the posterior probability $P(\omega_i|\boldsymbol{x})$ is maximized. In our case, we only consider two classes ω_1

and ω_2. However, classification can easily be extended to multiple classes as well. According to LDA theory, the equation for the optimal decision boundary between two classes (see Duda et al., 2001, pp. 117-121) is

$$\boldsymbol{w}^t \boldsymbol{x} + w_0 = 0 \tag{5.2}$$

where

$$w_0 = -\frac{1}{2}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)^t \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) + \ln\frac{\mathrm{P}(\omega_1)}{\mathrm{P}(\omega_2)} \tag{5.3}$$

is an additive constant and

$$\boldsymbol{w} = (w_1, w_2, \ldots, w_d)^t = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \tag{5.4}$$

are coefficients for each of the d features used. Here, $\boldsymbol{\mu}_i$ are the d-component class specific mean vectors and $\boldsymbol{\Sigma}$ is the $d \times d$ covariance matrix that is identical for both classes but otherwise arbitrary. $\mathrm{P}(\omega_i)$ denotes the prior probability of the class ω_i. The above equations hold only if the class specific densities $\mathrm{p}(\boldsymbol{x}|\omega_i)$ can be assumed to be multivariate normal distributions ($\mathrm{p}(\boldsymbol{x}|\omega_i) \sim \mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$). In Section 3.1 we show that this assumption is justified.

In summary, the decision rule for two classes ω_1 and ω_2 becomes

$$\boldsymbol{w}^t \boldsymbol{x} + w_0 = \begin{cases} > 0 \rightarrow \text{decide for } \omega_1 \\ < 0 \rightarrow \text{decide for } \omega_2 \end{cases} \tag{5.5}$$

which can be straightforwardly implemented even on a microprocessor. In the case of equality, i.e. $\boldsymbol{w}^t\boldsymbol{x} + w_0 = 0$, an engineering decision for one of the classes has to be made.

5.2.8 Feature selection methods

Feature selection means identifying a feature subset that delivers good classification rate while reducing the complexity of the overall process. The computation of the 19 features given in Tab. 5.1 on the microprocessor would be too time-consuming. Furthermore, the computation would require storing a lot of Hall sensor sample values, which is not possible due to memory constraints. Thus, we implemented a method to select the best subset from the original features. This reduction not necessarily decreases the overall classification rate, by deselecting detrimental features the result can even improve. A widely used method for selection or reduction of features is the principal component analysis (PCA), see (Duda et al., 2001). It identifies the major axes of variance within the feature space by a Karhunen-Loéve transform. Axes of low variance contribute less to the discriminative ability of the features and can thus be neglected, thereby reducing the feature space. In our case, however, the application of PCA has a major drawback. PCA needs all the input feature values in order to come up with a reduced set of features for classification. This means that the original set of 19 features would have to be computed on the embedded system before it is reweighed by the PCA coefficients. We cannot compute all these 19 features due to the real-time requirement. This means that we have to perform the feature reduction on the original set of 19 features, thus deciding for a subset of features that can directly be computed from our input signal and meets the real-time requirements. However, as we wanted to show whether our feature selection as described below is significantly inferior to feature reduction using PCA, we compared both methods in our experimental chapter (see Section 5.3.3).

For our proposed selection algorithm the obvious criterion for choosing the best subset is the overall classification rate for a given problem. This result is computed via leave-one out cross-validation to prevent any overfitting effects. One example is presented in the experiments Section 5.3. In this example, we collected data from 24 runners and performed leave-one-runner-out cross-validation and used LDA training and classification to compute the rates.

Our algorithm follows the principles of a beam search as proposed by (Bisiani, 1987). During initialization, we use all combinations of two features for training and evaluation of the LDA classifier. A total of $\binom{n_f}{2}$ combinations is tested where n_f is the number of features. The overall classification rates are stored for every feature pair. Subsequently, only a predefined (e.g. 20) number of feature pairs delivering the best results are promoted to the next algorithm step. In this step, the best pairs are combined with each of the remaining features, thus leading to feature triples. This process is iterated, in every iteration the overall classification results are computed, the best combinations are kept and then again combined with the remaining features. For each iteration, a subset is thus identified that delivers the best result amongst those combinations remaining in the pruned search space. While the beam search does not guarantee that the optimal solution is found, it is a very cost-effective search method and ensures a good tradeoff between computational complexity and classification rate. Moreover, we could show that in our case the optimal solution and the one identified by the beam search are identical, see Section 5.3.

An important effect of the feature reduction approach is that it gives a very good overview of classification rates for different feature subset sizes. If the hardware framework is not completely specified, a system designer can easily decide what classification rate is necessary for the particular application and give an estimate on computational complexity and thus the required hardware.

5.3 Experiments

In the following, the experimental evaluation of the proposed surface classifier is presented. The important framework requirements were that the algorithm works for

- different runners (w.r.t. height, weight, running style, training level),
- different shoe sizes and
- all shoe settings (e.g. hard, medium, soft).

We will show that the system for surface classification that was developed works for these conditions. Additionally, the design considerations that specifically aim at meeting the hardware architecture restrictions are given below.

5.3.1 Collected data

In order to get a sufficient random sample for the subsequent classification experiments, a test course was selected where the desired surface conditions were present. The test course is located on the campus of the Faculty of Engineering of the University Erlangen-Nuremberg. It is depicted in Fig. 5.6. All runners were asked to run 12 sections of about 150 m each. The test was divided into 2 parts of 6 sections. The first 6 and the second 6 sections were each made with a manually chosen shoe setting. No automatic cushioning adjustment was made to make sure that signal changes only derive from surface or speed changes. The shoe setting was only changed after the first part in order to have data generated with varying cushioning setting, then the running procedure from that part was repeated. The 6 sections of one part were

- two runs on soft surface (grass) with constant speed;
- two runs on hard surface (asphalt) with constant speed;

- one run on changing surface, starting on grass, then switching to asphalt, and finally running on grass again, all with constant speed;
- one run on hard surface with a change in running speed, starting with the same constant speed from the previous sections and accelerating to a fast jog after the first half of the distance.

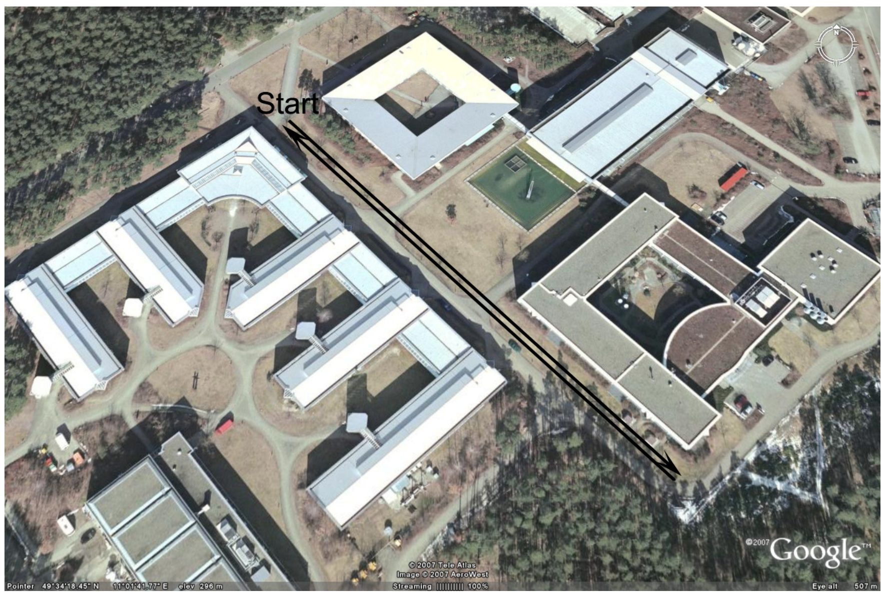

Fig. 5.6. Aerial view of the test course that was used for data collection.

Each participant was asked to run normally with a comfortable but constant speed for the first 5 sections. Shoe setting, time information and an athlete profile (weight, height, training frequency) was noted for every runner. In addition to the shoe signal, a Polar RS800 system with foot pod was used to get speed and step frequency information.

Altogether, 24 test runners participated in this data collection. Shoes with sizes 7, 9 and 11 were used for those experiments. A total of 106 datasets with different shoe cushioning settings was collected for the subsequent experiments. Steps were extracted using the automatic procedure described in Section 5.2.4. Tab. 5.2 shows the number of valid steps for each of the 24 test runners that were used for the classification experiments. They amount to a total of 22,910 single steps with a fraction of 50.6% on soft surface. The data was labeled as belonging to soft or hard surface using the GUI described in Section 5.2.5. The distribution of each of the 19 features for the two classes has been tested for normal distribution with a χ^2-test (e.g. Fukunaga, 1990, Chapter 3) after labeling. As a result we could say that the null hypothesis of normal distribution for both classes and each feature is true at a 95% significance level.

Tab. 5.2. Individual identifiers for the 24 test runners with number of valid steps given for each of them. Shoe size of each participant is given in brackets.

ABo (7) 1307	HH (9) 936	MW (11) 1165
ABr (9) 1152	JM (11) 541	RB (11) 670
AM (11) 1338	JP (9) 1013	RS (11) 384
BD (11) 781	KH (11) 903	SK (7) 1273
BE (11) 1206	KR (9) 1326	SW (9) 1240
CD (11) 911	MA (11) 898	TS (11) 612
DE (11) 1121	MP (9) 914	TT (11) 791
EK (7) 961	MS (9) 627	VD (11) 840

5.3.2 Classifier selection

In order to substantiate our choice of classifier, we tested the performance of the algorithms given in Section 5.2.7 on our 19-feature set. We used leave-one-runner-out cross-validation to compute the results that are summarized in Tab. 5.3. Algorithm settings were as follows. For the NNets, we used one hidden layer and 12 hidden nodes in the layer. Nearest Neighbor classification was performed with $k = 1, 3, 5, 11$ nearest neighbors. For the SVM evaluation, we chose linear polynomial kernels, as more complex kernels could not be implemented on the embedded system. AdaBoost.M1 was tested with decision stumps as weak classifiers. The number of training iterations, N_{it}, was set to 10, 30 and 50. PART was applied with at least 2 instances per rule (212 rules were trained in total). LDA and NB were tested as described in Section 5.2.7. As can be seen from Tab. 5.3, the results for LDA classification yielded comparable classification rates to most other algorithms that were tested. Nearest Neighbor classifiers were the only ones that performed significantly better. This indicates the existence of subclusters in the high-dimensional feature space, which can not be correctly classified using linearly-based decision boundaries. Despite this improved performance, in the context of the very restricted memory capacity of only 512 Bytes of the currently employed CY8C21634 microcontroller, NNeigh methods would be impossible to implement. We would need to store and compare with too many single data points or cluster centroids than is feasible. As we will point out in Section 3.3, we already used 98% of the available microcontroller program memory with our current approach using LDA and three features. All the same, we will certainly reconsider the choice of microprocessor in future product generations in the light of these results.

Three other algorithms (AdaBoost.M1 with N_{it} = 50 iterations, NNet and PART) outperform LDA less significantly. Our decision not to implement them on the microcontroller was made for computational reasons, too. We will briefly discuss them here. The 50 decision stumps trained by AdaBoost require a more complex decision system and more memory than is currently available. Furthermore, the number of required comparisons is undetermined, leading to a variable decision time. The constant number of operations required for LDA classification is preferable in our case. Even for a simple Neural Net as tested in our case, we would have a lot more multiplications (240 for the described Neural Network versus 19 for LDA). More importantly, we have to evaluate the sigmoid function or a comparable nonlinear function. These facts inhibit a NNet implementation on the employed embedded system. PART generates 212 rules with a total of 1305 possible comparisons, which is also impossible to implement on the embedded system that we utilize, given the fact that the available memory is already used with a much simpler solution. We did, however, compare all these results for the different classifiers in order to get good evidence of the performance of our proposed compromise of LDA classification. For other data and framework conditions, we expect that other solutions are more favorable. As we already stated in Section 5.2.7, is the experimental comparison of different solutions vital for a profound implementation decision.

Tab. 5.3. Cross-validated results for different classifiers computed on the complete set of 19 features.

Classifier	Classification rate (%)
NB	70.2
AdaBoost, N_{it} = 10	73.6
AdaBoost, N_{it} = 30	75.3
AdaBoost, N_{it} = 50	76.0
SVM	75.4
LDA	75.5
NNet	77.9
PART	78.4
NNeigh, k = 1	83.3
NNeigh, k = 3	84.9
NNeigh, k = 5	84.5
NNeigh, k = 11	83.6

5.3.3 Feature selection results

The results of the feature selection algorithm described in Section 5.2.8 are given in Tab. 5.4 (see Tab. 5.1 for details on the features). Only the combinations that perform best are shown. For this evaluation, we used the fact that the classification of single steps can be improved when additionally taking a context of preceding steps into account. In this case, a short context of three steps was used by casting a majority vote over the single decisions. In the implementation for the final product solution, a longer context can be used, which leads to even better classification results (see Section 5.3.6). We finally selected the feature triple 1, 12 and 17 for the implementation on the microcontroller for three reasons. First, it is the best three-feature combination and outperforms the two-feature classifier. Second, with the three-feature implementation we used 7816 Bytes program flash memory of the embedded CY8C21634 microcontroller. This corresponds to 98% of the available program memory, see Section 5.2.2. Implementation of a fourth feature would not have been feasible with the selected processor. The third reason for the implementation decision was that we could show that even with calculating the features and classification decision, we could still sample with maximum sample rate and therefore meet the real-time computation criterion.

Tab. 5.4. Results for the first 7 iterations of the feature selection algorithm.

Selected features	Classification rate (%)
1,12	75.4
1,12,17	76.3
1, 2, 14,17	76.9
1, 2, 5,14,17	77.0
1, 2, 7,12, 13,17	76.9
1, 2, 5,14, 15,16,17	76.8

As we already stated in Section 5.2.8, the beam search does not guarantee that the identified subset performs optimal. We therefore computed the classification rates for all 1140 possible three-feature combinations. We could thereby show that the selected feature triple represents the optimal solution. We also compared the results of our feature selection with feature reduction using PCA as described in

Section 5.2.8. A comparison of the classification rates of both methods for reduction to two to seven features is depicted in Fig. 5.7. For our data, it can be seen that our feature selection method outperforms PCA. In the comparable case of reduction to three features, a classification rate of 73.9% could be achieved with PCA and 76.3% using beam search.

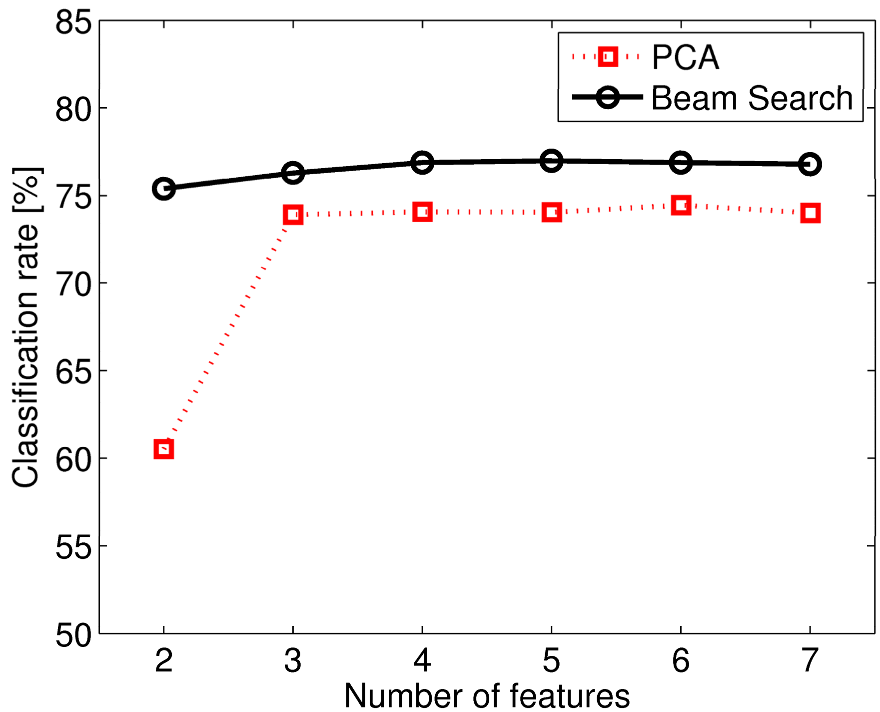

Fig. 5.7. Comparison of feature reduction using PCA and feature selection using beam search.

The confusion matrix for the selected feature combination is given in Tab. 5.5. Sensitivity is 77.7% and specificity is 73.6%. This result shows that no class is considerably favored over the other.

Tab. 5.5. Classification confusion matrix for the feature combination 1, 12 and 17.

	Class soft	Class hard
Classified as soft	9000	2583
Classified as hard	2993	8334

The characteristics of the selected feature subset become clear when visualizing the three-dimensional feature space. Runners on soft surface generally have

- smaller compression gradient (feature 1);
- higher step minima deviation (feature 12) and
- higher interstep time deviation (feature 17)

compared to runners on hard surface.

5.3.4 Classifier implementation

With the three-feature subset described in Section 5.3.3 we additionally performed experiments using all classifiers presented in Section 5.2.7 to confirm our choice of the LDA classifier. For cross-validation, 24 subsets were used, each consisting of the samples of one individual runner. The results of these experiments are presented in Tab. 5.6. Algorithm settings are the same as the ones given in Section 5.3.2, with the exception that only 4 hidden nodes were used for the NNet and that PART produced only 32 rules on the reduced feature set. It can be seen in Tab. 5.6 that only the Neural Network slightly outperforms the Linear Discriminant Analysis. However, the gain in classification rate is not statistically significant. Moreover, the problem of the complexity of the NNet classifier as already described in Section 5.3.2 remains, an implementation on the CY8C21634 microcontroller is thus not possible. We therefore decided to use the LDA classifier in the final application.

Tab. 5.6. Results of the experiments on the selected three-feature set using different classifiers. Classification rates are computed with a short context of three steps.

Classifier	Classification rate (%)
NNeigh, k = 1	71.2
NNeigh, k = 3	73.1
NNeigh, k = 5	75.3
NNeigh, k = 11	75.3
AdaBoost, N_{it} = 10	72.4
AdaBoost, N_{it} = 30	74.0
AdaBoost, N_{it} = 50	74.5
NB	74.6
PART	75.5
SVM	76.1
LDA	76.3
NNet	76.5

5.3.5 Effect of runner-dependent parameters

We already stated in Section 5.3 that our algorithm has to work independently of the runner, i.e. for different runner height, weight, shoe size and training level. The surface classification also has to work for different individual running speeds. We collected this information for all test participants. Tab. 5.7 shows the ranges for the parameters, as well as means μ and standard deviations σ. The training level was derived from the sports activity frequency of each individual on a range of one to four. On this scale, one indicates low or no regular sports activity whereas four stands for high running relevant activity, e.g. for marathon runners.

For the evaluation we computed the Spearman rank order correlation coefficient r_{spear} (Spearman, 1904) of the individual parameters and the classification rate of each runner. These correlation coefficients are ±1 for perfect positive or negative correlation and 0 if the samples are uncorrelated. The results of the evaluation are given in Tab. 5.7. In our case of 24 value pairs the null hypothesis that the samples are uncorrelated has to be rejected for $|r_{spear}| > 0{:}359$ at the 95% significance level (Olds, 1938). The results for r_{spear} given in Tab. 5.7 are all below this value and show that the value pairs are uncorrelated. If there was a significant correlation, we would have to assume that the runner-dependent parameters have some kind of influence on the recognition. For instance, the decision threshold would have to be adjusted for lighter runners if there was a correlation. However, by showing that there is no correlation between the individual parameters and the recognition rate we could assure that classification with the proposed system works independently of the runner.

Tab. 5.7. Individual runner parameters. Ranges are given for each parameter, as well as mean and standard deviation. The Spearman correlation r_{spear} of the individual parameters and the classification rates of each runner are also given.

Parameter	Range; mean; standard deviation	r_{spear}
Height [cm]	[156;196]; μ = 181.6; σ = 10.5	0.00
Weight [kg]	[46;125]; μ = 77.3; σ = 15.3	-0.19
Shoe size [US]	{7,9,11}; μ = 9.9; σ = 1.4	0.06
Training level	{1,2,3,4}; μ = 2.3; σ = 1.1	-0.03
Runner mean speed [km/h]	[8.3;15.0]; μ = 12.0; σ = 1.4	-0.15

5.3.6 Final evaluation on the microcontroller

It was important to implement our classification algorithm on the microcontroller that is employed in the product to verify our results. For these control experiments, we used the internal EEPROM (see Section 5.2.3) to store for each step only the classification decision derived with the described classifier. Longer contexts of 16 steps were used for the implementation. We let the test participants run totally freely, i.e. no requirements on running speed, step frequency or other parameters were made. An external observer counted and wrote down the number of steps on each of the surfaces that were tested during this evaluation. Thus, we could straightforwardly determine the classification rate. Tab. 5.8 shows the results of these experiments. Classification rates of more than 80% could be achieved.

Tab. 5.8. Description of the datasets that were used for the final evaluation on the microcontroller.

Dataset description	Number of recorded steps	Hard surface ratio (%)	Classification rate (%)
Park, grass and concrete	3480	61.5	82.8
Only asphalt surface	995	100.0	92.0
Forest soil, no inclination	4438	0.0	90.8
Forest soil and asphalt, running up/downhill	4448	65.9	80.3

5.4 Summary

For the realization of accurate surface classification using sensor output from the adidas_1, data was collected from 24 test runners on hard and soft surface. This data was labeled, and 19 features were extracted, which were chosen because they consistently represent the step information. A classification system using Linear Discriminant Analysis was then proposed. Using the classification rate as a criterion, a subset of three features was found that is suited to be implemented on the embedded system that is integrated in the running shoe. The system was evaluated with regard to the parameters shoe cushioning setting, runner height, runner weight and runner training level, which were all found to have no important effect on the accuracy of the classifier. The described classifier has been implemented in the current version of the adidas_1 running shoe.

During the complete analysis procedure, no assumptions regarding sampling rate of the sensor, memory or clock frequency have been made. This makes the methodology applicable to many other problems that require accurate embedded classification. The first important point is that computationally cheap features can be identified that well represent the sensor information. Secondly, a classifier has to be determined that establishes a good tradeoff between complexity and classification rate. Feature reduction is compulsory to provide a feature subset that is best suited for the problem at hand. Lastly, the result of the analysis that has been made on computationally powerful PC hardware has to be verified on the embedded system itself.

5.5 Future work

First results indicate that other important conditions can be classified using the shoe signal. One example includes the state of fatigue of a runner. An adaptation of the shoe hardness setting to a fatigued condition is therefore imaginable. Additionally, we will analyze the effect of elevation profile and speed changes in order to be able to classify these parameters, too.

We will furthermore investigate other application areas, for example accurate classification on microcontrollers in household appliances or for mobile phones. In the latter case, the computational framework conditions are not as critical as for microcontrollers. Still, a lot of effort similar to the one presented in this work has to be made to be able to implement pattern recognition algorithms on this kind of hardware.

CHAPTER 6 EMBEDDED CLASSIFICATION OF SPEED AND INCLINATION DURING RUNNING

6.1 Introduction

Smart sensors embedded in clothes and equipment for sports open novel opportunities to support and guide athletes. An example is the “adidas_1” running shoe, which is the first shoe that features an embedded system (see Fig. 6.1). This shoe was built to adapt to various running conditions. Examples for conditions that have to be taken into account include the prevailing surface situation, the fatigue state and the speed of the runner.

The adaptation was performed by changing the cushioning of the sole by a motor driven cable system inside the shoe. In order to recognize the current running situation, the heel compression of the shoe was continuously measured. The embedded microprocessor of the “adidas_1” processed this signal and performed a classification of the prevailing situation. Based on this classification result, a decision for a cushioning adaptation was made.

Pattern recognition methods in general were frequently used in recent locomotion related research (e.g. Schöllhorn, 2004; Wu and Wang, 2008). For example, a wavelet transformation was applied to electromyographic signals of runners (von Tscharner and Goepfert, 2003) for feature extraction. The resulting multi-muscle pattern could be employed for gender classification with high classification rate of 95%. In another study, the authors calculated three types of features (basic temporal/spatial, kinetic and kinematic) on human walking gait data (Begg and Kamruzzaman, 2005). The resulting set of 24 features was utilized to distinguish the gait of young and elderly subjects with a classification rate of 91.7%. Those application examples illustrated that pattern recognition algorithms can contribute considerably to data analysis tasks in locomotion related projects.

To the best of our knowledge, the embedded classification of running speed and surface inclination using the described compression measurements has previously not been investigated in the literature. Previous publications with the purpose of classifying these two variables used different sensor input and were not focused on embedded implementation. For example, a method for walking gait that was based on accelerometer measurements was presented (Aminian et al., 1995). For classification, the authors applied a neural network. The methodology was subsequently extended (Herren et al., 1999) for outdoor running. However, these approaches were based on triaxial accelerometry. The acceleration signal had implicitly included the running speed in its signal. Thus, the results from these studies could not be compared to results derived from compression measurements that were the basis for the running speed and surface inclination classification system that was developed in the present paper. Moreover, the measured signals were evaluated on PC hardware only. The complex mathematical calculations used for the complex neural networks that were employed (Aminian et al., 1995; Herren et al., 1999) may not have been possible with an embedded microprocessor. Nevertheless, the embedded classification of the speed and the track inclination variables were important in the “adidas_1” application scenario. Hence, the primary purpose of this paper was to use methods from pattern recognition to identify a classification system that distinguished three speed and three inclination classes based on the heel compression measurements.

In general, athletes can benefit from embedded classification systems. In the particular case of running with the “adidas_1”, the shoe could be adapted accordingly, setting itself into a cushioning state that was considered optimal for the given situation. However, the “adidas_1” shoe was just one example of smart sensors embedded in apparel and sport equipment. Comparable systems could be useful in other sports where an athlete can be actively supported by adapting the equipment to the prevailing situation. In a previous publication (Eskofier et al., 2009a), it was already demonstrated that accurate classification on an embedded microprocessor in sports was feasible. For this purpose, a framework for embedded classification was developed. This framework aimed at calculating features that described the originally measured signal well, while being at the same time efficiently calculable on embedded hardware. It was also discussed, which types of classifiers are suited for implementation on embedded hardware. The key idea that was followed was to conduct the various experiments on computationally powerful desktop computers, and to implement and validate only the most promising solution on the embedded hardware. Comparable systems could be useful in other sports where an athlete can be actively supported by adapting the equipment to the prevailing situation. Therefore, the secondary purpose of this paper was to further develop the previously employed (Eskofier et al., 2009a) general methods for embedded classification, so that the developed methodology could be more straightforwardly applied to other similar embedded classification tasks.

6.2 Methods

6.2.1 Data Collection

A total of 84 runners (30 female, 54 male) participated in a one-hour outdoor data collection. The age of the subjects was 32.9 ± 7.9 years (average, standard deviation). The subjects were not specifically chosen according to running experience; instead, the group contained runners of all activity levels. The measurement system consisted of three separate devices. Firstly, a “Polar RS800

Running Computer" (Polar Electro Oy, 2010) was used, which included an "S3 stride sensor" and a chest strap. This system was capable of measuring running speed, stride frequency and barometric height. The sampling interval for the collected signals was set to 5 s. These measurements formed the ground truth data, which means that the classes for the subsequent classification experiments were assigned according to these measurements.

Secondly, the heel compression signal f[*t*] of the runners was continuously measured using the "adidas_1" shoe (DiBenedetto et al., 2004; Eskofier et al., 2009a). The heel part of the shoe contained an adjustable cushioning element (Fig. 6.1). The amount of vertical compression that this element allowed was regulated by a motor-driven cable system (DiBenedetto et al., 2004). For the purpose of the data collection for this study, the cushioning element was manually put in a setting that allowed maximal heel compression and was not changed subsequently. This setting was chosen because the resulting compression signal had the highest possible signal-to-noise ratio. A hall sensor mounted at the top of the cushioning element detected the magnetic field strength induced by a small magnet at the bottom of the element. The sensor was sampled with a rate f_s = 342 Hz by the embedded microprocessor. The sensor-magnet distance d_m was computed from the measured field strength with an accuracy of ± 0.1 mm (Fig. 6.3).

Lastly, a specially programmed mobile phone (Eskofier et al., 2008a) was used to store the GPS position of the runner in intervals of 1 s. This allowed reconstructing all running situations after data collection. An example run is visualized in Fig. 6.2 based on the Google Earth (Google Inc., Mountain View, CA, USA) software. In this illustration, running speed is displayed as the height of the orange band along the running track. The software (Eskofier and Melzer, 2009) that was utilized to generate Fig. 6.2 is available for download from http://tinyurl.com/gervit.

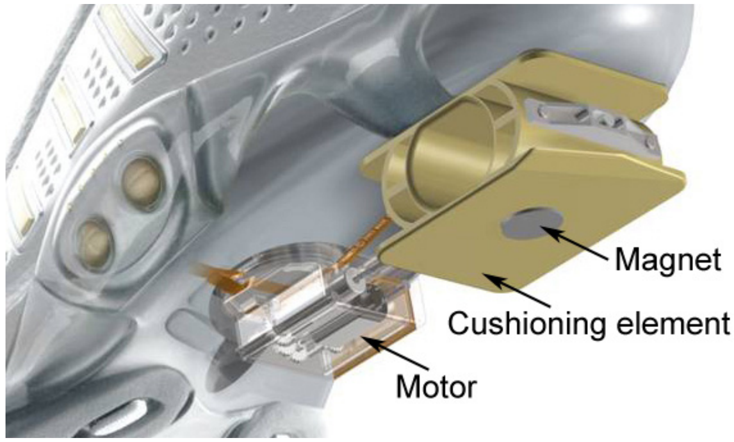

Fig. 6.1. The “adidas_1” shoe, its cushioning element, magnet and motor unit (DiBenedetto et al., 2004; Eskofier et al., 2009a).

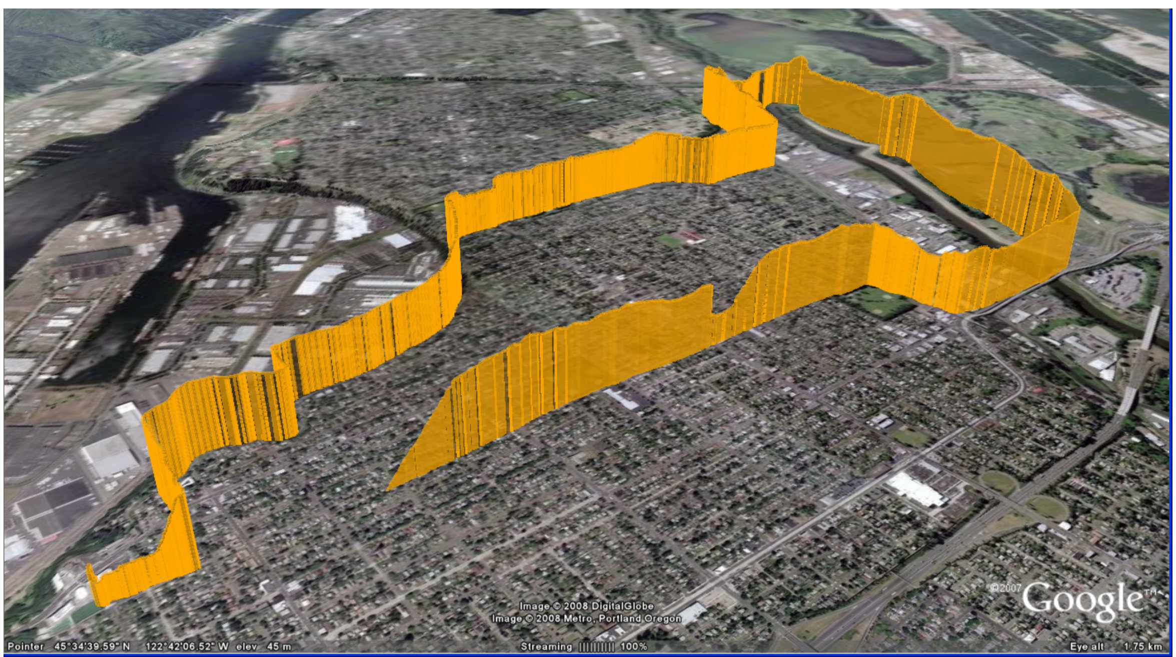

Fig. 6.2. Visualization of an example run in Portland, OR, USA. The height of the band represents the running speed.

After completion of the run, each participant was asked to fill in a questionnaire. Among other information, the questionnaire asked the test subjects whether they thought that the amount of equipment was in any way hindering to their run. Only two out of the 84 runners perceived a notable impediment by the equipment while running. This indicated that the collected data represented a free outdoor run very well (Eskofier et al., 2008a).

6.2.2 Data Processing

Out of the 84 study participants, 28 had to be excluded from further processing for various reasons. More specifically, five runners had incomplete data from the Polar RS800 system. The remaining 23 participants had to be excluded because of unusable data from the "adidas_1" shoe. In eight of these cases, data collection was not possible because the "adidas_1" was not present in all shoe sizes at the beginning of the study, and therefore the runners had to use other shoe models. In the remaining 15 cases (about 18% of all subjects), the runners were mid- or forefoot strikers. The measurement system of the "adidas_1" is located at the heel of the shoe and can therefore only capture meaningful data for rearfoot strikers, which represent more than 80% of the running population (Kerr et al., 1983).

6.2.3 Step Segmentation

Prior to feature extraction, the single strides were segmented by finding the deflection of the respective compression phases. Fig. 6.3 shows an example representation of the heel compression signal. The task of step segmentation was to find the beginning of the compression phase, i.e. the point in time when the runners started to compress the heel. A linear filter with the convolution vector

$$v_{con} = (-1, -1, -1, -1, -1, -1, -1, -1, 0, 1, 1, 1, 1, 1, 1, 1) \tag{6.1}$$

was used for this purpose. It was implemented with a moving window strategy in order to minimize the number of multiplications. That means that when a new sample was measured, the multiplication with the filter was only computed once. The multiplication result was stored in a ring buffer which contained only the preceding results according to the filter length. For each new sample, the filter output was then updated according to the elements in the buffer.

The filter was chosen for two reasons. First, noise was present in the signal during the time that the foot was in the air (Fig. 6.3). The filter had sufficient length to avoid misdetecting this noise as beginning of compression. Second, the filter yielded maxima for the beginning of the compression phase. The respective maxima after filtering marked the beginnings of the compression phase. In Fig. 6.3, the points where the maxima were located are depicted as red crosses.

The beginnings of each compression phase were defining the starts of consecutive strides $t_{s,i}$ and $t_{s,i+1}$, with i and i+1 indicating the respective stride number. Within the boundaries of one single stride, the point of maximum compression $t_{m,i}$ was identified by a linear search. The points of maximum compression are depicted as red circles in Fig. 6.3. The end of the compression phase $t_{c,i}$ (depicted with red stars in Fig. 6.3) was defined by the first sample value after maximum compression that was greater than the mean value before the actual compression phase minus two sample units. The reliability of this method was tested by visual inspection of 449 measured strides from six subjects (Eskofier et al., 2009a). The end of the compression phase was always identified at the correct position.

6.2.4 Feature Extraction

From every step, eleven hand-selected features were calculated (Tab. 6.1 and Fig. 6.3). These basic features were denoted by $F_1 \ldots F_{11}$. In order to add context information, the means μ_N and standard deviations σ_N over the features of the preceding N = {4, 8, 16} steps were also calculated. These features were denoted $\mu_N(F_n)$ and $\sigma_N(F_n)$, and were calculated from the n = 1, …, 11 basic features. For standard deviation calculation the unbiased version given in Eq. (6.2) was used. In this equation, c_m denotes a single calculated feature value for stride number m and $\bar{c}$ is the mean value of the respective feature values.

$$\sigma_N = \left(\frac{1}{N-1} \sum_{m=1}^{N} \left(c_m - \bar{c} \right) \right) \tag{6.2}$$

The gradients of all eleven basic features using N = 16 steps were also calculated and were denoted $g_{16}(F_n)$. Consequently, a total of N_f = 88 features were calculated. For feature extraction, the first five minutes of each run were not considered to ensure that the runners were warmed up and accustomed to data collection.

The obvious redundancy that was contained in these 88 extracted features was volitional. It was a goal from the start to use only a subset of the originally computed features in order to reduce complexity and to use only features with small or no mutual dependence. The feature subset selection algorithm will be presented below.

When conducting the speed classification experiments, it was noticed that a runner dependent feature rescaling considerably improved the result. A rescaling to the [0, 1] interval for each feature F_n of a runner according to

$$\hat{F}_n = \frac{F_n - \min(F_n)}{\max(F_n) - \min(F_n)} \tag{6.3}$$

was therefore implemented for each of the n = 1, …, N_f, = 88 features for the speed classification experiments. In this equation, $\hat{F}_n$ denominates the rescaled feature value.

6.2.5 Labeling

After consulting four sports experts, three classes according to running speed *v* and surface inclination, respectively, were defined (Tab. 6.2, Tab. 6.3). The class definition was chosen in a way that the resulting ranges of running speeds and inclinations covered an approximately equal amount of the distribution of the respective values for a typical training run. For this definition, an internal adidas report was used that had investigated relevant running speed and inclination distributions. In Tab. 6.2 and Tab. 6.3, k = 1…3 indicates the class number. Each detected step was labeled for the subsequent classification experiments according to these classes using the measured ground truth speed and surface inclination signals.

Tab. 6.1. Definition of the eleven basic features. From these, additional features were derived by computing context information over multiple steps. The resulting feature vector had 88 dimensions.

Nr.	Name	Formula
F_1	Inter step time	$t_{s,i+1} - t_{s,i}$
F_2	Time to peak	$t_{m,i} - t_{s,i}$
F_3	Maximum compression	$f\,[t_{m,i}]$ (measured value at $t_{m,i}$)
F_4	Compression time	$t_{c,i} - t_{s,i}$
F_5	Mean compression	$1/F_{4,i} \sum_{m=t_{s,i}}^{t_{c,i}} f[m]$
F_6	Step mass center	$1/F_{7,i} \sum_{m=t_{s,i}}^{t_{c,i}} \left(\left(m - t_{s,i}\right) f[m]\right)$
F_7	Step energy	$\sum_{m=t_{s,i}}^{t_{c,i}} f[m]$
F_8	Normalized compression time	F_4 / F_1
F_9	Normalized time to peak	F_2 / F_4
F_{10}	Compression gradient	$(f\,[t_{m,i}] - f\,[t_{s,i}]) / F_{2,i}$
F_{11}	Decompression gradient	$(f\,[t_{c,i}] - f\,[t_{m,i}]) / (t_{c,i} - t_{m,i})$

Tab. 6.2. Definition of the three classes according to running speed v.

Class	Class definition [m/s]
$\omega_{1,v}$	$0 \leq v < 2.5$
$\omega_{2,v}$	$2.5 \leq v < 3.6$
$\omega_{3,v}$	$3.6 \leq v$

Tab. 6.3. Definition of the three classes according to surface inclination α. A negative value indicates that the athlete was running downhill.

Class	Class definition [deg.]
$\omega_{1,\alpha}$	$\alpha < -3°$
$\omega_{2,\alpha}$	$-3° \leq \alpha \leq 3°$
$\omega_{3,\alpha}$	$3° < \alpha$

6.2.6 Classifiers

In the classification experiments, five different classifiers were compared in order to evaluate their performance on the measured data. The selected classifiers were chosen because each of them can be implemented on an embedded microprocessor. More specifically, these classifiers were used for the evaluation:

- Bayes Classifier (BC). The BC makes use of the assumption that all features are mutually independently distributed (Niemann, 1983). This assumption allows a straightforward estimation of the classifier parameters from the samples that are used for classifier training by direct mean and variance computation. The resulting linear discriminant function g_k can be computed by a multiplication of each feature with a weight factor, adding the results of the multiplications and comparing the sum against a threshold. Due to this simplicity, the BC is well suited for embedded implementation. The BC has been proven to perform well in many classification tasks (Domingos and Pazzani, 1997; Langley et al., 1992).

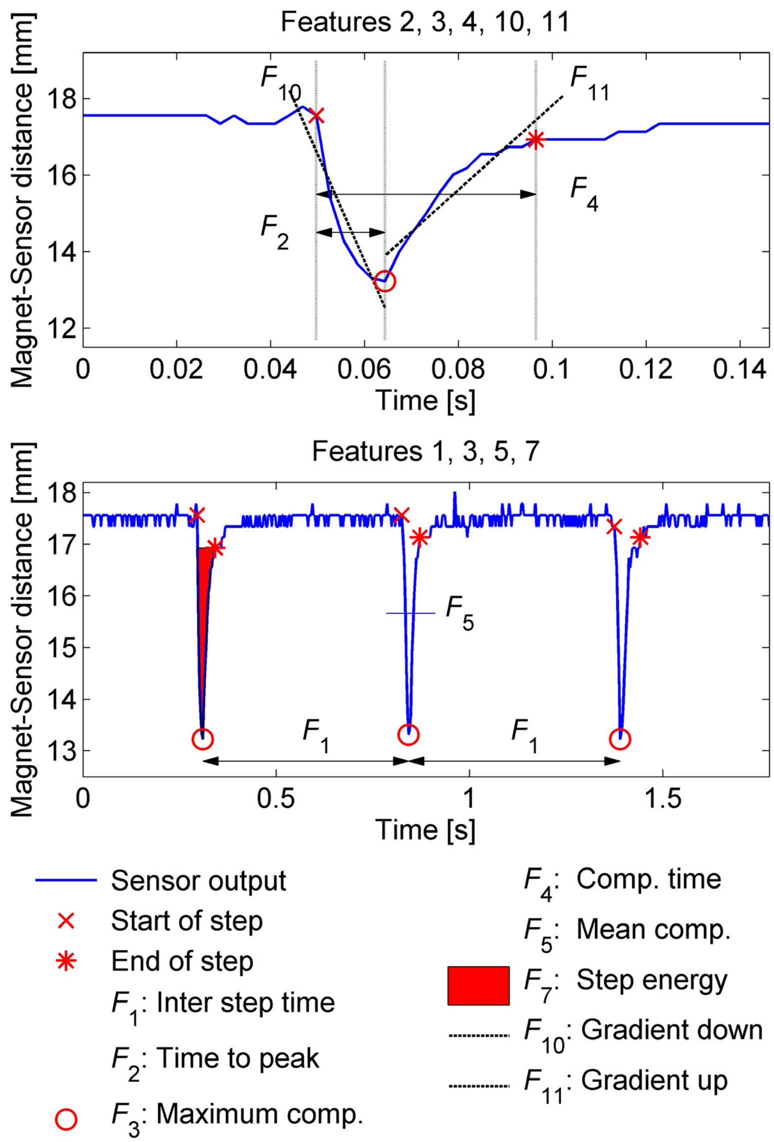

Fig. 6.3. Illustration of the eleven basic features. From these, additional features were derived by computing context information over multiple steps. The resulting feature vector had 88 dimensions.

- Linear Discriminant Analysis (LDA). The LDA classifier is based on Fisher's (Fisher, 1936) work on discriminant methods. It is a transformation that aims at minimizing the variability within a class, and maximizing the distance between classes. When LDA is applied for classification, the feature space is effectively projected onto a single axis. On this single axis, a linear decision boundary is applied for differentiation. In contrast to BC, the full covariance of the distribution is considered. However, the resulting linear discriminant function g_k can be implemented in the same simple way as for the BC. Applications of LDA can be found in a variety of fields, including face recognition (Lu et al., 2003) and document classification (Ye and Li, 2005).

- Polynomial Classifier (PC). The PC does, in contrast to BC and LDA, not use the parameters of the distribution of the features in the sample used for classifier training, but estimates the discriminant function g_k directly from this training sample (Niemann, 1983)). Different polynomial degrees can be used. In the present study, given that a simple classification rule for the embedded system had to be used, a linear polynomial was chosen. The estimation of the discriminant function g_k is then performed by solving a least squares systems of equations. The discriminant function g_k that results is again linear and can be implemented in the same simple way as for the BC and LDA classifiers. The PC has been shown to obtain good classification results in a variety of studies (e.g. Liu and Sako, 2006; Franke, 1997).
- Support Vector Machine (SVM). Support Vector Machines operate by first transforming the features into a high dimensional space (Vapnik, 1998). This transformation can be computed quite efficiently by different kernel functions (Schölkopf and Smola, 2002). In the present study a linear kernel was chosen, again due to reasons of computational simplicity. After the kernel transformation, a linear decision boundary with maximum margin is established in the resultant high dimensional space. While the process of training is complex, it is computed on a desktop PC and therefore not

relevant for the implementation of the classifier on the embedded system. A standard SVM implementation was used throughout the present study ("libSVM" (Chang and Lin, 2001), freely available on the web). Support Vector Machines obtain high classification rates in many pattern recognition tasks (Sapankevych and Sankar, 2009). Numerous applications of this classifier exist, including image classification (Chapelle et al., 1999) and email categorization (Drucker et al., 1999).

- Multilayer Perceptron Classifier (MLP). The MLP is built to simulate neuron interaction in the human brain (Specht, 1990). The neurons are implemented by multiple single nodes that are connected in multilayer nets (Duda et al., 2001). Each node has an input and an output. A feature value that is input into the node is subjected to a specified nonlinear function, e.g. a sigmoid function. Weights specify the contribution of individual nodes to the classification result. These weights are adjusted during classifier training according to different learning strategies (Hagan and Menhaj, 1994). The resulting discriminant function g_k is nonlinear. For the classifier implementation, the complete weight structure multiplication and the evaluation of the nonlinear (e.g. sigmoid) function needs to be performed on the embedded system. While this is still practicable, considerable higher computational demands are posed to the embedded system. MLP classifiers are frequently applied and several survey articles cover them (e.g. Baxt, 1995; Chua and Yang, 1988; Hunt et al., 1992).

With these classifiers, each vector of observed features $\boldsymbol{x} = (F_1 \ldots F_{88})$ was assigned to the class ω_k for that the discriminant function g_k of the respective classifier is maximal. In the experiments, five-fold cross-validation was performed in order to ensure generalizability of the results. In each of the cross-validation iterations the classifier was trained using all but the feature vectors from one specific fold. Subsequently, the feature vectors from the remaining fold (the test set) was classified according to maximum g_k. The mean classification accuracy was then computed as the average over all cross-validation iterations. To ensure that feature vectors were equally distributed over all classes, 10,000 vectors from each class were randomly selected from the collected data. The equal distribution of feature vectors per class allowed using equal priors e.g. for the Bayes Classifier.

Classification accuracies were deemed significant if the null hypothesis that classification was random could be rejected using a binomial test with significance level $\alpha = 0.01$.

6.2.7 Feature Selection

Due to the requirement that all computations had to be made in real-time on the employed microprocessor of the “adidas_1”, it was impossible to implement a classifier based on the complete set of 88 features. A feature selection algorithm was therefore implemented, the dynamic programming algorithm (Niemann, 1983). It required that the initial feature set was rather small, and that the scoring metric was monotone and separable. This is true for the Mahalanobis distance (Mahalanobis, 1936)

$$G_{k,l} = (\mu_k - \mu_l)^T \mathbf{\Sigma}^{-1} (\mu_k - \mu_l) \tag{6.4}$$

between two classes ω_k and ω_l. In Eq. (6.4), μ_k and μ_l denote the class means and $\mathbf{\Sigma}^{-1}$ is the common covariance matrix of all features. The dynamic programming algorithm was applied in multiple iterations using the Mahalanobis distance criterion. In each iteration, one single feature was added that gave the highest improvement for the worst class pair.

6.2.8 Microprocessor Implementation

The number of features that was possible to be computed in real-time on the employed microprocessor of the “adidas_1” was empirically found to be only two. Therefore, for the microprocessor implementation, the best performing two features had to be chosen. However, combinations of more features were also evaluated, because they could be implemented in future “adidas_1” versions that employ computationally more powerful microprocessors.

In order to demonstrate the ability of the developed methodology to perform accurately on the embedded microprocessor of the “adidas_1” shoe, the best classification system (according to the results on a desktop PC) was implemented on this microprocessor. The first important framework requirement for this implementation was the limited size of the internal memory (256 Bytes). This meant that the program had to be as short as possible to save ROM and that it had to economize on variables. Moreover the classification had to be done in real-time with the available computing power. The microprocessor of the “adidas_1” was clocked with 24 MHz, which posed considerable demands to the embedded classification algorithm. Finally, a floating point unit was lacking and therefore all computations had to work with integer operations only.

Considering this different hardware architecture, a final evaluation of the performance of the classification system on the embedded microprocessor was made. For this purpose, the classification decisions made by the microprocessor were compared with those of a desktop PC. For the multi-class decision system a one-against-one approach was used, where the decision for every class against each other was calculated. The one class that won the most decisions was the selected class. If two classes won exactly the same number of comparisons, the selection depended on the iteration sequence, and a decision for the first considered class was always made. In the case of a three class problem this was equal to a decision tree of depth 2. Therefore, two decision functions per step were calculated.

6.3 Results

6.3.1 Inclination Classification

The resulting classification rates for the inclination classification are given in Fig. 6.4. Fig. 6.4 shows the class-wise averaged classification rates that were obtained using one to six features that were selected according to the feature selection algorithm for each classifier.

The results of the feature selection algorithms showed that the most important features for this task were μ_{16} (F_2), μ_{16} (F_9) and μ_{16} (F_{11}). These are the mean values over 16 steps computed from time to peak, normalized time to peak and decompression gradient, respectively. Those features were, in the given order, selected in almost all cases for the classification. It can be seen that by using more features, better classification results were achieved in general. The best accuracy of 67.2 % class-wise mean accuracy was reached by using six features and the MLP classifier. This classification result is significantly different from random ($p < 0.001$). The confusion matrix for this case is given in Tab. 6.4.

Tab. 6.4. Confusion matrices for six features when using the MLP classifier for inclination classification. Classification accuracy values are given in %.

Classified as	$\omega_{1,v}$	$\omega_{2,v}$	$\omega_{3,v}$
labeled $\omega_{1,v}$	**54.6**	22.6	22.8
labeled $\omega_{2,v}$	5.5	**87.8**	6.7
labeled $\omega_{3,v}$	19.7	21.1	**59.2**

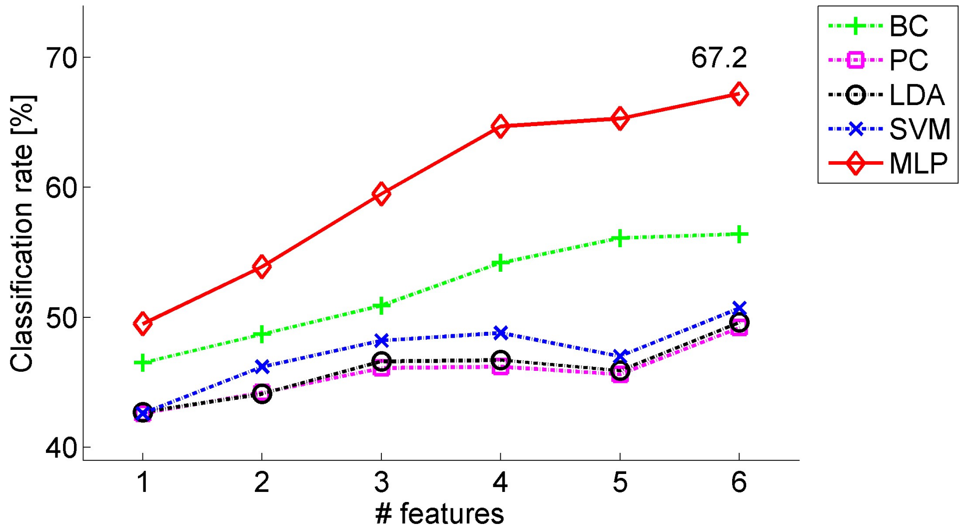

Fig. 6.4. Inclination classification results with feature selection. Depicted are the class-wise averaged rates.

6.3.2 Speed Classification

The resulting classification rates for the speed classification are given in Fig. 6.5. Fig. 6.5 shows the class-wise averaged classification rates that were obtained using one to six features that were selected according to the feature selection algorithm for each classifier.

The results of the feature selection algorithms showed that the most important features for this task were μ_{16} (F_3) and μ_{16} (F_1). These are the mean values over 16 steps computed from maximum compression and inter step time, respectively. Those features were, in the given order, selected in almost all cases for the classification. It can again be seen that by using more features, better classification results were achieved in general. The classification accuracies showed a noteworthy rise when using two features for all classification approaches. For some approaches (BC, SVM, MLP), another considerable rise in the classification accuracies was noticed when using four features. The best accuracy of 89.2 %

class-wise mean accuracy was reached by calculating six features and applying the MLP classifier. This classification result is significantly different from random ($p < 0.001$).

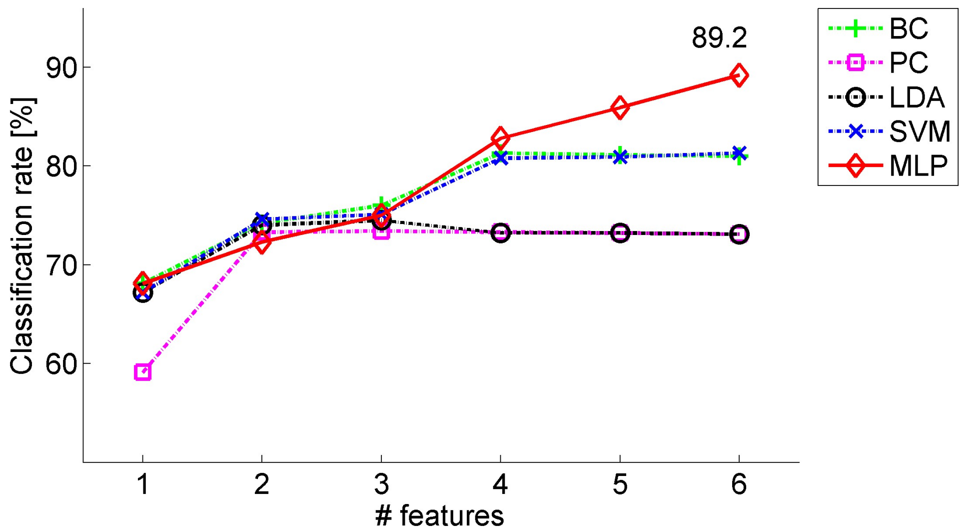

Fig. 6.5. Speed classification results with feature selection. Depicted are the class-wise averaged rates.

For the two-feature case (the number of features that could be computed on the currently employed microprocessor of the “adidas_1”) the best results were found using the SVM classifier and features μ_{16} (F_3) and μ_{16} (F_1). The class-wise mean accuracy was 74.6%, which is significantly different from random ($p < 0.001$). The confusion matrix for the SVM two-feature case is given in Tab. 6.5.

Tab. 6.5. Confusion matrices for two features when using the SVM classifier for speed classification. Classification rates are given in %.

Classified as	$\omega_{1,v}$	$\omega_{2,v}$	$\omega_{3,v}$
labeled $\omega_{1,v}$	**73.7**	11.7	14.7
labeled $\omega_{2,v}$	4.0	**74.9**	21.1
labeled $\omega_{3,v}$	7.7	17.0	**75.2**

6.3.3 Microprocessor Evaluation

Due to the low classification rates of the surface inclination system, only the SVM two-feature speed classification system was implemented on the product version of the "adidas_1" shoe. Only the trained classifier was implemented, using the Support Vectors in the decision process. For the speed classification case, the classification decisions made by the microprocessor were compared with those of a desktop PC. The tests showed that 99.2% of the classification decisions were the same.

6.4 Discussion

The inclination classification (Fig. 6.4) could not be performed with high classification rate. A major reason for this result was the fact that the quality of the signal measured with the "adidas_1" was decreasing with increasing inclination (both up- and downhill). This can be seen in the confusion matrix (Tab. 6.4) for this case, which showed to be unbalanced with a preference for the class for the low inclination range. Further examination of the data revealed that the reason for this unbalance was that the measurement sensor was located in the heel part of the shoe. When running up- or downhill at certain inclinations many runners tended to land more on the mid- or forefoot. In consequence, less overall compression was sensed. This resulted in a reduced signal to noise ratio and thus in a lower classification accuracy. To resolve this issue, at least a second sensor would have been needed in the front part of the shoe. Using such a sensor, additional information would have been available for classification. The incorporation of a second or even more sensors, therefore, is part of the future research work within this project.

The speed classification accuracies showed a considerable rise when using two features for all classification approaches (see Fig. 6.5). Another noteworthy rise was noticed when using four features for some approaches (BC, SVM, MLP). The result that adding more features, and therefore more information, to the classification process and to then obtain higher classification rates is often observed in pattern recognition (Duda et al., 2001; Theodoridis and Koutroumbas, 2009). The particular result in this study suggested using either two or four features for the final implementation on the microprocessor. Because of the limited hardware of the microprocessor, only the two feature approach was possible. In a future implementation on a computationally more powerful microprocessor, more features might also be implemented based on the results. The two selected features for classification were μ_{16} (F_3) and μ_{16} (F_1), as these were performing best.

Although the MLP classifier delivered results that were among the best for all conducted classification experiments, it was not chosen for the final implementation. First, it is computationally more demanding in a working classification system than the other classifiers. The BC, LDA, PC and SVM classifiers all have a different approach to classifier training, with typically increasing complexity. In the working classification system, however, all these classifiers pose a similar demand to the system they are implemented on. Only the MLP classifier is computationally considerably more demanding with its necessity of a more complex incorporation of the neuron weights and the requirement of the evaluation of the nonlinear function. Second, in the two feature case, the SVM classifier obtained the highest accuracies in any case. Thus, a decision was made to use SVM in the final microprocessor implementation. The confusion matrix in Tab. 6.5 for this case showed that the SVM yielded nearly equally good results for all classes. This meant that no speed class was considerable favored over another. This is an advantage of the system, because it prevents an overestimation of a certain prevailing speed condition.

The runner dependent feature rescaling was needed in order to obtain more accurate classification results. This rescaling thus had to be implemented on the microprocessor, which might be considered a disadvantage due to the additional calculational effort. However, the additional computational effort was low because only the current extreme values of the two features selected for implementation had to be stored in memory. Those were updated regularly, this way the shoe adapted to different runners. Moreover, the actual computation of the rescaling could be efficiently implemented and thus the real-time requirements could still be met.

The fact that the microprocessor classification results were the same as the results on a desktop computer in 99.2% of classified steps showed that it was feasible to do all the evaluations that require high computational effort on desktop computers while only evaluating the final product solution on the embedded microprocessor. The obtained classification results were practically the same on both systems. This confirms the results shown in a previous study (Eskofier et al., 2009a). The 0.8% of steps that were not classified in the same way as on the desktop computer were a negligible minority. Although a deficient implementation of the classification algorithms on the desktop PC could, in principle, also be the reason for the discrepancy, the different classification results were mainly ascribed to the different hardware architectures of both systems, e.g. a missing floating point unit on the microprocessor.

The core idea for enabling embedded classification was using computationally simple features and classifiers that could also be implemented on embedded microprocessors. Furthermore, all comparative experiments were performed on computationally powerful desktop computers, and only the best solution was implemented and validated on the embedded hardware. This approach was again successful, and an accurate embedded speed classification system could be developed. The proposed methodology will be helpful in many tasks in sports where classification on embedded systems has to be performed.

6.5 Summary

This research demonstrated the application of pattern recognition methods to detect running surface inclination and running speed using heel compression measured with the “adidas_1” running shoe. A set of 88 features was manually designed that was suited for the classification task at hand. The features were computationally inexpensive and could be calculated using the embedded microprocessor of the “adidas_1” shoe. Several classifiers that are suited for embedded implementation were compared with respect to their classification rate. Subsequently, it was shown how a subset of the original 88 features, which were most important for the classification task, could be identified. The applicability of the developed speed classification system was demonstrated by implementing and evaluating it on the embedded microprocessor of the “adidas_1”. Thus, a classification of the prevailing running speed was performed directly on the embedded system.

It was shown that in the three-class inclination case, a classification rate of 67.2% could be obtained using six features and a MLP classifier. Better performance was not possible due to the fact that only the heel compression was measured, and for the classification of some track inclinations this available sensor information was insufficient. However, it was demonstrated that if continuously good heel compression signals were available, as it was in the three-class speed classification case, acceptable classification rates of 74.6% could be achieved using only two and even 89.2% using six features. This result suggested that a trained automatic system could quite precisely support the athlete, for example by providing more shoe stiffness and thus more stability by the “adidas_1” running shoe when the sportsman was running faster.

6.6 Acknowledgments

The authors would like to thank the adidas innovation team *ait*. for the financial and technical support of this research. We are indebted to Pascal Kühner from the University of Koblenz-Landau for his help in data collection, and also thank the participants of the running study. Special thanks go to our colleagues that were proofreading this manuscript for their time and support.

CHAPTER 7 REAL TIME SURVEYING AND MONITORING OF ATHLETES USING MOBILE PHONES AND GPS

7.1 Introduction

Information about the subjective feeling of athletes is very important for many domains. One example is sports product testing. Details such as appearance, functionality, handling and ergonomics are important points that have substantial influence on the choice of the customer. The subjective feeling of athletes concerning their equipment therefore is an important criterion for the success of a product. Another example is perception research, where information about the perceptive state of an athlete is collected over a longer period of time. Training and performance optimization can also benefit from this information.

The common problem is to access the desired information while the athlete is in a typical situation. Real-time surveying is of course possible in a lab environment, e.g. on a treadmill. This has already been done for example by (Acevedo et al., 1996) and (O'Halloran et al., 2004) in psychological studies. The obvious disadvantage is that the results are biased due to the nonnatural lab situation. The more normal situation of a long distance outdoor run, for example, is much harder to assess because direct contact to the athlete is complicated or not possible at all. In most cases, the desired subjective as well as objective information is collected after the respective sports activity. (Abele and Brehm, 1985) have done this in a study where they wanted to assess the change in the mental state of athletes caused by a set of different sport activities. The participants had to answer questions concerning their subjective actual feeling-states before and after a 60 to

90 minute course of physical activity. When following this procedure, part of the information is lost because it is not possible for the athlete to memorize all individual details of his perception. It is more desirable to access the desired information at certain time points or after reaching for example a certain waypoint on a predefined route in real-time.

To achieve this, we designed and implemented a system for the surveying of runners using mobile communication equipment, i.e. a standard mobile phone. For this specific project we decided to use the Java Platform, Micro Edition (Sun, 2002b; Sun, 2007) as programming language because it is implemented on most mobile phones. The advantages of the cell phone hardware platform are manifold. It is lightweight, mobile and highly configurable. There is no extra cost associated with hardware development, only the software has to be adapted to the specific requirements at hand. Most mobile phones are highly suitable because of their advanced computational power. Communication and real-time data transmission could also be implemented easily if desired.

The system we implemented fulfills the following requirements:

- Predefined questions are handed over to the system as audio files, associated answers are recorded.
- The athlete is asked the questions at certain predefined time points.
- Alternatively we implemented the option to react to certain external events. This includes for example significant changes in running speed or altitude and the achievement of waypoints. External hardware like GPS receivers can be connected to the phone via Bluetooth to enable this.
- Headsets can be connected to the mobile phone via Bluetooth as well to assure maximum comfort for the athlete.
- If desired, arbitrary audio files (music) can be played between the question units for the purpose of motivating the sportsman.

- Configuration of the system is possible both directly on the cell phone or a personal computer.

Once the configuration is completed, the software requires no further interaction. That way, it could be used at anytime that is convenient for the test person. Starting the predefined survey program requires only the press of a button.

We will give a short overview of previous work on the topic of athlete monitoring. In the following, the important building blocks of our system will be explained. We will also show an experimental evaluation of our mobile monitoring solution with 84 runners. This evaluation was done within the scope of a larger psychological study for which subjective information during a one hour outdoor run was needed. As a result and conclusion we will show that our system is highly reliable, providing very valuable information about the psychological and physiological state of an athlete.

7.2 Previous Work

The authors know of no previous work that aims at implementing a sports monitoring and surveying device by using the capabilities of a mobile phone. There are, however, several publications that deal with the same topic. An obvious example are telemonitoring devices that rely on radio transmission. (Wang et al., 1992) showed the application of such a device in shell rowing. The disadvantage of such systems is that the athlete might get out of transmission range and information would be lost. An extensive review by (Armstrong, 2007) gives an overview about other applications of wireless connectivity for health and sports monitoring. None of the reviewed publications implements a method for getting real-time feedback about the subjective state of an athlete.

(Hallberg et al., 2004) present a system that monitors heart rate and location of an athlete via GPS. The information is sent via GPRS to a media server that provides an enriched media experience to viewers of sports events. They also showed the practical usability in an example for cross country skiers. However, no direct audio feedback from the athlete concerning the subjective fitness and psychological state is featured.

Another application of GPS and physiological information was presented by (Saupe et al., 2007). They also use Google Earth for the visualization of physiological parameters as well as information about endurance sport training activities on a large high resolution display. In contrast to our work, no direct subjective information is acquired for the analysis.

7.3 Methods and Materials

7.3.1 Java Platform, Micro Edition

One of our framework requirements was that our software should work with a broad range of mobile phones. The Java Platform, Micro Edition (Java ME) is preinstalled on most phones and therefore fulfills this requirement. We consequently chose this software platform for our implementation.

The capabilities of an environment for the Java Virtual Machine in the Micro Edition are defined by three important building blocks, see Fig. 7.1. The most basic is the device configuration. Most common for mobile phones is the Connected Limited Device Configuration (CLDC) as specified by (Sun, 2007). It specifies the minimum hardware requirement. In the current version 1.1 these requirements include a 16-bit or 32-bit processor, 32 kByte RAM and at least 160 kByte non-volatile memory. The high level programming interfaces are defined by profiles. The Mobile Information Device Profile (MIDP) is built on the CLDC and offers basic APIs for programmers. The current version 2.0 (Sun, 2002b) offers user interaction classes, security management and basic file connection capabilities. The third important building block for software development on mobile phones are the optional APIs. Phone manufacturers can decide, which of these packages called Java Specification Requests (JSR) they want to implement on their devices. Factually, a lot of these additional packages are standard and can be used on most phones. Important optional APIs for our software are the:

- Mobile Media API (JSR 135, (Sun, 2006b)) for playing and recording sound files and video processing.
- File Connection API (JSR 75, (Sun, 2004)) for file handling.
- Bluetooth API (JSR 82, (Sun, 2002a)) for Bluetooth connectivity.
- Location API (JSR 179, (Sun, 2006a)) for position determination.

Applications that build on the MIDP and any of the optional blocks are commonly referred to as MIDlets.

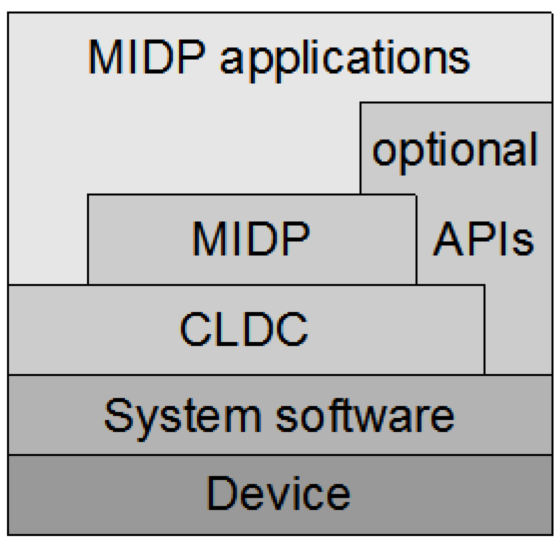

Fig. 7.1. High level Java ME architecture view.

7.3.2 Mobile Phone Hardware

For the development of our MIDlet software we had to restrict ourselves to mobile phones that offer a CLDC 1.1 compatible hardware and MIDP 2.0 with the optional APIs as stated above. Most of the current cellular phones fulfill this requirement. We wanted to show with our reference implementation that our software is working on different types of mobile phones. The companies Nokia and Sony Ericsson offered the best online support for developers, we therefore chose a Sony Ericsson W850i (116 g), a Nokia N70 (126 g), a Nokia E50 (104 g) and a Nokia 6110 Navigator (125 g), see Fig. 7.2. Each of the selected devices offers a slot for memory cards and thus enough capacity to store information even for very long studies. The phones are all lightweight and have high battery capacities for more than 4 hours of active use.

Fig. 7.2. Selected phone models. From left to right the Sony Ericsson W850i, the Nokia N70, the Nokia E50 and the Nokia 6110 Navigator are shown (Nokia, 2007; Sony-Ericsson, 2006).

7.3.3 Development Environment

Both selected phone manufacturers offer developer tools that provide device emulators and advanced debugging capabilities. This is very important for MIDlet development because error identification on the mobile platform can be very tedious. The software development itself was done with NetBeans 5.5 with mobility pack. The manufacturer SDKs can easily be integrated in this development environment, additional tasks like code obfuscating and optimization are thereby provided.

7.4 Implementation Details

7.4.1 Software Structure

The evaluation system had to be easily configurable and very flexible in order to support a lot of different devices and study options. Questions and position data had to be recorded as well as predefined sound files played to the athletes. The software had to work with minimum preparation time and no user interaction at all once the tests were running. The building blocks of our software that are shown in Fig. 7.3 will be explained in the following.

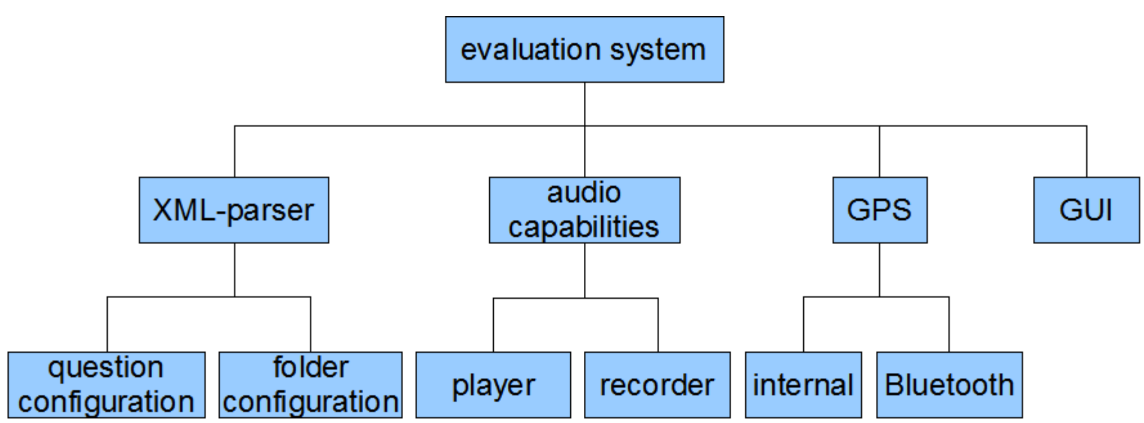

Fig. 7.3. Structural diagram of the software for the evaluation system.

7.4.2 XML-Parser

The configuration of question units and storage location for the recorded sound files and position information was done with a XML file. Additionally, we stored information like start time, recorded files, identification number of the mobile phone and other information in a XML info file after the survey was completed. Because XML parsing is only supported by JSR 172, which is seldom implemented, we had to come up with our own parser.

The system can be configured to play sound files at certain time points or in reaction to external events, e.g. when a predefined distance has been covered. Subsequent to the questions, answers can be recorded; in this case a short sound is played at the beginning and the end of the recorded time span. The system can also be configured not to record after playing a sound file in case the athlete should be briefed, e.g. to decrease the pace. Another option is silent recording, i.e. recording without playing any sound at all. We used this option to capture the breathing noise of the runner in order to be able to determine the respiratory frequency.

7.4.3 Audio capabilities

The audio part supports threaded playing and recording in order to allow for example seamless position information storing even during question units. The configuration is done in one single XML file. In case there is an overlap of sound files, i.e. in the event that the combined playback and recording duration is longer than the span to the desired start of the next unit, this overlap is automatically resolved. Fig. 7.4 illustrates this further. The order of the question units in the configuration file defines the precedence for the overlap resolution.

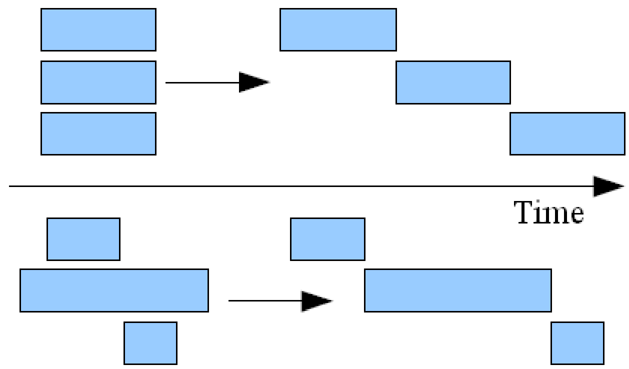

Fig. 7.4. Overlap resolution. Audio files are played in the order that is defined in the configuration file.

The audio codec is automatically selected dependent on the sound files to play. For recording, we used a codec suited for speech. We found that a bit rate of 128 kbit per second with a sample rate of 8 kHz was sufficient for our purposes.

7.4.4 GPS Integration

GPS integration was an integral part of the software development in order to have access to speed, altitude and position information. The software works for phone models with integrated GPS like the Nokia 6110 Navigator as well as with an external GPS receiver (e.g. a Nokia LD-3W) connected via Bluetooth. The GPS data is sent in an interval of approximately one second, which is sufficiently precise for the purpose of recording running position information. Each sample consists of longitudinal and latitudinal position information, speed of movement, altitude, time information and various precision and validity parameters.

The data is stored in the original NMEA (National Marine Electronics Association) format (Langley, 1995), as well as directly converted to the KML (Keyhole Markup Language) format used by Google Earth. This conversion allows for a quick and easy method of visualizing the run. Run parameters like speed and psychological state can be represented as height above ground (see Fig. 7.5 in the results section 7.6) or color coded.

7.4.5 Graphical User Interface

The GUI that we developed extends the limited window manager provided by Java ME. It allows changing several configuration options, to connect to the internal or external GPS device and view the current position information. Once the surveying process is started, no further user interaction is required in order to minimize interference with the athlete.

7.5 Experiments

Experimental evaluation of the system was performed in the context of a psychological study with 84 runners in Portland, Oregon (USA). While the details of the study itself are beyond the scope of this paper, the relevant points for the evaluation of our mobile surveying system will be given.

The objective of the study was to appraise the subjective feelings of the runners during a recreational run. Each athlete participating in the study was asked to run outdoors for one hour. They could freely choose their preferred route and speed as we could record these parameters with the GPS signal. We chose to use the Nokia 6110 Navigator cell phones for the purpose of this study as they have an inbuilt GPS receiver. This prevented that the runners had to carry an external GPS receiver as extra equipment. The phones were placed in a belt that was attached to the upper arm of the participants. The runners also wore a Bluetooth headset to ensure maximum comprehensibleness.

Before starting the run, an audio file with instructions was played to the participants, followed by a first set of 8 questions. After each question, a short sound was played to indicate the start of the recording interval. The end of the three second recordings was marked by another sound. The athletes were instructed to answer each question about their subjective state with a self-rated grade as given in Tab. 7.1. An example question is “Do you feel motivated?”.

Tab. 7.1. Grades for the athlete self-rating.

Spoken answer	Meaning
0	not at all
1	very little
2	little
3	somewhat
4	rather
5	very
6	extremely

Directly following this first question unit the runners were asked to start their one hour run. During this run, question units identical to the first one were posed with an interval of 5 minutes between the start of each unit. A total of 13 question units with 8 answers per unit were thus recorded for each athlete.

7.6 Results

The mobile surveying system worked without technical difficulties for all 84 runners. A total of 8736 sound files with self-rated subjective state information were recorded. We transcribed the audio files by listening to them and then manually entering the spoken answers in a data matrix. We found that 355 sound files (4.1%) were unusable, i.e. containing no meaningful answers. The main reason for this was that at the beginning of the study, we did not clearly enough emphasize the fact that the answers should be spoken in between the two sounds indicating the recording time. Consequently, a lot of runners spoke their answers right after the questions were asked when we started the evaluation. We therefore changed the set of instructions after the 19^{th} participant so that the record interval was clearly explained. After this change, only 62 entries could not be acquired, mostly because the runners were exhausted at the end of their runs and did not answer in time. In summary it can be said that as long as the athletes gave their answers during the recording time, the information was audible and could be transliterated. No audio sample was lost due to malfunctioning of the mobile phone.

We also collected a questionnaire after completion of each run. Among other details, we wanted to know how much impeded the athletes felt by the additional equipment. The results can be seen in Tab. 7.2. It can be seen that most runners perceived the cell phone and headset as very little or little impeding. Only 4 out of 84 athletes found the equipment to be hindering.

Tab. 7.2. Impediment by the additional equipment as perceived by the 84 study participants.

Perceived impediment	Number of runners
very little	51
little	29
some	2
much	2
very much	0

It was also very important that the GPS signal recording worked in order to get reliable position and running speed information for our study. After analysis of the recorded data, we found that only 0.07% (173 out of 260214) of the position samples were unusable. Because of the fact that in no case two consecutive samples were missing, it was straightforwardly possible to interpolate the unavailable position information with a linear estimation strategy.

Fig. 7.5 illustrates the GPS information for one example runner. The image is based on Google Earth. The speed is displayed as the height of the colored band along the running track. Thus, the chosen running track and speed can easily be analyzed. We can also show the subjective states, in the example of Fig. 7.5 the state of perceived fatigue is displayed color coded. Yellow means little or no perceived fatigue. The redder the band becomes, the higher is the perceived fatigue state of the athlete. It can clearly be seen that the perceived tiredness is increasing during the run. This visualization allows for a straightforward and convenient analysis of the interplay between various parameters like elapsed time, speed, elevation circumstances and subjective state. Additional data, e.g. heart rate, can easily be integrated into the visualization if present.

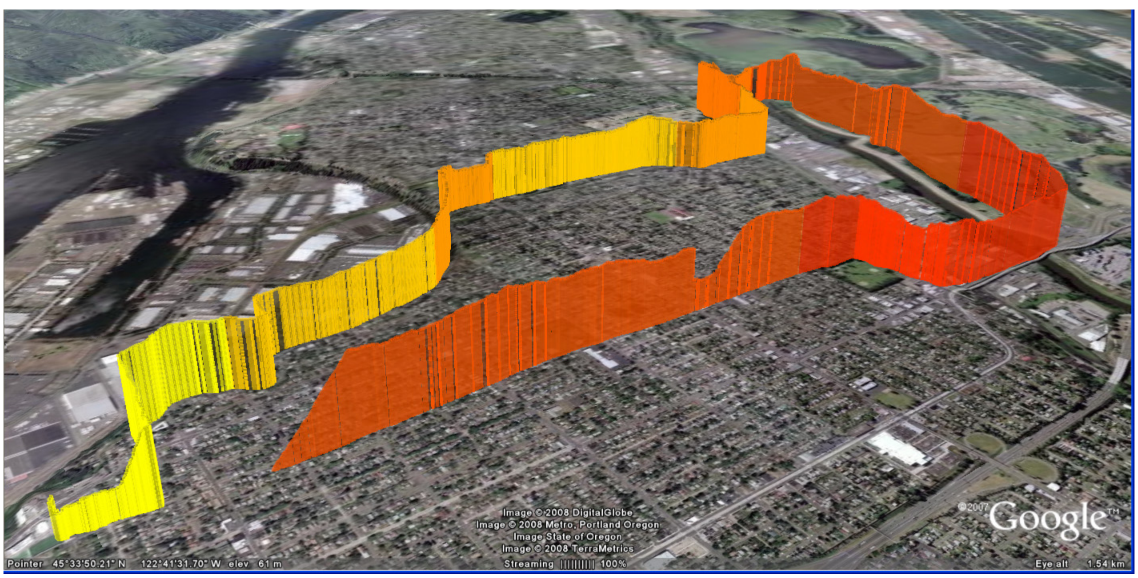

Fig. 7.5. Visualization of running speed for a 1 hour example run in Portland, Oregon, USA.

7.7 Conclusion and further work

We designed and realized a system for collecting real-time subjective, physiological and other information about a sports session. For the implementation we made use of the advanced computational power and the multimedia capabilities of mobile phones, which offer a high adaptability through software packages tailored to the problem at hand. Our system is capable of asking questions about the subjective state of an athlete as defined in a configuration file or as a reaction to external events. Other information like speed and position can be collected via an internal or external GPS receiver. It is also possible to connect other sensors like heart rate monitors using Bluetooth connection.

The system has already proven its usability in practice. The system has been found to be not hindering to the sports activity of running by a majority of 84 athletes. Run information has been collected for an hour for each of the athletes with 100% reliability for the audio information and 99.93% reliability for the position information. The position and other information can very conveniently be visualized using Google Earth. The data of this ongoing study is currently analyzed, the results will be the topic of another presentation.

Our system could also be used for evaluation of other outdoor and endurance sports like rowing, cross-country skiing and biking. It is highly mobile, lightweight and applicable even for long studies due to extendable memory and high battery capacities. To our knowledge, it is the first time that a surveying system has been implemented on a mobile phone.

7.8 Acknowledgements

The authors would like to thank the adidas innovation team *ait*. for the financial and technical support of this research. We also thank the reviewers for their useful comments and suggestions.

CHAPTER 8 SUMMARY AND DISCUSSION

In this chapter, a summary of the thesis work and a general discussion of the methodology is presented. The results corresponding to the research questions that were defined in Chapter 1 and the related hypotheses are summarized first. Then, a general discussion of strengths and weaknesses of the applied methods is conducted, followed by a synopsis of the significance of the findings. Subsequently, suggestions for future work are presented and finally the thesis is concluded with a short overall summary.

8.1 Summary of results

A summary of the results of this thesis with reference to the six research questions (cf. Section 1.3) and corresponding hypotheses is presented below.

8.1.1 Research result one

Research question one was concerned with group classification using generic features for biomechanical data.

In the research work for this thesis, generic feature sets for biomechanical locomotion data were developed. The generic features could be calculated from the ‘analog’ time-dependent locomotion data directly. Two approaches were employed for this purpose. In the biomechanical classification study in Chapter 3, locomotion data was first processed using an inverse dynamics approach to yield kinematic and kinetic time series. From them, generic features were extracted using basic, regression and frequency domain features. In the second biomechanical classification study in Chapter 4, measured data from 3D marker positions was directly used for generic feature computation by utilizing PCA dimensionality reduction.

Both approaches allowed group classification with acceptable classification rates. In Chapter 3, a classification of gender and shod vs. barefoot classes was conducted with the generic features that were developed.

Hypothesis H1 stated that using generic features for biomechanical data, a class-wise mean classification rate of at least 80% is possible for gender classification. In the experiments, a classification rate of 84.7% was obtained for two groups with 40 subjects in each group. This result was significantly different from random ($p < 0.001$).

Hypothesis H2 stated that using generic features for biomechanical data, a class-wise mean classification rate of at least 80% is possible for shod versus barefoot classification. For these groups, a classification rate of 98.3% was obtained. This result was also significantly different from random ($p < 0.001$).

However, gender and shod versus barefoot groups were not the only ones that were examined. In Chapter 3, groups of runners that developed patellofemoral pain syndrome and matched runners that did not during a prospective study were also examined. For these groups, all trials of each runner could be assigned to the correct group. This corresponds to a classification rate of 100%. This result was significantly different from random ($p < 0.001$). In Chapter 4, groups composed of young and elderly subjects could be differentiated with a classification rate of 95.8%. Although not explicitly reported in the study, this result was also significantly different from random.

8.1.2 Research result two

Research question two was concerned with feature selection and ranking methods for biomechanical data. It was examined whether these methods are capable of identifying the measured variables containing the discriminating information.

In Chapter 3, it was shown that this identification is possible and that the reduced feature set still facilitated acceptable classification rates. For the purpose of feature ranking, the AdaBoost classifier was applied in a specific implementation with decision stumps as weak classifiers. These weak classifiers employed thresholds in single feature dimensions for classification. In this process, the feature with the most important contribution to class discrimination was selected. By counting how often a feature contributed to the discrimination, a ranking of the calculated features was facilitated. The results of the study in Chapter 3 showed that using this feature selection and ranking method, previous results that were published in the literature could be confirmed. Additionally, previously unidentified variables that were important for group discrimination could be identified, for example the hip abduction moment with respect to the development of PFPS. A further discussion of the application of feature selection in biomechanical studies is conducted in Section 8.2.1 below.

8.1.3 Research result three

Research question three was concerned with the ability of pattern recognition methods to show high-dimensional dependencies of classes on features for questions of biomechanical group classifications.

It was shown in both biomechanical group classification studies that for most groups that were examined (with the single exception of the patellofemoral pain classification in Chapter 3, see below), combinations of features were needed to achieve acceptable group differentiation. In the study in Chapter 3, three different groups were investigated. For the groups based on gender, it was shown that two features from different kinematic measurements were consistently selected for group differentiation. More specifically, these two variables were the variance of the hip flexion-extension moment and the variance of the vertical ground reaction force (Fig. 3.2). Although other variables were selected as well, these two were required in any case to obtain a high classification rate. Both variables have individually been shown to be different for gender groups (cf. Section 3.4). However, to the knowledge of the author, it was previously not shown that a combination of these variables is beneficial for gender group differentiation. This is a clear advantage of the data mining aspect of the thesis. Exactly the same reasoning is valid for the two features that were consistently selected for shod/barefoot group classification, which were the quadratic polynomial component of the foot sagittal plane angle and the linear polynomial component of the shank sagittal plane angle (Fig. 3.3). The necessity for combination of these features for this differentiation task had, to the knowledge of the author, also not been revealed previously.

The results in Chapter 3 also showed that only one feature (Fig. 3.4) was necessary for perfect group differentiation for runners that did or did not develop patellofemoral pain syndrome. This result illustrates that a high-dimensional combination of features is not necessary if the information for perfect group classification is already present in one variable. Nevertheless, the advantage of the proposed classification methodology is that such single feature is also identified automatically and unbiased. Moreover, only a small population of runners that developed the investigated injury type was available in that particular study (cf. Chapter 3). It is possible (and typical for larger samples) that in a larger population of injured and non-injured subjects, the group classification may again depend on a combination of features. However, it was demonstrated in the other group classification tasks that these combinations of features can be revealed by the chosen pattern recognition approach. It can therefore be speculated that the algorithms that were applied would also successfully identify a combination of features that would be needed for group classification in a larger sample.

In the study in Chapter 4, it was additionally shown that a combination of features leads typically to better classification rates compared to using only single features, which was stated in Hypothesis H3. In that study, which was concerned with young vs. elderly gait classification, the classification rate when using only a single feature (Fig. 4.3) was 62.5%. The maximum classification rate that could be achieved was 95.8% using 36-39 features. This change in classification rate is statistically different ($p < 0.001$). This is in support of Hypothesis H3.

A different novel approach for showing high-dimensional dependencies for biomechanical group differentiation tasks was also presented in this study. With the applied methodology, it was possible to show group differences in gait directly in the original marker space. This was made possible by employing the classification decision boundary, i.e. the information about the largest group differences, as a basis for the movement difference visualization. The groups under investigation in this study were clinically not very relevant, but the proposed difference visualization might be a promising technique for the analysis of pathological gait.

8.1.4 Research result four

Research question four was concerned with the suitability of pattern recognition methods for obtaining acceptable classification rates on embedded microprocessor hardware.

For this purpose, a general methodology was presented (Chapter 5 and Chapter 6) that allowed the implementation of classification algorithms on embedded microprocessors. The considerations were general and applicable to a wide range of embedded classification tasks. Firstly, considerations were made regarding what types of features were suited for implementation on microprocessors. These features needed to be computationally inexpensive but nevertheless they had to represent the information available from the sensor measurements as well as possible. Secondly, requirements that a classifier must fulfill in order to be suitable for a limited performance hardware environment were identified. Feature reduction with the chosen classifier was proposed in order to keep the computational demand for the embedded device as small as possible. All computationally expensive calculations were performed on PC hardware, and only the final solution was then implemented and reevaluated with the embedded microprocessor.

An example task of embedded data classification in sports was investigated and successfully solved. It was shown that the types of features and classification algorithms were able of obtaining acceptable classification rates on an embedded microprocessor. In Chapter 5, the classification of two surface classes was conducted with the methodology that was developed. Hypothesis H4 stated that using this methodology, an on-system classification rate of at least 80% is possible for this task. In the experiments, it was shown that a classification rate of more than 80% could be consistently obtained in different scenarios (Tab. 5.8). Although not explicitly reported, this result is significantly different from random and in support of Hypothesis H4. In Chapter 6, the classification of three speed classes was considered with the same methodology. Hypothesis H5 stated that using this methodology, an on-system classification rate of at least 80% is possible for the speed classification task. In the experiments, it was demonstrated that a classification rate of 89.2% is indeed feasible (Fig. 6.5). This result was shown to be significantly different from random ($p < 0.001$). The result thus supports Hypothesis H5.

8.1.5 Research result five

Research question five was concerned with the ability to effectively test the developed algorithms for microprocessors on the embedded hardware.

For this purpose, the important key concept (cf. Chapter 5 and Chapter 6) was to perform all the comparative experiments using computationally powerful desktop computers. Only the final trained classifier solution that was most promising was then implemented on the embedded hardware and only this implementation was tested. Given that all the considerations that were made in the methodological part were correct, the results of the experiments on the desktop computers and those on the embedded hardware had to be identical. It could then also be assumed that the proposed algorithm represented the optimal solution for the given embedded classification task.

In Chapter 5 and Chapter 6, this methodology was used to verify the results on the embedded microprocessor. Hypothesis H6 stated that by using this methodology the classification rate on the embedded device and the one achieved during testing on a desktop machine leads to the same result. In the experiments conducted in this study, it was shown that both classification rates were indeed identical, which is in support of Hypothesis H6.

8.1.6 Research result six

Research question six studied the ability of an embedded system to collect high-quality data for classification purposes while at the same time being unobtrusive.

An embedded system implementation was developed for this purpose (Chapter 7). It was programmed to specifically perform athlete monitoring, data collection and subsequent information processing. It was designed so that it was applicable in a range of data collection tasks. Its data collection ability was shown in real-world studies, and in these studies it was evidenced that it was not hindering an athlete during sports activity.

During the research work for this thesis, it additionally became apparent that a software tool for visualizing the collected sports data was desirable. Such a tool was implemented and its usability for data visualization was demonstrated. This tool for outdoor sports data visualization and its experimental evaluation was published earlier (Eskofier and Melzer, 2009).

8.2 Strengths and weaknesses

The different methods that were applied in the studies presented in this thesis have several general strengths and weaknesses. First, the pros and cons of the biomechanical group classification studies that form Chapter 3 and Chapter 4 are discussed, followed by the same discussion for the embedded classification papers of Chapter 5 and Chapter 6.

8.2.1 Biomechanical group classification studies

In Chapter 3, successful group classification was reported for gender, shod versus barefoot and injury groups. However, the application of the proposed classification methodology is not restricted to these specific groups. Other group classifications, including clinical or performance groups, can be conducted straightforwardly. For these other group classifications, the classification system as a whole does not have to be changed. It only has to be provided with the group information, i.e. labels have to exist that define the group membership. The classification system can then be retrained using these labels and group differentiability can be tested. The same is true for the classification algorithms in Chapter 4. Clinically more relevant groups can be straightforwardly analyzed if different group labels are available.

In the study in Chapter 3, generic features were calculated using time series of kinematic and kinetic variables as an input to the classification system. However, in Chapter 4, it was argued that a classification system that is directly based on marker measurements has advantages for the analysis of biomechanical gait data. The question why kinematic and kinetic variables were used as basis for the classification study in Chapter 3 therefore needs to be discussed. The disadvantage of a classification system that is only based on markers is that the moment and force distribution within the human body can not be taken into account. Several functional aspects that may, for instance, be a reason for injury development, can therefore not be accounted for. As an example, the result of the study in Chapter 3 showed that group classification was possible with perfect classification rate for groups composed of runners that developed patellofemoral pain syndrome and runners that were asymptomatic. Further investigation revealed that one single feature, the mean value of the hip abduction moment, was sufficient for this group classification. The appropriate manipulation of this variable may therefore play a preventive role for patients who are predisposed to patellofemoral pain. In order to generate this clinically relevant statement, it is required to include the moment and forces distribution within the human body. Whether to use the marker-based or the kinematics/kinetics based approach is therefore dependent on the specific research task. In Chapter 3, the latter approach was deliberately chosen, also because the marker-based methodology may be helpful in future studies (Section 8.4) relating to the direct analysis of gait.

The study that was included in Chapter 4 claimed that anthropometric differences were removed by the analysis procedure that was described there. The mean position from each marker was removed in order to only use the positional variation of each marker for the analysis. However, all aspects of anthropometric differences could not be claimed to be removed by this procedure. The variation of each marker position might also be affected by anthropometric differences. For example, a tall person might exhibit more variation in the general hand movement than a shorter one. Nevertheless, the main aspects of height difference could certainly be claimed to be considered by the mean marker position removal. Furthermore, the difference in positional variation that was originating from aspects of anthropometric differences could be assumed to be small, and they could not explain the high classification rates that could be obtained in Chapter 4 alone.

A shortcoming of the included biomechanical group classification studies can be seen in the fact that just one specific feature selection and ranking strategy was applied. However, the important point to note is not what feature selection is applied, but that some selection scheme should be considered in a classification study. This is because feature selection reduces the computational complexity of the classifier (Begg and Kamruzzaman, 2005) and can lead to better classification rates (Wu et al., 2007). Furthermore, the result of Chapter 3 showed that feature selection can be employed for pointing out group discriminating variables, for example for sport injuries. Such variables that have been identified to discriminate pathological from asymptomatic subjects can then be used for further investigation into the causes for developing a specific injury. In order to obtain these advantages, it is therefore advisable for researchers to consider at least one of the available feature selection and ranking strategies (cf. Section 2.1.3.3) in a biomechanical group classification study.

The last point about the functional interpretation of the reasons for group differentiation also indicates a different shortcoming of the biomechanical studies that were presented in Chapter 3 and Chapter 4. The methodologies that were proposed were all aimed at an information representation that was as complete as possible. This approach, which is typical for pattern recognition systems, has the advantage that all important aspects are included. The group classification rates that can be obtained are therefore typically high. However, in this pattern recognition approach, the functional understanding of the reasons for group differentiation may not be obvious. This is because the outcomes of the pattern recognition approach may be variables that are difficult to interpret. This interpretation may be easier when individual variables with a more functional meaning (i.e. knee angle at toe off, hip moment at mid stance) are used for analysis. This more traditional approach provides group identification factors that are typically easy to interpret. The disadvantage of the more traditional approach is that the researcher never knows whether the right variables have been selected. Furthermore, the group identification levels are frequently found to be low. In summary, it can be said that both approaches have complementary strengths and weaknesses. The introduction of techniques from pattern recognition into biomechanical analysis procedures might therefore be considered as valuable additional tools to gain this complementary information.

8.2.2 Embedded classification studies

In the embedded classification studies (Chapter 5 and Chapter 6), an approach for microprocessor implementation of pattern recognition algorithms was presented. This approach was constructed to be generally applicable. Nevertheless, it needs some adaptation when transferred to a different embedded system. This is because certain specific considerations must be made for every pattern recognition system (cf. Section 2.1), which is independent of the embedded classification task. Possible changes may be required for the preprocessing methods, the applied features and the selected classifiers. Thus, the presented approach is not directly applicable in every conceivable application. However, the core ideas that were discussed play a role in any application on embedded hardware, and they are transferable to general embedded classification tasks. More specifically, the design of manual features allows embedded classification systems to be computationally simple, while still representing the available information in the measured sensor signals adequately. Feature selection allows identifying a set of features for implementation on the embedded microprocessor that can be computed in real-time and obtains high classification rates. Finally, the selection of classifiers that are suited for embedded implementation and their comparison with regard to the obtainable classification performance allows a sound embedded classification system design.

The manual feature generation that was employed in the embedded classification studies in Chapter 5 and Chapter 6 has some disadvantages that require further discussion. Manual feature design is more time intensive than generic feature extraction approaches (cf. Section 2.1.3.2). Furthermore, manually designed features are specific to a given application and not straightforwardly transferable like generic features. In some scenarios where computationally more powerful microprocessors are employed, it is conceivable to use generic feature extraction strategies like Fourier transformations or Principal Component Analysis. However, for computationally less powerful systems, generic feature extraction strategies are often prohibitive. By following the general considerations made in this thesis, the additional work for manual feature design can be minimized. With the availability of these designed features, their benefits (i.e. their low computational complexity and dense information representation) can be exploited for embedded classification.

Another shortcoming of the work done in this thesis can be seen in the fact that the applicability of the developed embedded classification strategy was only shown in a single real world application, the ‘adidas_1’ shoe (Chapter 5 and Chapter 6). Although different groups were considered in the related studies, the input sensor signal and hardware environment was always the same. However, the purpose of this thesis was focused on reporting the general ideas for embedded classification system implementation. These ideas are straightforward to apply to various different embedded classification tasks. Indeed, the developed methodology has already been employed in other application scenarios than the ‘adidas_1’ shoe, e.g. for fatigue classification in a mobile application (Horz, 2008), for the classification of user interaction in household appliances (Haas, 2008) and for movement classification based on gyroscopic sensor data from runners (Tüxen, 2009). In all these scenarios, the general ideas that were presented in Chapter 5 and Chapter 6 were applied and the classification tasks were successfully solved.

The classification rates that have been reported in the embedded studies were acceptable for the questions at hand. Since the embedded classification systems were already employed in real-word application scenarios, there is of course a desire to further improve these results. It was, however, not possible to improve the classification rate given the embedded system presented in Chapter 5 and Chapter 6. Nevertheless, only one sensor was used for classification in these studies. Better classification rates can therefore be expected when data from multiple sensors becomes available. This has also been evidenced in the above-mentioned additional embedded classification studies (Haas, 2008; Horz, 2008). In these studies that utilized the methodology presented in this thesis, it was shown that better classification rates than the ones reported in Chapter 5 and Chapter 6 could be obtained due to the availability of more sensor information.

The data collection study that was included in Chapter 7 was not a dedicated classification study. At first view, it could be argued that the content of this particular chapter does not fit fully into the thesis. However, when regarding the classification flow chart (Fig. 1.1) it becomes obvious that the data input for classification is an integral part of any pattern recognition system. High classification rates can not be obtained without first collecting high-quality data for the classification task. In the light of this argument, a study that specifically targeted data collection using embedded systems for locomotion studies was an integral part of the discussion of the complete classification procedure that was conducted in this thesis.

8.3 Significance of findings

The attempt of this thesis was to contribute to the body of knowledge regarding group classification including:

1. Different strategies for feature generation and selection, both in the context of biomechanical and embedded group classification.
2. The demonstration of the ability of various classifiers to produce significant classification rates in different biomechanical group classification tasks.
3. The identification of specific combinations of features needed in order to differentiate groups based on biomechanical data.
4. The demonstration of a single specific feature that is an important indicator for the classification of patellofemoral pain.
5. The introduction of a method to use the resulting classification boundary for visualizing gait group differences directly.
6. The demonstration of the benefits of a comparative analysis of different feature extraction and classification methods for embedded classification.
7. A discussion how to implement the complete classification pipeline on restricted embedded hardware environments.
8. The development of implementation strategies for employing mobile embedded devices for self-sufficient data collection from athletes.

Overall, the thesis answered new, important questions. The results have the potential to be implemented in future embedded classification applications and in analysis strategies for biomechanical studies. To the best of the author’s knowledge, the discussed aspects were either only partially addressed or never addressed in the previous literature.

8.4 Future work

Several ideas for future studies resulted from the research for this thesis. First, it was demonstrated that indicators for sport specific injuries can be identified when applying pattern recognition methods to prospective biomechanic data sets (Chapter 3). One of the intended goals of future work is the evaluation of the proposed methods based on kinematic/kinetic variables on further studies including:

1. The identification of injury probability due to anatomical characteristics. For this purpose, groups containing cases of defined clinically relevant groups and asymptomatic subjects need to be analyzed.
2. The development of an injury severity indicator. So far, due to the limited amount of data, it was only attempted to classify two groups (injured or not injured). Having more data available, the severity of the injury could additionally be taken into account in the classification.

The gait pattern classification based on direct feature calculation from 3D marker positions in Chapter 4 will also be applied to clinically more relevant groups. The next steps in this research direction will be:

1. The analysis of the effect of cognitive tasks on gait. Cognitive and motor control tasks that have to be executed at the same time have already been shown to influence balance and stability (Huxhold et al., 2006; Lindenberger et al., 2000). Gait data for such dual tasking situations has already been collected at the Human Performances Lab. The analysis method presented in Chapter 4 is planned to be used for the further evaluation of this study.
2. The classification of clinically more relevant groups. So far, groups composed of elderly and young subjects were analyzed. However, the walking gait classification method can also be applied to pathological gait groups, for

example cerebral palsy gait. The gait differences in pathological groups vs. healthy subject groups can also be visualized using the methodology in Chapter 4.

3. The employment of nonlinear relationships for gait classification. The study in Chapter 4 used only linear relationships for group differentiation. However, the applied PCA and SVM algorithms can both take nonlinear relationships into account by employing nonlinear kernel functions (Schölkopf and Smola, 2002; Wu et al., 2007). The application of such methods can be speculated to increase the classification result further.

Future work is also intended on aspects of embedded classification. Microprocessors are already being used in a manifold of sports and biomechanics related applications (Baca et al., 2009). The research work on embedded classification that was performed in this thesis (Chapter 5 and Chapter 6) is intended to be further employed in applications in:

1. Real-time classification systems for athlete support. A concrete example is the classification of the fatigue state of an athlete.
2. Embedded classification systems in areas outside of sports. An example is the classification on microprocessors in household appliances. Manufacturers of household appliances have a desire to prevent dangerous situations during the operation of their devices. For instance, children should not be able to operate electrical ovens. The recognition of these potentially dangerous states can reduce the risk for consumers to sustain damage.

It is also planned to further employ embedded systems for monitoring purposes. A software solution tailored for mobile phones for a running data collection study was already presented in Chapter 7. Similar systems could be used in:

1. The evaluation of other outdoor and endurance sports. It is currently planned to extend the application to rowing, skiing and biking.

2. The monitoring of risk patients. In Chapter 7, a mobile phone provided the basis for the implementation of the data collection tool. It is intended to further exploit the networking capabilities of the mobile phone for the purpose of, for instance, sending an alarm to a monitoring station in the case of an emergency.

8.5 Summary

This thesis had two main purposes.

The first purpose was to show the applicability of pattern recognition methods to biomechanical data, which was demonstrated with two different studies that resulted in high classification rates for different group classification tasks. Two possible methods to generate generic features for biomechanical data were presented, both from kinetic/kinematic measurements and directly from marker position data. Furthermore, it was demonstrated that pattern recognition methods are capable of identifying what combinations of features are needed to differentiate the groups. Additionally, it was shown that the resulting difference information can be visualized for the purpose of further analysis.

The second purpose of this thesis was to discuss a general methodology for embedded classification. It was demonstrated that manual feature design, feature selection and a careful choice and implementation of classifiers for implementation are important factors for embedded classification. Using the proposed approach, the ability to produce acceptable results in several sports biomechanics related classification tasks was demonstrated.

REFERENCES

ABELE, A. & BREHM, W. 1985. Einstellungen zum sport, präferenzen für das eigene sporttreiben und befindlichkeitsveränderungen nach sportlicher aktivität. *Psychologie in Erziehung und Unterricht,* 32(1), 263-270.

ACEVEDO, E. O., GILL, D. L., GOLDFARB, A. H. & BOYER, B. T. 1996. Affect and perceived exertion during a two-hour run. *International Journal of Sport Psychology,* 27(3), 286-292.

AHMED, N., NATARAJAN, T. & RAO, K. R. 1974. Discrete cosine transfom. *IEEE Transactions on Computers,* C-23(1), 90-93.

ALDROUBI, A. & UNSER, M. 1996. *Wavelets in medicine and biology,* Boca Raton, FL, CRC-Press.

AMINIAN, K., ROBERT, P., JÉQUIER, E. & SCHUTZ, Y. 1993. Level, downhill and uphill walking identification using neural networks. *Electronics Letters,* 29(17), 1563-1565.

AMINIAN, K., ROBERT, P., JÉQUIER, E. & SCHUTZ, Y. 1995. Estimation of speed and incline of walking using neural network. *IEEE Transactions on Instrumentation and Measurement,* 44(3), 743-746.

ARABSHAHI, S. A. & JIANG, Z. 2005. Development of a tactile sensor for braille pattern recognition: Sensor design and simulation. *Smart Materials and Structures,* 14(6), 1569-1578.

ARMSTRONG, S. 2007. Wireless connectivity for health and sports monitoring: A review. *British Journal of Sports Medicine,* 41(5), 285-289.

ASSFALG, J., BERTINI, M., COLOMBO, C. & DEL BIMBO, A. 2002. Semantic annotation of sports videos. *IEEE Multimedia,* 9(2), 52-60.

BACA, A., DABNICHKI, P., HELLER, M. & KORNFEIND, P. 2009. Ubiquitous computing in sports: A review and analysis. *Journal of Sports Sciences,* 27(12), 1335-1346.

BARCLAY, C. D., CUTTING, J. E. & KOZLOWSKI, L. T. 1978. Temporal and spatial factors in gait perception that influence gender recognition. *Perception and Psychophysics,* 23(2), 145-152.

BARTON, J. G. & LEES, A. 1995. Development of a connectionist expert system to identify foot problems based on under-foot pressure patterns. *Clinical Biomechanics,* 10(7), 385-391.

BARTON, J. G. & LEES, A. 1997. An application of neural networks for distinguishing gait patterns on the basis of hip-knee joint angle diagrams. *Gait and Posture,* 5(1), 28-33.

BAXT, W. G. 1995. Application of artificial neural networks to clinical medicine. *Lancet,* 346(8983), 1135-1138.

BEGG, R. & KAMRUZZAMAN, J. 2005. A machine learning approach for automated recognition of movement patterns using basic, kinetic and kinematic gait data. *Journal of Biomechanics,* 38(3), 401-408.

BEGG, R. K., PALANISWAMI, M. & OWEN, B. 2005. Support vector machines for automated gait classification. *IEEE Transactions on Biomedical Engineering,* 52(5), 828-838.

BEKEY, G. A., CHI-WU, C., PERRY, J. & HOFFER, M. M. 1977. Pattern recognition of multiple EMG signals applied to the description of human gait. *Proceedings of the IEEE,* 65(5), 674-681.

BENBASAT, A. Y. & PARADISO, J. A. 2002. An inertial measurement framework for gesture recognition and applications. Revised Papers from the International Gesture Workshop on Gesture and Sign Languages in Human-Computer Interaction, 2002. Springer, 9-20.

BENBASAT, A. Y. & PARADISO, J. A. 2004. Design of a real-time adaptive power optimal system. *In:* ROCHA, D., SARRO, P. M. & VELLEKOOP, M. J., eds. Proceedings of the 3rd IEEE International Conference on Sensors (IEEE Sensors 2004), October 24-27 2004, Vienna. 48-51.

BERTANI, A., CAPPELLO, A., BENEDETTI, M. G., SIMONCINI, L. & CATANI, F. 1999. Flat foot functional evaluation using pattern recognition of ground reaction data. *Clinical Biomechanics,* 14(7), 484-493.

BICCIATO, S., LUCHINI, A. & DI BELLO, C. 2003. PCA disjoint models for multiclass cancer analysis using gene expression data. *Bioinformatics,* 19(5), 571-578.

BISHOP, C. 1996. *Neural networks for pattern recognition,* Oxford, Oxford University Press.

BISIANI, R. 1987. Beam search. *In:* SHAPIRO, S. C. (ed.) *Encyclopedia of artificial intelligence.* 1st ed. New York, NY: Wiley & Sons.

BLANKE, D. J. & HAGEMAN, P. A. 1989. Comparison of gait of young men and elderly men. *Physical Therapy,* 69(2), 144-148.

BLANZIERI, E. & MELGANI, F. 2008. Nearest neighbor classification of remote sensing images with the maximal margin principle. *IEEE Transactions on Geoscience and Remote Sensing,* 46(6), 1804-1811.

BOLAND, P. J. 1989. Majority systems and the condorcet jury theorem. *Journal of the Royal Statistical Society. Series D (The Statistician),* 38(3), 181-189.

BONI, A., GASPARINI, L., PIANEGIANI, F. & PETRI, D. 2005. Low-power and low-cost implementation of SVMs for smart sensors. Proceedings of the IEEE Instrumentation and Measurement Technology Conference, May 16-19 2005, Ottawa, ON. 603-607.

BOTROS, N. M. 1988. A microprocessor-based pattern recognition algorithm for in-vivo tissue differentiation. *Journal of Clinical Engineering,* 13(2), 115-120.

BUCKLEY, T., PITSIKOULIS, C., BARTHELEMY, E. & HASS, C. J. 2009. Age impairs sit-to-walk motor performance. *Journal of Biomechanics,* 42(14), 2318-2322.

BURGES, C. J. C. 1998. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery,* 2(2), 121-167.

CHANG, C. & LIN, C. 2001. Libsvm: A library for support vector machines. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

CHAPELLE, O., HAFFNER, P. & VAPNIK, V. N. 1999. Support vector machines for histogram-based image classification. *IEEE Transactions on Neural Networks,* 10(5), 1055-1064.

CHAU, T. 2001a. A review of analytical techniques for gait data. Part 1: Fuzzy, statistical and fractal methods. *Gait and Posture,* 13(1), 49-66.

CHAU, T. 2001b. A review of analytical techniques for gait data. Part 2: Neural network and wavelet methods. *Gait and Posture,* 13(2), 102-120.

CHAZAL, P. D., O'DWYER, M. & REILLY, R. B. 2004. Automatic classification of heartbeats using ecg morphology and heartbeat interval features. *Biomedical Engineering, IEEE Transactions on,* 51(7), 1196-1206.

CHEN, H. C., ASHTON-MILLER, J. A., ALEXANDER, N. B. & SCHULTZ, A. B. 1991. Stepping over obstacles: Gait patterns of healthy young and old adults. *Journals of Gerontology,* 46(6), 196-203.

CHEN, K. & KURGAN, L. 2007. Pfres: Protein fold classification by using evolutionary information and predicted secondary structure. *Bioinformatics,* 23(21), 2843-2850.

CHILDERS, D. G. & WU, K. 1991. Gender recognition from speech. Part ii: Fine analysis. *Journal of the Acoustical Society of America,* 90(4), 1841-1856.

CHRISTODOULOU, C. I. & PATTICHIS, C. S. 1999. Unsupervided pattern recognition for the classification of EMG signals. *IEEE Transactions on Biomedical Engineering,* 46(2), 169-178.

CHUA, L. O. & YANG, L. 1988. Cellular neural networks: Applications. *IEEE Transactions on Circuits and Systems,* 35(10), 1273-1290.

COVER, T. & HART, P. 1967. Nearest neighbor pattern classification. *Information Theory, IEEE Transactions on,* 13(1), 21-27.

CUTTING, J. E., PROFFITT, D. R. & KOZLOWSKI, L. T. 1978. A biomechanical invariant for gait perception. *Journal of Experimental Psychology: Human Perception and Performance,* 4(3), 357-372.

DASH, M. & LIU, H. 1997. Feature selection for classification. *Intelligent Data Analysis,* 1(3), 131-156.

DAUBECHIES, I. 1992. *Ten lectures on wavelets,* Philadelphia, PA, Society for Industrial and Applied Mathematics (SIAM).

DE BOOR, C. 1978. *A practical guide to splines,* New York, NY, Springer.

DE WIT, B., DE CLERCQ, D. & AERTS, P. 2000. Biomechanical analysis of the stance phase during barefoot and shod running. *Journal of Biomechanics,* 33(3), 269-278.

DEKRUGER, D. & HUNT, B. R. 1994. Image processing and neural networks for recognition of cartographic area features. *Pattern Recognition,* 27(4), 461-483.

DEVEREAUX, M. D. & LACHMANN, S. M. 1984. Patello-femoral arthralgia in athletes attending a sports injury clinic. *British Journal of Sports Medicine,* 18(1), 18-21.

DEVITA, P. & HORTOBAGYI, T. 2000. Age causes a redistribution of joint torques and powers during gait. *Journal of Applied Physiology,* 88(5), 1804-1811.

DIBENEDETTO, C., OLESON, M. A., ROTH, C. & THOMPSON, M. C. 2004. *Intelligent footwear systems*. United States patent application US20040177531.

DIBENEDETTO, C., OLESON, M. A., ROTH, C. & THOMPSON, M. C. 2005. *Intelligent footwear systems*. European patent application EP1582108.

DOMINGOS, P. & PAZZANI, M. 1997. On the optimality of the simple bayesian classifier under zero-one loss. *Machine Learning,* 29(2), 103-130.

DRAPER, B. A., BAEK, K., BARTLETT, M. S. & BEVERIDGE, J. R. 2003. Recognizing faces with PCA and ica. *Computer Vision and Image Understanding,* 91(1), 115-137.

DRUCKER, H., WU, D. & VAPNIK, V. N. 1999. Support vector machines for spam categorization. *IEEE Transactions on Neural Networks,* 10(5), 1048-1054.

DUDA, R. O. & HART, P. E. 1973. *Pattern classification and scene analysis,* New York, NY, Wiley & Sons.

DUDA, R. O., HART, P. E. & STORK, D. G. 2001. *Pattern classification,* New York, NY, Wiley & Sons.

ELBLE, R. J., THOMAS, S. S., HIGGINS, C. & COLLIVER, J. 1991. Stride-dependent changes in gait of older people. *Journal of Neurology,* 238(1), 1-5.

ENGLEHART, K. & HUDGINS, B. 2003. A robust, real-time control scheme for multifunction myoelectric control. *IEEE Transactions on Biomedical Engineering,* 50(7), 848-854.

ESKOFIER, B. M., FEDEROLF, P., KUGLER, P. & NIGG, B. M. 2010a. Marker-based classification of young-elderly gait pattern differences via direct PCA feature extraction and SVMs. Submitted to: *Journal of Biomechanics*.

ESKOFIER, B. M., HARTMANN, E., KÜHNER, P., GRIFFIN, J., SCHLARB, H., SCHMITT, M. & HORNEGGER, J. 2008a. Real time surveying and monitoring of athletes using mobile phones and GPS. *International Journal of Computer Science in Sports,* 7(1), 18-27.

ESKOFIER, B. M., HOENIG, F. & KUEHNER, P. 2008b. Classification of perceived running fatigue in digital sports. Proceedings of the 19th International Conference on Pattern Recognition (ICPR 2008), December 8-11 2008b, Tampa, FL. No pagination.

ESKOFIER, B. M., HORNEGGER, J., OLESON, M., MUNSON, I., KRABBE, B. & DIBENEDETTO, C. 2008c. Classification of running surface on an embedded system - a digital sports example application. *In:* THÖMMES, T., WESSEL, N. & WOLF, W., eds. Innovationen bei der Erfassung und Analyse bioelektrischer und biomagnetischer Signale (Biosignalverarbeitung 2008), July 16-18 2008c, Potsdam. Berlin: Physikalisch-Technische Bundesanstalt, 147-150.

ESKOFIER, B. M., KORNHUBER, J. & HORNEGGER, J. 2008d. Embedded qrs detection for noisy ecg sensor data using a matched filter and directed graph search. *In:* BAUERNSCHMITT, R., CHAPLYGIN, Y., FEUßNER, H., GULYEAV, Y., HORNEGGER, J., MAYR, E., NAVAB, N., SCHOOKIN, S., SELISHCHEV, S. & UMNYASHKIN, S., eds. 4th Russian-Bavarian Conference on Biomedical Engineering, July 8-9 2008d, Zelenograd. No pagination.

ESKOFIER, B. M., KRAUS, M., WOROBETS, J. T., STEFANYSHYN, D. J. & NIGG, B. M. 2010b. Pattern classification of kinematic and kinetic running data to distinguish gender, shod/barefoot and injury groups with feature ranking. Submitted to: *Computer Methods in Biomechanics and Biomedical Engineering*.

ESKOFIER, B. M. & MELZER, D. 2009. Gervit - a tool for sports data visualization employing Google Earth. *International Journal of Computer Science in Sports,* 8(2), 32-39.

ESKOFIER, B. M., OLESON, M., DIBENEDETTO, C. & HORNEGGER, J. 2009a. Embedded surface classification in digital sports. *Pattern Recognition Letters,* 30(16), 1448-1456.

ESKOFIER, B. M., WAGNER, M., MUNSON, I. & OLESON, M. 2009b. Classification of changes in speed and inclination during running. International Symposium on Computer Science in Sports (IACSS 2009), September 22-25 2009b, Canberra. No pagination.

ESKOFIER, B. M., WAGNER, M., MUNSON, I. & OLESON, M. 2010c. Embedded classification of speed and inclination during running. Submitted to: *International Journal of Computer Science in Sports*.

FAN, Y. & WANG, R. 2002. An image retrieval method using dct features. *Journal of Computer Science and Technology,* 17(6), 865-873.

FISHER, R. A. 1936. The use of multiple measurements in taxonomic problems. *Annals of Eugenics,* 7(2), 179-188.

FRANK, E. & WITTEN, I. H. 1998. Generating accurate rule sets without global optimization. *In:* SHAVLIK, J., ed. Proceedings of the 15th International Conference on Machine Learning, July 24-27 1998, Madison, WI. Morgan Kaufmann Publishers, 144-151.

FRANKE, J. 1997. A comparison of gaussian distribution and polynomial classifiers in a hidden markov model based system for the recognition of cursive script. *In:* GLOGER, J. M., KALTENMEIER, A. & MANDLER, E., eds., 1997. 515-515.

FREDERICSON, M., COOKINGHAM, C. L., CHAUDHARI, A. M., DOWDELL, B. C., OESTREICHER, N. & SAHRMANN, S. A. 2000. Hip abductor weakness in distance runners with iliotibial band syndrome. *Clinical Journal of Sport Medicine,* 10(3), 169-175.

FREUND, Y. & SCHAPIRE, R. E. 1996. Experiments with a new boosting algorithm. Proceedings of the 13th International Conference on Machine Learning, July 3-6 1996, Bari. 148-156.

FREUND, Y. & SCHAPIRE, R. E. 1997. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences,* 55(1), 119-139.

FUKUNAGA, K. 1990. *Introduction to statistical pattern recognition,* San Diego, CA, Academic Press.

FURUI, S. 1986. Speaker-independent isolated word recognition using dynamic features of speech spectrum. *IEEE Transactions on Acoustics, Speech, and Signal Processing,* 34(1), 52-59.

FURUI, S. 2004. Fifty years of progress in speech and speaker recognition. *The Journal of the Acoustical Society of America,* 16(4), 2497-2498.

GRIMMER, R., ESKOFIER, B., SCHLARB, H. & HORNEGGER, J. 2009. Comparison and classification of 3d objects surface point clouds on the example of feet. *Machine Vision and Applications*, Article in press.

HAAS, W. 2008. *Innovative operator systems - classification of user interaction using embedded systems.* Computer Science Master Thesis, University of Erlangen-Nuremberg.

HACKER, C., BATLINER, A. & NÖTH, E. 2006. Are you looking at me, are you talking with me: Multimodal classification of the focus of attention. 9th International Conference on Text, Speech and Dialogue (TSD 2006), September 11-15 2006, Brno. 581-588.

HAFED, Z. M. & LEVINE, M. D. 2001. Face recognition using the discrete cosine transform. *International Journal of Computer Vision,* 43(3), 167-188.

HAGAN, M. T. & MENHAJ, M. B. 1994. Training feedforward networks with the marquardt algorithm. *IEEE Transactions on Neural Networks,* 5(6), 989-993.

HALL, E. H. 1879. On a new action of the magnet on electric currents. *American Journal of Mathematics,* 2(3), 287-292.

HALLBERG, J., SVENSSON, S., ÖSTMARK, A., LINDGREN, P., SYNNES, K. & DELSING, J. 2004. Enriched media-experience of sport events. Proceedings of the 6th IEEE Workshop on Mobile Computing Systems and Applications (WMCSA'04), December 2-3 2004, Windermere. 2-9.

HARALICK, R. M., SHANMUGAM, K. & DINSTEIN, I. 1973. Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics,* 3(6), 610-621.

HERREN, R., SPARTI, A., AMINIAN, K. & SCHUTZ, Y. 1999. The prediction of speed and incline in outdoor running in humans using accelerometry. *Medicine and Science in Sports and Exercise,* 31(7), 1053-1059.

HOLZREITER, S. H. & KOHLE, M. E. 1993. Assessment of gait patterns using neural networks. *Journal of Biomechanics,* 26(6), 645-651.

HORZ, V. 2008. *Analysis and classification of fused sensor data from a running study.* Mechatronics Thesis, University of Erlangen-Nuremberg.

HSU, C. W. & LIN, C. J. 2002. A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks,* 13(2), 415-425.

HUBERT, M. & ENGELEN, S. 2004. Robust PCA and classification in biosciences. *Bioinformatics,* 20(11), 1728-1736.

HUNT, K. J., SBARBARO, D., ZBIKOWSKI, R. & GAWTHROP, P. J. 1992. Neural networks for control systems - a survey. *Automatica,* 28(6), 1083-1112.

HUXHOLD, O., LI, S. C., SCHMIEDEK, F. & LINDENBERGER, U. 2006. Dual-tasking postural control: Aging and the effects of cognitive demand in conjunction with focus of attention. *Brain Research Bulletin,* 69(3), 294-305.

JAMES, S. 1995. Running injuries to the knee. *Journal of the American Academy of Orthopaedic Surgeons,* 3(6), 309-318.

JANSSEN, D., SCHÖLLHORN, W. I., LUBIENETZKI, J., FÖLLING, K., KOKENGE, H. & DAVIDS, K. 2008. Recognition of emotions in gait patterns by means of artificial neural nets. *Journal of Nonverbal Behavior,* 32(2), 79-92.

JAREMKO, J. L., PONCET, P., RONSKY, J., HARDER, J., DANSEREAU, J., LABELLE, H. & ZERNICKE, R. F. 2002. Genetic algorithm-neural network estimation of cobb angle from torso asymmetry in scoliosis. *Journal of Biomechanical Engineering,* 124(5), 496-503.

KARHUNEN, K. 1947. Über lineare methoden in der wahrscheinlichkeitsrechnung. *Annales Academiae Scientiarum Fennicae, Series A1: Mathematica-Physica,* 37(1), 3-79.

KERR, B. A., BEAUCHAMP, L., FISHER, V. & NEIL, R. 1983. Footstrike patterns in distance running. *In:* NIGG, B. M. & KERR, B. A. (eds.) *Biomechanical aspects of sport shoes and playing surfaces.* 1st ed. Calgary, AB: University Printing Calgary.

KERRICK, D. D. & BOVIK, A. C. 1988. Microprocessor-based recognition of handprinted characters from a tablet input. *Pattern Recognition,* 21(5), 525-537.

KERRIGAN, D. C., TODD, M. K. & DELLA CROCE, U. 1998. Gender differences in joint biomechanics during walking: Normative study in young adults. *American Journal of Physical Medicine and Rehabilitation,* 77(1), 2-7.

KOHONEN, T. 1990. The self-organizing map. *Proceedings of the IEEE,* 78(9), 1464-1480.

KOPRINSKA, I., POON, J., CLARK, J. & CHAN, J. 2007. Learning to classify e-mail. *Information Sciences,* 177(10), 2167-2187.

KUDO, M. & SKLANSKY, J. 2000. Comparison of algorithms that select features for pattern classifiers. *Pattern Recognition,* 33(1), 25-41.

LAI, D. T. H., LEVINGER, P. T., BEGG, R. K., GILLEARD, W. L. & PALANISWAMI, M. 2009. Automatic recognition of gait patterns exhibiting patellofemoral pain syndrome using a support vector machine approach. *IEEE Transactions on Information Technology in Biomedicine,* 13(5), 810-817.

LANGLEY, P., IBA, W. & THOMPSON, K. 1992. An analysis of bayesian classifiers. Proceedings of the 10th National Conference on Artificial Intelligence (AAAI-92), July 12-16 1992, San Jose, CA. 223-228.

LANGLEY, R. B. 1995. Nmea 0183: A GPS receiver interface standard. *GPS World,* 6(7), 54-57.

LAUKKANEN, R. M. T. & VIRTANEN, P. K. 1998. Heart rate monitors: State of the art. *Journal of Sports Sciences,* 16(1), 3-7.

LEE, Y. 1991. Handwritten digit recognition using k nearest-neighbor, radial-basis function, and backpropagation neural networks. *Neural Computation,* 3(3), 440-449.

LEE, Y., LIN, Y. & WAHBA, G. 2004. Multicategory support vector machines: Theory and application to the classification of microarray data and satellite radiance data. *Journal of the American Statistical Association,* 99(465), 67-81.

LI, S. Z. & ZHANG, Z. Q. 2004. Floatboost learning and statistical face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 26(9), 1112-1123.

LINDENBERGER, U., MARSISKE, M. & BALTES, P. B. 2000. Memorizing while walking: Increase in dual-task costs from young adulthood to old age. *Psychology and Aging,* 15(3), 417-436.

LIU, C.-L. & SAKO, H. 2006. Class-specific feature polynomial classifier for pattern classification and its application to handwritten numeral recognition. *Pattern Recognition,* 39(4), 669-681.

LIU, M. 2010. Fingerprint classification based on adaboost learning from singularity features. *Pattern Recognition,* 43(3), 1062-1070.

LIU, M. M., HERZOG, W. & SAVELBERG, H. H. C. M. 1999. Dynamic muscle force predictions from EMG: An artificial neural network approach. *Journal of Electromyography and Kinesiology,* 9(6), 391-400.

LOÈVE, M. 1977. *Probability theory,* New York, NY, Springer.

LOZANOVA, S. & ROUMENIN, C. 2010. Angular position device with 2d low-noise hall microsensor. *Sensors and Actuators, A: Physical*, Article in Press.

LU, J., PLATANIOTIS, K. N. & VENETSANOPOULOS, A. N. 2003. Face recognition using LDA-based algorithms. *IEEE Transactions on Neural Networks,* 14(1), 195-200.

LUN, V., MEEUWISSE, W. H., STERGIOU, P. & STEFANYSHYN, D. 2004. Relation between running injury and static lower limb alignment in recreational runners. *British Journal of Sports Medicine,* 38(5), 576-580.

MAHALANOBIS, P. C. 1936. On the generalized distance in statistics. *Proceedings of the National Institute of Sciences of India,* 2(1), 49-55.

MALEK, M. M. & MANGINE, R. E. 1981. Patellofemoral pain syndromes: A comprehensive and conservative approach. *Journal of Orthopaedic and Sports Physical Therapy,* 2(3), 108-116.

MALHI, A. & GAO, R. X. 2004. PCA-based feature selection scheme for machine defect classification. *IEEE Transactions on Instrumentation and Measurement,* 53(6), 1517-1525.

MALLET, Y., COOMANS, D., KAUTSKY, J. & DE VEL, O. 1997. Classification using adaptive wavelets for feature extraction. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 19(10), 1058-1066.

MATHER, G. & MURDOCH, L. 1994. Gender discrimination in biological motion displays based on dynamic cues. *Proceedings of the Royal Society B: Biological Sciences,* 258(1353), 273-279.

MESSIER, S. P. & PITTALA, K. A. 1988. Etiologic factors associated with selected running injuries. *Medicine and Science in Sports and Exercise,* 20(5), 501-505.

MEZGHANI, N., HUSSE, S., BOIVIN, K., TURCOT, K., AISSAOUI, R., HAGEMEISTER, N. & DE GUISE, J. A. 2008. Automatic classification of asymptomatic and osteoarthritis knee gait patterns using kinematic data features and the nearest neighbor classifier. *IEEE Transactions on Biomedical Engineering,* 55(3), 1230-1232.

MOISIO, K. C., SUMNER, D. R., SHOTT, S. & HURWITZ, D. E. 2003. Normalization of joint moments during gait: A comparison of two techniques. *Journal of Biomechanics,* 36(4), 599-603.

MORI, S., SUEN, C. Y. & YAMAMOTO, K. 1992. Historical review of ocr research and development. *Proceedings of the IEEE,* 80(7), 1029-1058.

NARENDRA, P. M. & FUKUNAGA, K. 1977. A branch and bound algorithm for feature subset selection. *IEEE Transactions on Computers,* 26(9), 917-922.

NIEMANN, H. 1983. *Klassifikation von Mustern,* Berlin, Springer.

NIEMANN, H. 2003. *Klassifikation von Mustern,* Berlin, Springer.

NIGG, B. M. 2007. Pressure distribution. *In:* NIGG, B. M. & HERZOG, W. (eds.) *Biomechanics of the musculo-skeletal system.* 3rd ed. New Jersey, NJ: John Wiley & Sons.

NIGG, B. M. & BOYER, K. A. 2007. Acceleration. *In:* NIGG, B. M. & HERZOG, W. (eds.) *Biomechanics of the musculo-skeletal system.* 3rd ed. New Jersey, NJ: John Wiley & Sons.

NIGG, B. M., COLE, G. K. & WRIGHT, I. C. 2007. Optical methods. *In:* NIGG, B. M. & HERZOG, W. (eds.) *Biomechanics of the musculo-skeletal system.* 3rd ed. New Jersey, NJ: John Wiley & Sons.

NIGG, B. M., FISHER, V. & RONSKY, J. L. 1994. Gait characteristics as a function of age and gender. *Gait and Posture,* 2(4), 213-220.

NIGG, B. M., STEFANYSHYN, D., COLE, G., STERGIOU, P. & MILLER, J. 2003. The effect of material characteristics of shoe soles on muscle activation and energy aspects during running. *Journal of Biomechanics,* 36(4), 569-575.

NOKIA. 2007. *Devices* [Online]. Available: http://www.nokia.com/press/media_resources/photos/devices [Accessed December 2007].

NUSSBAUM, M. A., MARTIN, B. J. & CHAFFIN, D. B. 1997. A neural network model for simulation of torso muscle coordination. *Journal of Biomechanics,* 30(3), 251-258.

NYAN, M. N., TAY, F. E. H., SEAH, K. H. W. & SITOH, Y. Y. 2006. Classification of gait patterns in the time-frequency domain. *Journal of Biomechanics,* 39(14), 2647-2656.

O'HALLORAN, P. D., MURPHY, G. C. & WEBSTER, K. E. 2004. Mood during a 60-minute treadmill run: Timing and type of mood change. *International Journal of Sport Psychology,* 35(4), 309-327.

OH, I. S. & SUEN, C. Y. 1998. Distance features for neural network-based recognition of handwritten characters. *International Journal on Document Analysis and Recognition,* 1(2), 73-88.

OHANIAN, P. P. & DUBES, R. C. 1992. Performance evaluation for four classes of textural features. *Pattern Recognition,* 25(8), 819-833.

OLDS, E. G. 1938. Distributions of sums of squares of rank differences for small numbers of individuals. *The Annals of Mathematical Statistics,* 9(2), 133-148.

ORENDURFF, M. S., SEGAL, A. D., BERGE, J. S., FLICK, K. C., SPANIER, D. & KLUTE, G. K. 2006. The kinematics and kinetics of turning: Limb asymmetries associated with walking a circular path. *Gait and Posture,* 23(1), 106-111.

ORPONEN, P. 1994. Computational complexity of neural networks: A survey. *Nordic J. of Computing,* 1(1), 94-110.

PEARSON, G. L. 1948. A magnetic field strength meter employing the hall effect in germanium. *Review of Scientific Instruments,* 19(4), 263-265.

POLAR ELECTRO OY. 2010. *Rs800 training computer* [Online]. Polar Electro Oy. Available: http://www.polar.fi/en/products/earlier_products/RS800 [Accessed January 15 2010].

POLIKAR, R. 2006. Ensemble based systems in decision making. *Circuits and Systems Magazine, IEEE,* 6(3), 21-45.

PONTIL, M. & VERRI, A. 1998. Support vector machines for 3d object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 20(6), 637-646.

POPOVIC, R. S. 1989. Hall-effect devices. *Sensors and Actuators,* 17(1), 39-53.

PROAKIS, J. G. & MANOLAKIS, D. G. 1992. *Digital signal processing (2nd ed.): Principles, algorithms, and applications,* Indianapolis, IN, Macmillan Publishing Co., Inc.

RAHMAN, A. S. A., YAP, M. M. S., SHAKAFF, A. Y. M., AHMAD, M. N., DAHARI, Z., ISMAIL, Z. & HITAM, M. S. 2004. A microcontroller-based taste sensing system for the verification of eurycoma longifolia. *Sensors and Actuators, B: Chemical,* 101(1), 191-198.

RÄTSCH, G., ONODA, T. & MÜLLER, K. R. 2001. Soft margins for adaboost. *Machine Learning,* 42(3), 287-320.

RIPKA, P. 1994. Magnetic sensors for industrial and field applications. *Sensors and Actuators: A. Physical,* 42(1), 394-397.

SADEGHI, H., ALLARD, P., SHAFIE, K., MATHIEU, P. A., SADEGHI, S., PRINCE, F. & RAMSAY, J. 2000. Reduction of gait data variability using curve registration. *Gait and Posture,* 12(3), 257-264.

SANDERSON, C. & PALIWAL, K. K. 2002. Fast feature extraction method for robust face verification. *Electronics Letters,* 38(25), 1648-1650.

SAPANKEVYCH, N. & SANKAR, R. 2009. Time series prediction using support vector machines: A survey. *IEEE Computational Intelligence Magazine,* 4(2), 24-38.

SAUPE, D., LÜCHTENBERG, D., RÖDER, M. & FEDEROLF, C. 2007. Analysis and visualization of space-time variant parameters in endurance sport training. Proceedings of the 6th International Symposium on Computer Science in Sports, June 3-6 2007, Calgary, AB. 252-256.

SCHAALE, M. & FURRER, R. 1995. Land surface classification by neural networks. *International Journal of Remote Sensing,* 16(16), 3003-3031.

SCHAPIRE, R. E., FREUND, Y., BARTLETT, P. & LEE, W. S. 1998. Boosting the margin: A new explanation for the effectiveness of voting methods. *Annals of Statistics,* 26(5), 1651-1686.

SCHAPIRE, R. E. & SINGER, Y. 1999. Improved boosting algorithms using confidence-rated predictions. *Machine Learning,* 37(3), 297-336.

SCHMIDT, R. O. 1986. Multiple emitter location and signal parameter estimation. *IEEE Transactions on Antennas and Propagation,* 34(3), 276-280.

SCHÖLKOPF, B., SMOLA, A. & MÜLLER, K. R. 1998. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation,* 10(5), 1299-1319.

SCHÖLKOPF, B. & SMOLA, A. J. 2002. *Learning with kernels: Support vector machines, regularization, optimization, and beyond,* Cambridge, MA, MIT Press.

SCHÖLLHORN, W. I. 2004. Applications of artificial neural nets in clinical biomechanics. *Clinical Biomechanics,* 19(9), 876-898.

SCHÖLLHORN, W. I., NIGG, B. M., STEFANYSHYN, D. J. & LIU, W. 2002. Identification of individual walking patterns using time discrete and time continuous data sets. *Gait and Posture,* 15(2), 180-186.

SEPULVEDA, F., WELLS, D. M. & VAUGHAN, C. L. 1993. A neural network representation of electromyography and joint dynamics in human gait. *Journal of Biomechanics,* 26(2), 101-109.

SHENOI, B. A. 2006. *Introduction to digital signal processing and filter design,* Hoboken, NJ, Wiley-Interscience.

SILVER, A. E., LUNGREN, M. P., JOHNSON, M. E., O'DRISCOLL, S. W., AN, K. N. & HUGHES, R. E. 2006. Using support vector machines to optimally classify rotator cuff strength data and quantify post-operative strength in rotator cuff tear patients. *Journal of Biomechanics,* 39(5), 973-979.

SONG, Y., SHIN, S., KIM, S., LEE, D. & LEE, K. H. 2007. Speed estimation from a tri-axial accelerometer using neural networks. 9th Annual International Conference of IEEE-EMBS, Engineering in Medicine and Biology Society (EMBC 2007), August 22-26 2007, Lyon. 3224-3227.

SONY-ERICSSON 2006. White paper w850. Lund: Sony Ericsson Mobile Communications AB.

SPEARMAN, C. 1904. The proof and measurement of association between two things. *American Journal of Psychology,* 15(1), 72-101.

SPECHT, D. F. 1990. Probabilistic neural networks. *Neural Networks,* 3(1), 109-118.

STAMKOPOULOS, T., DIAMANTARAS, K., MAGLAVERAS, N. & STRINTZIS, M. 1998. Ecg analysis using nonlinear PCA neural networks for ischemia detection. *IEEE Transactions on Signal Processing,* 46(11), 3058-3067.

STEFANYSHYN, D. J., STERGIOU, P., LUN, V. M., MEEUWISSE, W. H. & WOROBETS, J. T. 2006. Knee angular impulse as a predictor of patellofemoral pain in runners. *American Journal of Sports Medicine,* 34(11), 1844-1851.

STEWART, J. & WANG, C. J. 2003. Waveform recognition of back-electro motive force for real-time embedded system applications using support vector machines. *In:* DAGLI, C. H., BUCZAK, A. L., GHOSH, J., EMBRECHTS, M. & ERSOY, O., eds. Proceedings of the Artificial Neural Networks in Engineering Conference (ANNIE 2003), November 2-5 2003, St. Louis, MO. 951-956.

STIRLING, L. M., KUGLER, P. & VON TSCHARNER, V. 2009. Support vector machine classification of muscle intensity patterns during prolonged running. 22nd Congress of the International Society of Biomechanics, July 5-9 2009, Cape Town. No pagination.

SUN. 2002a. *Javatm apis for bluetoothtm wireless technology (jsr-82)* [Online]. Sun Microsystems. Available: http://jcp.org/aboutJava/communityprocess/final/jsr082/index.html [Accessed February 1 2010].

SUN. 2002b. *Mobile information device profile for javatm 2 micro edition* [Online]. Sun Microsystems. Available: http://jcp.org/aboutJava/communityprocess/final/jsr118 [Accessed February 1 2010].

SUN. 2004. *Fileconnection api (jsr-75)* [Online]. Sun Microsystems. Available: http://jcp.org/aboutJava/communityprocess/final/jsr075 [Accessed February 1 2010].

SUN. 2006a. *Location api (jsr-179)* [Online]. Sun Microsystems. Available: http://jcp.org/aboutJava/communityprocess/final/jsr179 [Accessed February 1 2010].

SUN. 2006b. *Mobile media api (jsr-135)* [Online]. Sun Microsystems. Available: http://jcp.org/aboutJava/communityprocess/mrel/jsr135 [Accessed February 1 2010].

SUN. 2007. *Connected limited device configuration (cldc); jsr 30, jsr 139 overview* [Online]. Sun Microsystems. Available: http://java.sun.com/products/cldc/overview.html [Accessed February 1 2010].

TAUNTON, J. E., RYAN, M. B., CLEMENT, D. B., MCKENZIE, D. C., LLOYD-SMITH, D. R. & ZUMBO, B. D. 2002. A retrospective case-control analysis of 2002 running injuries. *British Journal of Sports Medicine,* 36(2), 95-101.

TAUSCHEK, G. 1929. *Reading machine*. United States patent application 2026329. Dec 1935.

THEODORIDIS, S. & KOUTROUMBAS, K. 2009. *Pattern recognition,* Amsterdam, Elsevier Academic Press.

TROJE, N. F. 2002. Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. *Journal of Vision,* 2(5), 371-387.

TÜXEN, S. 2009. *Analysis and classification of gyroscopic sensor data from runners.* Information and Communication Technology Bachelor Thesis, University of Erlangen-Nuremberg.

TZANETAKIS, G. & COOK, P. 2002. Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing,* 10(5), 293-302.

UMAPATHY, K. & KRISHNAN, S. 2006. Modified local discriminant bases algorithm and its application in analysis of human knee joint vibration signals. *IEEE Transactions on Biomedical Engineering,* 53(3), 517-523.

UNSER, M. 1995. Texture classification and segmentation using wavelet frames. *IEEE Transactions on Image Processing,* 4(11), 1549-1560.

VAN MECHELEN, W. 1992. Running injuries. A review of the epidemiological literature. *Sports Medicine,* 14(5), 320-335.

VAPNIK, V. N. 1998. *Statistical learning theory,* New York, NY, Wiley & Sons.

VARELA, C., TIMP, S. & KARSSEMEIJER, N. 2006. Use of border information in the classification of mammographic masses. *Physics in Medicine and Biology,* 51(2), 425-441.

VIOLA, P. & JONES, M. J. 2004. Robust real-time face detection. *International Journal of Computer Vision,* 57(2), 137-154.

VON TSCHARNER, V. 2000. Intensity analysis in time-frequency space of surface myoelectric signals by wavelets of specified resolution. *Journal of Electromyography and Kinesiology,* 10(6), 433-445.

VON TSCHARNER, V. 2009. Spherical classification of wavelet transformed EMG intensity patterns. *Journal of Electromyography and Kinesiology,* 19(5), 334-344.

VON TSCHARNER, V. & GOEPFERT, B. 2003. Gender dependent EMGs of runners resolved by time/frequency and principal pattern analysis. *Journal of Electromyography and Kinesiology,* 13(3), 253-272.

VON TSCHARNER, V. & HERZOG, W. 2007. EMG. *In:* NIGG, B. M. & HERZOG, W. (eds.) *Biomechanics of the musculo-skeletal system.* 3rd ed. New Jersey, NJ: John Wiley & Sons.

WANG, P. Y., AN-LIAN, Q. & KANG, H. G. 1992. A multiparameter telemetering system used in shell rowing study. Proceedings of the IEEE Conference of the Engineering in Medicine and Biology Society, October 29 - November 1 1992, Paris. 1287-1288.

WELCH, R. M., SENGUPTA, S. K., GOROCH, A. K., RABINDRA, P., RANGARAJ, N. & NAVAR, M. S. 1992. Polar cloud and surface classification using avhrr imagery: An intercomparison of methods. *Journal of Applied Meteorology,* 31(5), 405-420.

WILLIAMS, R. J. & ZIPSER, D. 1989. A learning algorithm for continually running fully recurrent neural networks. *Neural Computation,* 1(2), 270-280.

WILLIS, A. J. & MYERS, L. 2001. A cost-effective fingerprint recognition system for use with low-quality prints and damaged fingertips. *Pattern Recognition,* 34(2), 255-270.

WITTEN, I. & FRANK, E. 2005. *Data mining: Practical machine learning tools and techniques,* San Francisco, CA, Morgan Kaufmann.

WOLD, E., BLUM, T., KEISLAR, D. & WHEATON, J. 1996. Content-based classification, search, and retrieval of audio. *IEEE Multimedia,* 3(3), 27-36.

WOLF, W., OZER, B. & LV, T. 2002. Smart cameras as embedded systems. *Computer,* 35(9), 48-53.

WU, J. & WANG, J. 2008. PCA-based SVM for automatic recognition of gait patterns. *Journal of Applied Biomechanics,* 24(1), 83-87.

WU, J., WANG, J. & LIU, L. 2006. Kernel-based method for automated walking patterns recognition using kinematics data. *In: Advances in natural computation.* 1st ed. Berlin: Springer.

WU, J., WANG, J. & LIU, L. 2007. Feature extraction via kpca for classification of gait patterns. *Human Movement Science,* 26(3), 393-411.

YANG, J., ZHANG, D., YONG, X. & YANG, J. Y. 2005. Two-dimensional discriminant transform for face recognition. *Pattern Recognition,* 38(7), 1125-1129.

YE, J. & LI, Q. 2005. A two-stage linear discriminant analysis via qr-decomposition. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 27(6), 929-941.

ZHANG, K., SUN, M., LESTER, D. K., PI-SUNYER, F. X., BOOZER, C. N. & LONGMAN, R. W. 2005. Assessment of human locomotion by using an insole measurement system and artificial neural networks. *Journal of Biomechanics,* 38(11), 2276-2287.

ZHANG, Y. & ROCKETT, P. I. 2009. A generic multi-dimensional feature extraction method using multiobjective genetic programming. *Evolutionary Computation,* 17(1), 89-115.

ZHOU, H., LIU, Y., LI, L. & WEI, B. 2009. A clustering approach to free form surface reconstruction from multi-view range images. *Image and Vision Computing,* 27(6), 725-747.

APPENDIX

Two classification approaches (Support Vector Machines and AdaBoost) were not thoroughly explained within the thesis due to word limitations of the respective included manuscripts. Those approaches were, however, important for the results that were obtained within the thesis. Therefore, further information about these algorithms is given here.

Support Vector Machine

A Support Vector Machine (SVM) uses a set of m training feature vectors $\mathbf{x}_i \in \Re^n, i = 1,...,m,$ with each vector belonging to one of two classes $y_i = \{-1,1\}$, to find a decision hyperplane of dimension n-1 that linearly separates both classes (Vapnik, 1998; Burges, 1998). The distance of the training vectors that are closest to this hyperplane (these are called the Support Vectors), is tried to be maximized. For this reason the SVM is called a maximum margin classifier.

The decision function of the SVM has the form $f(\mathbf{x}) = \mathrm{sgn}(\mathbf{w}^t\mathbf{x} + b)$, see also Fig. 4.2. In this decision function, **w** is the normal vector of the decision hyperplane and b is the distance from the origin of the hyperplane. From all possible decision hyperplanes, the SVM identifies the one with the minimal quadratic norm of **w**, thus, the problem of finding the optimal hyperplane can be rewritten as

$$\min_{\mathbf{w},b}\left(\frac{1}{2}\mathbf{w}^t\mathbf{w}\right), \tag{A.1}$$

subject to $y_i\left(\mathbf{w}^t\mathbf{x}_i + b\right) \geq 1,\ i = 1...m,$ which is equal to maximizing the distance of the Support Vectors to the plane. The normal vector **w** can then be expressed as a linear combination

$$\mathbf{w} = \sum_1^m \alpha_i y_i \mathbf{x}_i \ , \tag{A.2}$$

of the training vectors. With this, the decision function can be rewritten as

$$f(\mathbf{x}) = \mathrm{sgn}\left(\sum_1^m \alpha_i y_i \mathbf{x}_i^t \mathbf{x} + b\right). \tag{A.3}$$

A feature vector is then assigned to one of the two classes depending on the sign of the decision function.

Usually, a linear separation in the given feature space is not possible, thus, the feature vectors are mapped to a space of higher dimension. This is done by using a kernel function K that is a dot product of the feature vectors as

$$K(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i)^t \Phi(\mathbf{x}_j), \tag{A.4}$$

where $\Phi(\mathbf{x})$ denotes a function which maps a feature vector to a (usually) higher dimensional space. The underlying idea is that in a higher dimensional space, it is more likely possible to find a hyperplane that separates the classes linearly. Different kernel functions are available for this purpose (Schölkopf and Smola, 2002), for example linear, polynomial or radial basis function kernels.

For non-linearly separable cases even in higher dimensions, slack variables are introduced into the optimization task. The optimization task is a convex programming problem, and can be solved using standard optimization packages. A well written tutorial publication about all the different aspects of Support Vector Machine implementation can be found in (Burges, 1998). A reference implementation (‘libSVM’) that covers many aspects of the classifier as well as allows to test and compare different kernel functions is freely available in the web (Chang and Lin, 2001).

AdaBoost

For AdaBoost and, specifically, its often used variant AdaBoost.M1, additional information to Chapter 3 can be given using pseudocode. This was shown in a very well organized overview publication (Polikar, 2006), which covers also other aspects of boosting. The pseudocode given in this publication looks as follows:

- Input: Sequence of N examples $S = [(\mathbf{x}_i, y_i)]$, $i = 1,\ldots,N$ with labels $y_i \in \Omega$, $\Omega = \{\omega_1,\ldots,\omega_C\}$, where C is the number of classes.
- Input: Weak learning algorithm **WeakLearn**.
- Input: Integer T specifying number of iterations.
- Initialize: $D_1(i) = 1/N$, which is the initial weight distribution.
- Do: for $t = 1,\ldots,T$:
    - 1. Select a training data subset S_t, drawn from the distribution D_t.
    - 2. Train **WeakLearn** with S_t, receive hypothesis h_t.
    - 3. Calculate the error of h_t: $\varepsilon_t = \sum_{i:h_t(\mathbf{x}_i)\neq y_i} D_t(i)$. If $\varepsilon_t < 1/2$, abort.
    - 4. Set $\beta_t = \varepsilon_t/(1-\varepsilon_t)$.
    - 5. Update distribution $D_t : D_{t+1}(i) = \frac{D_t(i)}{Z_t} \cdot \begin{cases} \beta_t & \text{if } h_t(\mathbf{x}_i) = y_i \\ 1, & \text{otherwise} \end{cases}$, where $Z_t = \sum_i D_t(i)$ is a normalization constant chosen so that D_{t+1} becomes a proper distribution function.
- Test: Weighted Majority Voting. Given an unlabeled instance $\mathbf{x}$,
    - 1. Obtain total vote received by each class $V_j = \sum_{t:h_t(\mathbf{x})=\omega_j} \log\frac{1}{\beta_t},\ j = 1,\ldots,C.$
    - 2. Choose the class that receives the highest total vote as the final classification.

GLOSSARY

Accelerometer

A measurement instrument used to determine acceleration. Usually, they possess a small mass that is connected to a stiff spring. Typically, the spring deflection is measured when the mass is accelerated.

AdaBoost

A classifier. AdaBoost is a meta classifier that combines multiple simple classifiers to a strong one. The resulting decision boundary is nonlinear.

Anthropometry

The study concerned with measuring the proportions, size, and weight of the human body.

Artificial Neural Network

See Neural Network.

Biomechanics

A scientific discipline that examines forces that act upon and within biological structures. Furthermore, the effects that are produced by such forces are investigated.

Class

One of a discrete number of categories that a classifier assigns an object to. The object is represented by a feature vector.

Classification

The process of assigning a feature vector to a class.

Classification rate

A measure of the ability of a classification system to correctly assign an unknown object to its class.

Classifier

An algorithm from pattern recognition that facilitates the classification of objects into a number of classes.

Coronal plane

An imaginary plane that travels vertically from the top to the bottom of the body, dividing it into forward (anterior) and backward (posterior) sections. It is perpendicular to the transverse and sagittal planes.

Cross-validation

A method for evaluating a statistical model that has free parameters. The training data is divided into several parts, and in turn one part of the data (the test set) is used to test the procedure that is fitted to the remaining parts (the training sets). Cross-validation also provides an estimation of the generalization performance of a classification system.

Curse of dimensionality

The problem caused by the exponential increase in volume associated with adding extra dimensions to a (mathematical) space.

DCT

See Discrete Cosine Transform.

Decision boundary

A hypersurface that partitions the feature space into several subspaces. In a classification problem with two classes, two subspaces are for example constructed. The classifier will classify all the points on one side of the decision boundary as belonging to one class and all points on the other side as belonging to the other class.

Decision criterion

See decision boundary.

Decision stumps

A simple classifier. It performs a threshold decision in one or multiple feature dimensions. It is not commonly used on its own, since very few problems can be accurately classified using a simple threshold. Typically, multiple decision stumps are used in the AdaBoost learning algorithm as simple classifiers. The decision boundary that results from a single decision stump is linear.

Discrete Cosine Transform

A transform that is used to represent the frequency content of a time series.

Electromyography

A technique for evaluating and recording the muscle activation. EMG measurement is performed using an instrument called an electromyograph, to produce a record called an electromyogram.

Embedded system

A computer system designed to perform only a few dedicated functions, frequently with real-time computing constraints. It often consists of a microprocessor, sensing unit, additional hardware and mechanical parts. The microprocessor is thus embedded as part of a complete device. By contrast, a personal computer, is designed to be flexible and to meet a wide range of end-user needs. Embedded systems control many devices in common use today.

EMG

See electromyography.

Feature

The individual measurable properties of the objects being observed. Choosing independent and discriminating features is key to any pattern recognition algorithm in order to be successful in classification. Features are usually numeric, however structural features (graphs, strings, etc.) are sometime also employed. The classification success is critically dependent on the choice of features. Therefore it is helpful to consider multiple representations of the same data (i.e. different features).

Feature extraction

The process of computing a feature representation from an input signal.

Feature ranking

The process of assigning each individual feature a measure of importance for the classification task. Often, feature selection is done according to the feature ranking result.

Feature reduction

Reducing the number of features for classification by combining features, for example using a principal component analysis. In contrast to feature selection, the information from all computed features in contained in the reduced representation.

Feature selection

Reducing the number of feature for classification by removing features that are irrelevant for classification, or by selecting features that are most relevant for classification. The criterion is usually the obtainable classification rate. In contrast to feature reduction, only the information from the most relevant computed features in contained in the selected representation.

Feature space

The space where each measured object is represented as a point in a high-dimensional space. The dimension of the space is determined by the number of features used to describe the objects. Similar objects from the same class are ideally grouped together.

Feature vector

A representation of data suitable for classification. Each individual feature is concatenated into the feature vector representation. The feature vector for classification is typically of constant length and defines the dimensionality of the feature space.

Force

It is not possible to define force. However, the effects of force can be defined. Force is normally represented by a vector, with magnitude and direction.

Force plate

An measurement instrument used to determine the magnitude and direction of the ground reaction force beneath the foot during gait.

Frontal plane

See coronal plane.

Fundamental frequency

The predominant frequency in a complex waveform.

Gait

Describes style and manner of locomotion, rather than the walking process itself.

Generalization performance

The ability of a classification system to correctly classify new samples, i.e. feature representations from objects that have not been used in the training set of the classification system.

Generic feature

Features that are not specifically adapted to the sensor input, as opposed to manual features.

Inverse dynamics

A method for computing forces and moments based on the kinematics of a body and the body's inertial properties (mass and moment of inertia). In practice, inverse dynamics computes the forces and moments from measurements of the motion of limbs and external forces such as ground reaction forces.

Kinematics

Describes motion without reference to the forces that are involved. An example of a kinetic instrument is a camera, which can be used to observe the motion of the trunk and the limbs during walking, but which gives no information on the forces.

Kinetics

The study of mass, force, moments, and acceleration without the complete knowledge of the orientation and position of the objects that are involved. A force plate is normally used for measuring force during gait.

Label

An unambiguous, discrete identifier for a feature vector that assigns it to a specific class. It is unique for each class.

LDA

See Linear Discriminant Analysis.

Linear Discriminant Analysis

A classifier. It computes statistics from features before a classification with a linear decision boundary is conducted.

Locomotion

The process of movement from place to place through an act of self-propulsion.

Mahalanobis distance

A distance measure in statistics introduced by P. Mahalanobis. It determines the similarity of an unknown sample set to a known set. In contrast to Euclidean distance, it takes into account the correlations of the data set.

Manual features

Features that are specifically designed to represent characteristics of the input measurements, as opposed to generic features.

Median plane

See sagittal plane.

Microprocessor

An electronic component that is made from miniaturized transistors and other circuit elements. These components are placed on a single semiconductor integrated circuit. It is a complete processing unit and capable of performing arithmetic and logical operations.

Motion analysis

The process of measuring locomotion, typically gait. Motion analysis is often conducted by attaching reflective markers to anatomical landmark positions. The 3D position of these markers is then measured as a time series.

Naïve Bayes

A classifier. It assumes that the components of the feature vector are conditionally independent for each class. The resulting decision boundary is a quadratic function.

Nearest Neighbor classifier

A classifier. It assigns a new feature vector to the class that the k nearest neighbors of this feature vector possess in feature space. The resulting decision boundary is a nonlinear and unconnected.

Neural Network

A classifier. It employs a network of neurons that usually implement a logistic scaling function. The neurons are connected in the network by weight functions. These are adapted during training, and a nonlinear decision boundary results.

Object

In pattern recognition, an object is a real-world entity that is desired to be automatically assigned to a class. Examples are: audio signals in speech recognition, handwritten characters in character recognition and kinetic/kinematic measurements of subjects from defined groups for classification in biomechanics.

Object detection

A subfield of pattern recognition. In object detection, images of real-world scenes are typically analyzed. The task of object detection is, for example, to recognize humans or cars in the images.

Optical character recognition

A subfield of pattern recognition. In optical character recognition, images of printed or handwritten characters are typically analyzed. The task of optical character recognition is to automatically transcribe the printed or handwritten characters to an electronic representation.

Patellofemoral pain syndrome

An inflammation of the patellar tendon in the knee. It results in pain and/or discomfort.

Pattern classification

See pattern recognition.

Pattern recognition

A scientific discipline with the goal to classify objects into a number of categories or classes. Pattern recognition and pattern classification are used synonymously throughout this thesis.

PFPS

See patellofemoral pain syndrome.

Recognition rate

See classification rate.

Regression

The process of predicting the value of a random variable *y* from a measurement *x*. Regression is generalizing classification since *y* can be any quantity, including a class label. Many classification algorithms can be seen as thresholding the output of a regression. Curve fitting is the common special case of regression.

Sagittal plane

An imaginary plane that travels vertically from the top to the bottom of the body, dividing it into left and right sections. It is perpendicular to the coronal and transverse planes.

Sensor

A device that measures a physical quantity and converts it into an electronic representation. Sensors are often used in pattern recognition for measuring an object. Sensors can be part of embedded systems.

Speech recognition

A subfield of pattern recognition. In speech recognition, audio signals containing words or sentences are typically analyzed. The task of speech recognition is to automatically transcribe the utterance to an electronic textual representation.

Support Vector Machine

A classifier. It subjects features to a nonlinear transformation before a classification with a linear decision boundary is conducted. The linear decision boundary is computed by applying a maximum margin criterion.

SVM

See Support Vector Machine.

Test set

A set comprising of feature vectors from one or different objects that is used for testing the classifier. The test set is unknown to the classifier (but labeled) and therefore allows evaluating the generalization performance and the classification rate of the classifier. This process is often automated using cross-validation.

Time series

A measurement of a variable that may change as a function of time.

Training set

A set comprising of feature vectors from different objects that is used for training the classifier, i.e. for defining the decision criterion. For supervised classifier training, the training set is labeled. After training, the test set is used to evaluate the generalization performance and the classification rate of the classifier. This process is often automated using cross-validation.

Transverse plane

An imaginary plane that travels horizontally, dividing the body into upper (superior) and lower (inferior) sections. It is perpendicular to the coronal and sagittal planes.

Ubiquitous computing

A concept of computing that is present in many aspects of daily life, often not noticed by the users. Ubiquitous computing may involve multiple different embedded devices in various appliances that often operate in the background.

Wavelet

A one-dimensional wave-like oscillation that is typically delimited in time. Attributes of a Wavelet are amplitude, frequency and phase. Wavelet transformations are used to represent time series simultaneously in frequency and time.