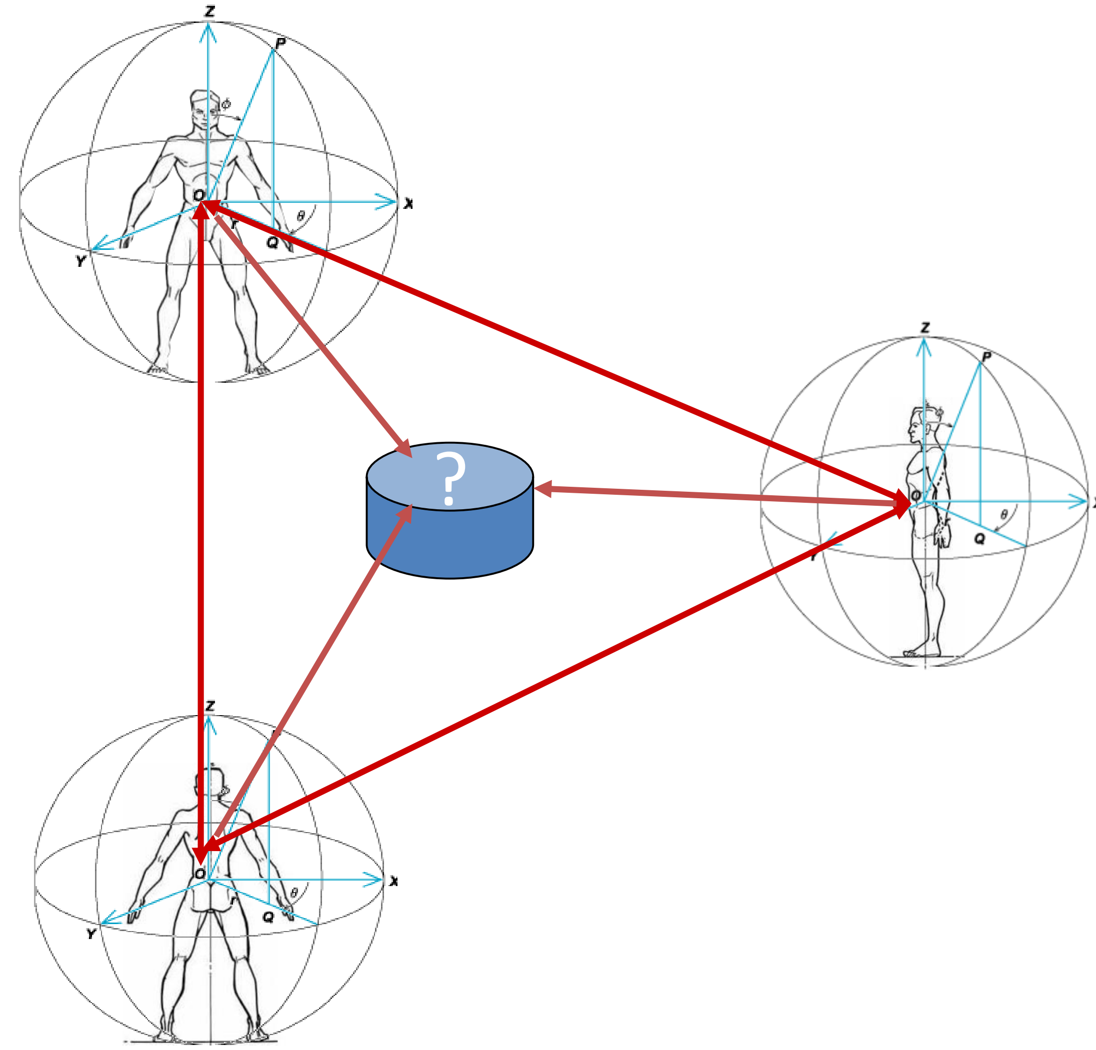


Kamrad Khoshhal Roudposhti

PROBABILISTIC-BASED HUMAN BEHAVIOUR ANALYSIS USING A HIERARCHICAL FRAMEWORK

Tese de Doutoramento na área científica de Engenharia Electrotécnica e de Computadores, especialidade de Automação e Robótica, orientada pelo Senhor Professor Doutor Jorge Manuel Miranda Dias e apresentada ao Departamento de Engenharia Electrotécnica e de Computadores da Faculdade de Ciências e Tecnologia da Universidade de Coimbra.

Maio de 2014

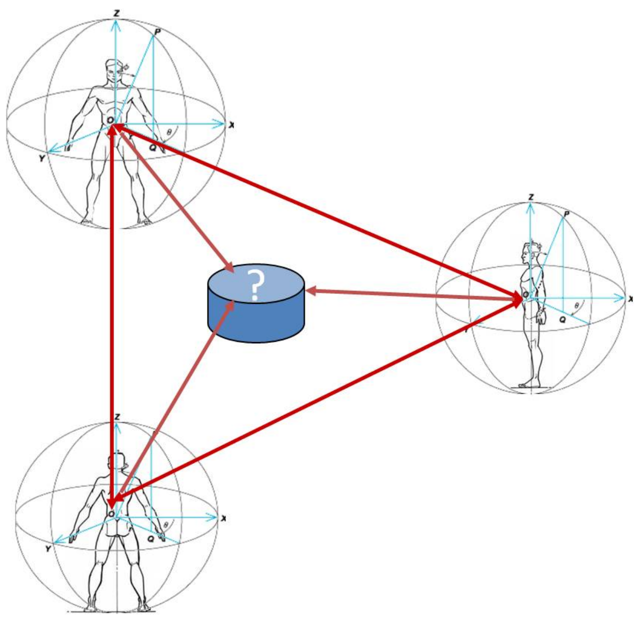
· U C ·

UNIVERSIDADE DE COIMBRA

Kamrad Khoshhal Roudposhti
PROBABILISTIC-BASED HUMAN BEHAVIOUR ANALYSIS USING A HIERARCHICAL FRAMEWORK
UNIVERSIDADE DE COIMBRA

University of Coimbra
Faculty of Science and Technology
Department of Electrical and Computer Engineering

Probabilistic-based Human Behaviour Analysis using Hierarchical Framework



PhD Thesis

Kamrad Khoshhal Roudposhti

Coimbra, May 2014

University of Coimbra
Faculty of Science and Technology
Department of Electrical and Computer Engineering

Probabilistic-based Human Behaviour Analysis using Hierarchical Framework

Thesis submited:

to the Electrical and Computer Engineering Department of the Faculty of Science and Technology of the University of Coimbra in partial fulfillment of the requirements for the Degree of Doctor of Philosophy.

Kamrad Khoshhal Roudposhti
Coimbra, May 2014

This thesis was supervised by

Professor Doctor Jorge Manuel Miranda Dias

Professor at the
Faculty of Science and Technology, University of Coimbra

To my beloved parents and brothers

Resumo

Os recentes progressos das tecnologias e infra-estruturas de monitorização e compreensão do comportamento humano sugerem novas e interessantes aplicações. A presente tese de doutoramento tem por objectivo a analise do comportamento humano em diferentes contextos, utilizando os movimentos do corpo humano. As primitivas associadas aos movimentos do corpo humano são menos restritivas que outras possíveis (ex. expressões faciais ou voz), no entanto a sua analise é mais complexa. O desafio prende-se com a grande dimensionalidade dos movimentos dos elementos do corpo, complexidade dinâmica, interdependência, e elevados requisitos computacionais para analise e estimação do comportamento humano.

Inicialmente, propõe-se uma metodologia hierárquica para reduzir a complexidade do processo em diferentes camadas. Em seguida, propõe-se a utilização do reconhecido descritor de movimentos humanos, o Laban Movement Analysis (LMA), para extrair e decompor as primitivas em cinco componentes. As componentes LMA garantem um numero mínimo de primitivas para analise de qualquer movimento humano, colmatando a lacuna entre as Low Levels Features (LLFs) e o movimento. Finalmente, a metodologia é modelizada recorrendo a redes bayesianas (Bayesian Network (BN), Dynamic BN (DBN) e Hidden Markov Model (HMM)) permitindo lidar com a incerteza dos dados, utilizar processos de aprendizagem em pequena escala, fundir diferentes tipos de primitivas (no domínio da frequência e espaço) e obter um modelo de dependências suficientemente flexível entre as varias primitivas e camadas. A metodologia proposta nesta tese explora diferentes actividades e comportamentos humanos, nomeadamente: movimento de elementos do corpo humano, analise humana individual, interacção homem/objecto, interacção humano/humano, comportamento interpessoal e o papel social das pessoas.

A metodologia proposta foi modelizada e construída seguindo uma abordagem de baixo para cima (bottom-up). A camada de analise de baixo nível foi decomposta em dois domínios: frequência e espaço. Esta camada permite estimar componentes LMA, obter um conhecimento de baixo nível dos movimentos e fornece dados para hierarquias superiores de analise de actividades humanas. Por exemplo, a componente de esforço (Effort), responsável pela dinâmica dos movimentos do corpo humano foi modelizada no domínio da frequência, a componente forma (Shape), responsável pela deformação 3D dos elementos do corpo durante os movimentos foi modelizada através de primitivas espaciais.

A etapa de analise humano/objecto e interacção humano/humano (baseada no contexto) requer o conhecimento individual de cada acção humana e respectivas inter-relações. A modelização das inter-relações, foi inspirada na "Relation Ship component" , uma das componentes LMA pouco explorada, e permite obter uma metodologia e um modelo de interacção humano/objecto e humano/humano capaz de explorar as actividades humanas num dado contexto (scene understanding). As actividades humanas num contexto social foram igualmente analisadas utilizando informações do nível LMA. O sistema é suficientemente flexível para comportar as complexas inter-relações entre as várias primitivas utilizadas na estimação dos movimentos resultantes dos comportamentos interpessoais e respectivo papel social (inspirado no trabalho de investigação "Honest Signals" de Alex Pentaland).

Uma abordagem e uma metodologia hierárquica apresentam várias vantagens, tais como modelos flexíveis, generalização a aplicações relacionadas, expansibilidade da tecnologia sensorial, capacidade de lidar com a incerteza e produção de informação semântica em todas as camadas de analise.

A abordagem e o sistema proposto garantem uma monitorização automática do comportamento humano e tem uma nível de aplicabilidade muito alto em muitos cenários com pessoas, ex. estudos clinico, sistemas de segurança, cuidados a idosos, vigilância, treino desportivo, realidade virtual, coreografia, etc.

O sistema proposto, foi experimentalmente testado com dados reais obtidos através de um fato com sensores de movimento capazes de gerar informação da posição 3D das várias partes do corpo humano a uma frequência de 120Hz. O dataset produzido é o resultado da utilização desse fato por varias pessoas enquanto executavam actividades bem definidas e foi calibrado individualmente. Isto significa que os dados gerados para cada pessoa nos diferentes trials podem ser diferentes consoante a calibração.

Os resultados obtidos em cada etapa, demonstram as capacidades da abordagem e metodologias propostas nos vários níveis de analise da actividade humana. Em síntese, a tese propõe uma abordagem global e descritiva, suficientemente flexível, capaz de explorar e estimar os vários níveis dos movimentos do corpo baseados na actividade.

Abstract

The recent progress in technologies and infrastructures of monitoring and understanding human behaviour suggest new and interesting applications. This PhD research aims at analysing human behaviour in different contexts, through the motions of the human body. The features of human movements are less restricted than the other possible features (e.g. facial expressions and voice) in many real world applications, however they are more complicated to be analysed. Due to the large dimensions of body parts movements and the complex dynamics and dependencies between them, high computational processing resources are needed to analyse and estimate human behaviours.

Firstly, a hierarchical framework is proposed to reduce the complexity of the process in different layers. Secondly, a well-known human movement descriptor, Laban Movement Analysis (LMA), which provides different types of needed features in five components, is proposed. The LMA components prepare minimum needed features that can assist us to analyse any kind of human activities, and fill the gap between the Low Level Features (LLFs) and human movements analysis. Finally, for modeling the framework, Bayesian-based approaches (Bayesian Network (BN), Dynamic BN (DBN), and Hidden Markov Model (HMM)), are defined to deal with the uncertain data, to apply learning processes using small data, to fuse different types of features (in frequency and spatial domains) and to have enough flexibility for modeling the different dependencies between different features and layers. In this study we explored several different human activities and behaviours through the framework, namely; body parts movements, human individual analysis, human-object interaction, human-human interaction, interpersonal behaviour, and social role of people.

The mentioned framework is modeled and constructed in a bottom-up strategy. During the process, different approaches are proposed to solve the problems. From the lowest level of analysis, the study is divided into two different domains; frequency and spatial. This study is performed to estimate LMA components, which provide enough knowledge in body motion level, to be able to explore more high level of human activities analysis. For instance; Effort component, which explains human body movement dynamics, is modeled in frequency domain, and Shape component, which explains human body shape deformation in 3D space during any movement, is modeled by spatial based features. Based on the mentioned components, individual human action level is modeled.

In the next step we need to analyse human-object and human-human interaction (context-based), each individual human action and the relations between them. Those relations are modeled by inspiration of the Relationship Component which is one of the less explored LMA component. We were thus able to model human-object and human-human interactions by modeling Relationship components through the framework to explore human activities concerning the context (i.e., scene understanding).

With respect to the proposed framework, human activities in social context are explored by information of the LMA level. The system was flexible enough to provide the complex existing dependencies between different features to estimate body-motion based interpersonal behaviours and a social role, inspired by Alex Pentland's investigation in "Honest signals" book.

The hierarchical framework presents many capabilities such as; flexibility of modeling, generalizing to different related applications, extendability by progressing the sensory technology, dealing with uncertainly in all levels of analysis, and providing semantic-based information for all layers of analysis. The proposed framework provides an automatic human behaviour monitoring system which is a very highly required application in fields like clinical study, security systems, elder-care, surveillance systems, sport training, virtual reality, choreography, etc.

In the experimental process, a motion tracker suit which provides 3D position of human body parts in maximum 120 Hz resolution, is used. To prepare a dataset, several people dressed the suit and performed the defined activities. The attached sensors of the suit for each record trial, need to be calibrated. It means the data for each person in different trials can be different with respect to the calibration process.

The obtained results in each step, present the capability of the mentioned framework in different levels of human movement activity analysis. Furthermore, a descriptive global framework to explore and estimate various levels of human body-motion based activities, is proposed.

Acknowledgment

I would like to thank Professor Doctor Jorge Dias for his supervision, support and encouragement. I should also thank ISR-UC for the facilities provided and Foundation of Science and Technology of Portugal for the granted scholarship SFRH/BD/70640/2010, and PROMETHEUS European project's group for their knowledge sharing, during my study.

My thanks to the professors and colleagues from ISR, like Anìbal T. de Almeida, Paulo Menezes, Rui Rocha, Jorge Lobo, João Filipe Ferreira, Omar Tahri, Luis Santos, João Quintas, Paulo Freitas, Pedro Trindade, Pedro Machado, Diego Faria, Jose Prado, Amilcar Ferreira, David Portugal, Abed Malti, Tito, Ricardo Martin, Filipe Ferreira, Catia Pinho, Paula Lopes, Rita Catarino, Jörg Rett, José Marinho, Hugo Faria, Alexandre Malhão, Micael Couceiro, André Araújo, Bruno Patrão, Tiago Docílio, Paulo Drews Jr, Joao Monteiro, Carlos Simplício, Oswaldo Ludwig, Cristiano Premebida, Luciano Oliveira, Nono Martins, , and Christiana Tsiourti.

My Spanish colleagues from University of Extremadura namely Leandro Serrano, Luis Manso, Pedro Nunez and Professor Antonio Bandeira from University of Malaga. My French colleagues from Probayes® research company namely Kamel Mekhnacha and Julien Ros. My colleague from Technical University of Munich, Martin Hofmann.

I am very grateful to my all friends specially Iranian friends for their supports; Afshin Ghanbari, Doctor Hadi Aliakbarpour, Shima Naseh, Sadra Aliakbarpour, Doctor Mahmoud Tavakoli, Leila Sadeghi, Diana Tavakoli, Doctor Ali Marjovi, Sarvenaz Choobdar, Ehsan Asadi, Ehsan Ghafari, Saba Nikrahi, Behrang Chenari, Shiva Saadatian, Seyed Jafar Hosseini, Sirvan Khalighi, Amir Safaei, Elsa Gomes, Amir Azhdari, Samaneh Abbasi, Maryam Abbasi, Naghmeh Ivaki, Ashkan Ivaki, Mojtaba Bandarabadi, Mohsen Bandarabadi, Homa Maleki, Maryam Bandari, Sousan Siassipour and her family, Doctor Nima Garmanjani, Parisa Ahani, Ebrahim azhdari, Aram Emami, Jahed Naghipour and his family, Mohammad Dodangeh and his family, Behzad Bayat and his family, Marzieh Bakhshandeh and her family, Maryam Khoshhal and her family, Doctor Masoumeh Ghalkhani, Saeid Rastegar, Morad Daneshvar, Ivano Irrera.

I would like to thank my wife, Maryam, whose love and devotion have been incredible. Special thanks to my faithful and thoughtful dad, my late mother, and my brothers, who were my first teachers and their love and encouragement inspired my passion for learning.

Contents

List of Figures

List of Tables

List of Algorithms

Abbreviations

2D	Two-Dimensions/Two-Dimensional
3D	Three-Dimensions/Three-Dimensional
ATM	Automatic Teller Machine
BN	Bayesian Network
BP	Bayesian Programming
DBN	Dynamic Bayesian Network
DFT	Discrete Fourier Transform
DHMM	Discrete Hidden Markov Model
FFT	Fast Fourier Transform
HB	Human Behaviour
HBA	Human Behaviour Analysis
HMI	Human Machine Interaction
HMM	Hidden Markov Model
HRI	Human Robot Interaction
IB	Interpersonal Behaviour
IMU	Inertial Measurement Unit
LLF	Low Level Feature
LMA	Laban Movement Analysis
NIR	Near-Infrared
PCA	Principal Component Analysis
PCR	Positive Classification Rate
PS	Power Spectrum
SR	Social Role
SSP	Social Signal Processing
SVM	Support Vector Machine
XML	Extensible Markup Language

Chapter 1

Introduction

Automatic monitoring of human activities in different applications such as; surveillance, elder-care , social behaviour analysis, etc., is probably one of the most important needs in near future. There are several works which have been done for each aspect of human motion-based activity analysis, for a specific application, but not for more general applications which can be estimated from human movements. For instance, in rehabilitation application we need low-level information of body movements, in a surveillance system we need a sequence of human movements, while in human action and interaction analysis we need a sequence of human movements related to the environmental parameters. The problem is that almost no system is capable of having all that different information in one level, moreover, there are several dependencies between those levels of information which can connect them in a bottom-up framework. Each level of information can be used as an input for higher ones, however, each level of the framework has its own informative features for a specific application. For instance; body motion level for rehabilitation and choreography applications, human movement level for human-robot interaction, gesture recognition, and sign language analysis applications, human action and interaction level for human-robot interaction, scene understanding, surveillance, and elder-care applications, social context-based behaviour level for social robot and security applications, are useful.

In this study, we propose a framework and an approach that estimates the various levels of information through human movements features. There are several problems in having model with all of those expected advantages. For instance; feature selection, feature categorization, analysing the existing dependencies between features to extract higher information, dealing with uncertainty, exploring through the connection between different complexity levels of human movement activities, exploring through the relationship between context-based knowledge and human movement, and complexity of modeling. Thus, to solve those problems, we propose a hierarchical framework, to illustrate the connection between various levels of analysis and simplify the complexity of modeling and learning process. The hierarchical framework is provided by using Bayesian Network (BN) based approaches, to deal with uncertainty and study on the existent features dependencies statistically. In order to have a proper feature selection and categorization system for the mentioned purpose, a standard human movement descriptor, Laban Movement Analysis (LMA), is used.

Fig.1.1 presents the diagram of different levels of human behaviour analysis process. The first level defines the used sensory data related to human body motion. The second level, LLFs are extracted in a couple of domain categories; frequency and spatial domains. In Chapter 2, the second level is discussed in detail. The third level is the LMA level (description level) information, which is estimated from LLFs (Chapter 2). In Chapter 3, the fourth level, which analyses human movement, and the human action and interaction's section of the fifth level, are discussed. Chapter 5 consists of the interpersonal behaviour section of the fifth level and the sixth level, which is about social behaviour analysis. The seventh level, which concerns about interest scenarios, is involved in Chapter 3, Chapter 5 and Chapter 4. In this study, Chapter 4 is a preprocessing research, which improves other chapters results. Its improvement on the results which are obtained on the Chapter 3, are presented in Chapter 4.

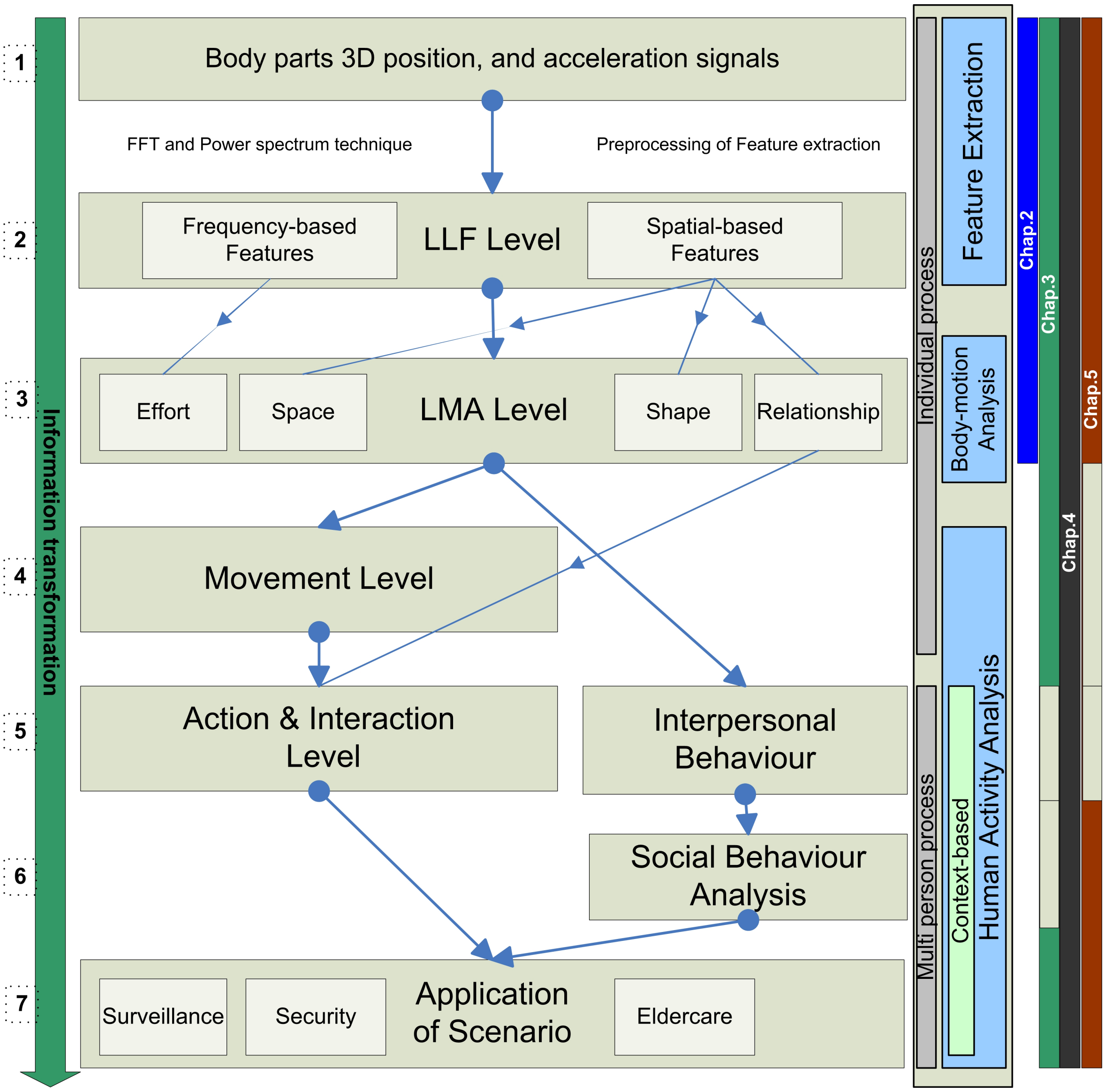


Figure 1.1: Human behaviour analysis in different levels with respect to the following chapters

1.1 Motivation

Different applications are highly interested in human behaviour analysis. Table 1.1 presents some of the existing related applications. In related research of analyzing human behaviour, using facial expression, voice and body motions are popular. However in reality, in most of the mentioned applications, it is not possible to use facial expression and voice-based features to understand human behaviours. Thus body motion-based features are going to be more utilized in the related area. Generally body motion-based analysis can be categorized in 2D and 3D domains. Most of the works were done in 2D based features which are faster, but not reliable in several cases. On the other hand, 3D-based approaches are more reliable but more time consuming. Recently, by progressing computer vision through 3D-based analysis, we can have proper 3D-base features real-time (such as the Kinect). We are thus able to investigate human complex behaviours by using advantageous 3D-based approaches.

1.2 Applications of Human behavior analysis

There are many reasons why it is useful to understand human behaviour. Virtual reality is a popular application for human behaviour analysis. In order to present the human behavior more naturally through virtual space, human behaviour analysis techniques are very important [Dev06].

Surveillance is another popular application of human behaviour analysis. Many places (e.g., parking lots, supermarkets, airports, secure zones) in the world, monitor the areas with security guards using many cameras and screens. Alternatively, we can use human behaviour analysis with smart machines to recognize human actions automatically and detect potential hazards.

Recently in sports, teams, players, coaches, etc, are using different technologies more and more to improve their quality. Human behaviour analysis can help them to analyze their games. Also there are several applications in the medical field such as rehabilitation, which are very valuable. Table. 1.1 presents some of the existing related applications.

General domain	Specific area
Virtual reality	-interactive virtual worlds -virtual studios -games -teleconferencing -character animation
“Smart” surveillance systems	-access control -building lobby, bank lobbies -parking lots -train station ticket offices -public building entry areas -library space -student lab -airport -supermarkets, department stores -vending machines, ATMs -coffee room, lounge/play room -traffic -suspicious behaviour identification
Motion analysis	-Content-based indexing of sports video footage -clinical studies of orthopedic patients -choreography of dance -personalized training in golf, tennis, etc.

Table 1.1: Human behaviour analysis applications

Authors/Ref/year	Description
Aggarwal/[ACLS94]/1994	Articulated and elastic non-rigid motion.
Cedars et al./[CS95]/1995	Body parts config. and motion estimation.
Gavrila/[Gav99]/1999	Human motion analysis in 2D and 3D .
Pentland/[Pen00]/2000	Human motion analysis and its applications.
Wang et al./[WHT03]/2003	Human detection, tracking and activity.
Moeslund et al./[MbK06]/2006	Initialization, tracking, pose estimation.
Poppe/[Pop10]/2010	Human motion and human action analysis.
Aggarwal et al./[AR11]/2011	Human high level activity analysis.
Hotle et al./[HMNP11]/2011	Human 3D pose and activity estimation.

Table 1.2: A number of surveys around body motions-based investigations

1.3 Related works

Body parts motions convey different levels of information during human daily activities. Scientists explore that information depending on their application. For example; Foroud and Whishaw [FW06] analysed body parts motions through kinematic structure and non-kinematic features of movements during skilled reaching after stroke. In a more high-level process, Ryoo and Aggarwal [RA09] have a sequence of works which analyse different levels of human activities in 2D from human movement to group activity understanding. In the last couple of decades, different body motion-based features were extracted and used to understand various kinds of human activities. Based on the used techniques and approaches, their applications and solving the existing constraints, several surveys were published in different taxonomies (see Table.1.2).

1.3.1 Hierarchy-based approach

The hierarchy-based approach is an approach which solves a complex problem through a number of simple ones in a sequence of levels (multi-level). This approach is very useful, and well-known to model complex problems, however, defining the multi-level of the process and the proper dependencies between the levels, are the issues which need to be addressed precisely. There are several works which attempted to solve the mentioned problems directly from Low Level features (LLFs) to human movements or even higher level of analysis (Table. 1.2). The scientists had several problems which a hierarchical approach can solve. In hierarchy-based approaches the biggest problem divides into some smaller ones. This property has several advantages in the related applications, and the main one is the flexibility of the model. For estimating a higher level of human behaviour information, we don't need any changes in previous levels (no remodeling, relearning and reclassification process for previous steps) we only need the output of the previous model and add needed knowledge to define a new level of analysis on the top of the model.

The structure of each level and how to define the dependencies between those levels in a hierarchical framework is very important. The recent related survey from Aggarwal and Ryoo [AR11], mentioned a number of works which use the advantages of hierarchical concept, through their focus was more on the last level's information , and usually their mid-level analysis was not informative. However, there is plenty of research around the subject, but there is no general framework that is able to present all that different information in one general model.

Imagine a person is walking while keeping one hand in his pocket and shakes hands with another person. The information which can be important to different applications can be listed as:

1: Body part motions level; analysis of normality of hands and feet movements by observing the probabilities of each body parts motions states (rehabilitation).

2: Movement level; if hands and feet are moving periodically as a normal walking, or the walking is like a drunk person, or the walking is in one hand in pocket state, etc. (gait and gesture recognition).

3. Action or interaction level; the person is walking to reach to the other person and do handshaking (human action and interaction analysis).

Based on the various level of interpretations, the person before the handshaking is in a "reaching" state (action and interaction level), in a "walking" state (movement level), and the left hand and feet are in normal motion but the right hand is in "static" state (body part motions level).

To have one framework with all that information, we propose LMA parameters in body part motion level, and a hierarchy model which can use lower level information for higher ones. Through this transformation the process of learning and classification will reduce several times.

1.3.2 Scene context-based Human behaviour analysis

Behaviour ascribes to an object actions with respect to its environment. Human behaviour takes from different behaviour sources imposed by attitudes, emotions, culture, etc. As a matter of fact, human behaviour can range from being considered normal or admissible to abnormal or inadmissible in a specific situation.

Human behaviour analysis has a very wide meaning, making it a rather complex problem. Take for example, the recognition of specific movements due to variations of one action or gesture performed by one individual, and even the considerable differences in the dynamic of the same gesture performed by different people. Examples of different approaches to this subject would be: gait recognition, action analysis [RYS02], gesture recognition [MA07], facial expression recognition [MGIJ03], explicit body movement based communication, namely sign language recognition [KBOZ04], etc, illustrating the numerous facets of human movement analysis. Hence, there are several notational systems to obtain the expressive content of human movements theoretically which can help us find useful features to analyze human movements and behaviours.

Arsic et al. in [AWSR05] discussed a real-time behaviour detection method which is video based. Their application is for passenger behaviour detection in public transports such as airplanes. They defined some special human behaviour such as aggressive, nervous, tired, kid and talk. Thus they defined some low level features (LLF) such as global motion (e.g. head movements, sit down,rise, etc.), face position, eye movement (e.g. blinking), lip movement, gaze which gained by using near infrared (NIR) illumination and red eye effect. For the detection of each feature, they used an existing real-time method. They also noted their LLF detection outputs may have high error rates, so they used a trained Bayesian network.

Hongeng et al. in [HNB04] illustrated an approach for complex multi-agent human activity (e.g. stealing). They supposed each human activity includes a few action threads done by an actor. Thus, they defined a Bayesian network which uses Low level features (trajectory and shape of moving blob) to recognize the action threads. To achieve the goal, they defined a number of temporal constraints and event threads likelihoods to understand a few multi-person activities (e.g. stealing and moving an object).

Jahnson and Hogg in [JH02] showed a probabilistic method for stochastic behaviour modeling (e.g. pedestrian trajectory modeling scenario). They used a set of estimated parameters which are given from a training set of joint density distribution. They defined a Gaussian mixture by state history to model non-linear behaviours and used an auto-regressive process which comprises model conditioning, state change selection, and state updated to solve the problem.

Cohen and Li works in [CL03] described a method to identify a shape posture by 2D and 3D shape of human body classification. They used four cameras and for each one, a silhouette of a human body, and then by the set of silhouettes, a 3D shape of a human body can be achieved. In order to distinguish between the different body postures, they described a 2D/3D shape description system by changeability of human body proportions, and defined a learning step based on Support Vector Machine (SVM) to solve the problem of identifying similar postures, which are different, by different people.

Leo et al. in [LDG+04, LDS04] proposed an approach for complex human activities detection from image sequences in outdoor environment such as archaeological sites (e.g. walking, probing the subsoil by a stick, damping the ground with a tank and picking-up objects from the ground). They estimated human body postures frame by frame, and then modeled the temporal sequences of the detected postures statistically by Discrete Hidden Markov Models. They used a Basic Competitive Learning Scheme (BCLS) which is an unsupervised clustering algorithm fed by features which are horizontal and vertical histograms of the binary shapes. Finally they used the Manhattan distance for both clusters building and run-time classification.

Khalid and Naftel in [KN05] presented an approach for clustering and classifying spatio-temporal object trajectories by a neural network learning algorithm. They used a Discrete Fourier Transform (DFT) method to achieve Fourier coefficients for modeling motion trajectories. Then they used the Fourier coefficient feature to recognize similar motion patterns. They defined a Self-Organizing Map (SOM) by a Neural Network which has one layer of input nodes connected directly to a single 1-dimensional output layer. Each input node which was fed by the coefficient of the basis functions is connected to every output node with the weighted connection. The SOM is an unsupervised system which is able to learn similarities between object trajectories. Finally they used a Mahalanobis classifier for recognizing abnormal trajectories [Mcl04].

In [dLS02], Leon and Sucar presented a method using Fourier transform and Bayesian network for human activity recognition concerning the variety velocity of similar activities by different people and missing some parts of trajectories because of occlusion.

They supposed that one does not need to know the beginning and the end of activities.

Mecocci et al. in [MPF03] proposed a system which is able to automatically adapt to different scenarios without any human intervention, thus they used self learning techniques to automatically learn the “typical” behaviour of the targets in each specific operative environment. With these learned data, the system can detect abnormal behaviors of object.They used an Altruistic Vector Quantization algorithm (AVQ) to achieve self-learning capabilities which they believed can describe the trajectories of objects in complex, not structured and outdoor scenarios automatically.

Nascimento et al. in [NFM05] described a method for recognizing some human activities in a shopping space (e.g. entering, exiting, passing and browsing). They used human motion patterns which are achieved from a sequence of displacements of each human’s blob center. They modeled trajectories of the human’s blob by using a method which they called a multiple dynamical model with a switching mechanism. Finally they estimated the identification of the models which connected with a trajectory.

Medioni et al. in [MCB$^+$01] discussed analysis of moving objects behaviours through movies which are obtained from an airborne moving platform on a scene. They defined two steps for that; the first step used an approach for detecting and tracking motion regions in the movie, which infer their trajectories. The second step used the trajectories and information which was provided by a user in the form of geospatial and goal context to instantiate likely scenarios.

Remagnino and Jones in [RJ01] presented a high-level description of events which happen in a typical surveillance scenario, and they grouped each event into type and behaviour. They determined “type” from event characteristics (e.g. height, width and velocity) by a Bayesian system. In addition, they used a Hidden Markov Model system to estimate some typical event behaviours such as vehicle-entering, person-entering, vehicle-exiting and person-exiting in a parking lot.

Yue et al. in [YZC03] proposed an image-based method which used human body part segmentation approach to cover the visual hull method weak [Lau94] for concave regions reconstruction. The first result was a virtual silhouette image fit to the given viewing direction which was used in human body part localization method. They produced the body parts separately in a virtual view from the corresponding input views and then assembled them together. For removing the separate and squeezed regions in the final view, they used the last silhouette image.

Yacoob and Black in [Yac99] discussed a method which used a number of atomic activities such as a set of measurements for modeling a few human activities such as walking, running, etc. They parametrized body parts (arm, foot, torso, thigh and calf) motions in human activities, and then they modeled similar human activities by achieving the body parts motions parameters.

Naylor et al. have some reports from ADVISOR project which was related to intelligent surveillance from video outputs [Nay06], detecting abnormal human behaviours to improve security systems by vision-based methods. They work in several subjects which are image capture, behaviour recognition, motion detection, crowd monitoring, people tracking, archiving, search and retrieval, human computer interface, communications over an IP infrastructure. In this project they used the mentioned techniques in a subway environment, however they believed the system can be fit also to other places such as airport and railway station.

There has been several attempts which used movement descriptors and probabilistic frameworks. For instance; Bregler [Bre97] introduced mid-level descriptors through a probabilistic framework for human movements classification. Low-level motion clusters were defined to analyse high-level gait categories. Those low-level motion clusters are limited to show the relations through the data of physiological studies of human movements, properly. To overcome this weakness, Rett et al. have tied their descriptors to a well established notational framework: Laban Movement Analysis (LMA) [RD07].

Norman Badler attepted to re-formulate the LMA descriptions using computational models [BPW93]. In addition, the work of Zhao & Badler [ZB05] used the computational models to learn motion qualities from human gestures.

Rett & Dias in [RDA08] presented a system that analyzes human movements realtime, based on the concept of Laban Movement Analysis. The implementation used a Bayesian model for learning and classification. They presented the Laban Movement Analysis as a concept to identify useful features of human movements to classify human actions. As mentioned before, the movements were extracted using both vision and a magnetic tracker. The descriptor opened possibilities to study expressiveness and emotional content. They used a Bayesian framework as it offers an intuitive approach to learning and classification. It also provides the possibility to anticipate the performed action given the observed features.

Addressing group-based behaviour analysis, Allbeck et al. in [AKA+02] presented

	Gathering	Dispersing	Obj. Referential	Formation	Milling
Shape					
Advancing					
Retreating					
Spreading		X			
Enclosing	x				
Effort					
Slow					
Fast					
Sudden					
Sustained					x
Direct	x	X	x	x	
Indirect					x
Free					
Bound					
Other					
Obj Referent			x		
Structured				x	

Table 1.3: Feature table for aggregate movements [AKA+02]

a system for analysing people aggregate movements in a virtual environment using a high-level interface. The major components contain: textquotedblleftan interactive interface for aggregate control based on a collection of parameters extending an existing movement quality model, a feature analysis of aggregate motion verbs, recognizers to detect occurrences of features in a collection of simulated entities, and a clustering algorithm that determines subgroups". They used LMA components (Effort and Shape) to describe aggregate movements as shown in the Tab. 1.3.

1.3.3 Enhanced human activity segmentation approach

A recent survey by Weinland et al. [WRB11], has identified three major action segmentation categories: Sliding Window, Boundary detection and Grammar Concatenation. The already reviewed Sliding windows are used to divide a motion sequence into multiple overlapping segments, which are bounded by the window limits. The information within the window, may or may not be processed for alternative representations. Each candidate segment (or equivalent representation) is then used for sequential classification. The success of this approach strongly depends on the discriminant abilities of the generated representations. As mentioned this technique is easily integrated with the majority of static and dynamic classifiers. The major drawbacks of this technique are computational burden, and the need of multiple window sizes to overcome the variability problem.

Boundary detection methods generally identify discontinuities or local extrema in observed motion signals. The boundaries usually define an implicit basic action taxonomy, without depending on specific class definitions. A branch of works identify boundary at the cost of the dynamics of the observed signal, such as [VKD08, LLF00]. Others depend on geometric property changes observed through techniques like Principal Component Analysis [BSP+04] or piecewise arc fitting models [AM09, BN10]. A related research addresses the segmentation problem from the subspace separation perspective, exploring the so called Agglomerative Lossy Compression [RTVM10]. In [RA00], the authors apply a Singular Value Decomposition (SVD) decomposition to a long sequence of optical flow images in order to detect trajectories discontinuities within SVD component trajectories. Ogale et al. [OKGfA04] also explore optical flow of body silhouettes, performing segmentation by detecting minimum and maximum values of the absolute value sequence.

A method using features from visual hulls is developed in [WRB06]. This category of approaches is very sensitive to noise and other related errors (e.g. camera perspectives). Additionally, it allows generic segmentation, but is not particularly suitable for labelling purposes. The focus is on boundary identification rather than interpretation of intermediate data.

Weinland et al. [WRB11] identify Grammars as another category. The common approach is to model state transitions between actions, where HMMs are a popular approach.

Multiple methods can be used to generate features. Some examples are curvature scale space and centroid distance function [Bea06], joint angles alone [LN06, PVW07], or together with velocity profiles [KI95], dynamic system representations [PR00, THB00, Kah03] and geometrical property encoding [OSK02]. These are applied to segment and label action sequences, at the expense of computing a minimum-cost path through the model using techniques like Viterbi path, Conditional Random Fields or Markov Models. However, these methods rely on the comprehensiveness of state grammars, which may jeopardize the model effectiveness and the generalization purpose, if large amount of training data is not available.

1.4 Contributions

The primary contributions of this research are as follows:

- Human movement was analysed in LMA framework by conjugating frequency and spatial-based domains features, and Effort component of LMA was formulated and modeled by exploring through frequency-domain features.
- A global framework to analyse human action and interaction in a scene was proposed inspired by the Relationship component of the LMA. Relationship component was studied to provide a framework which performs the existing dependencies between human movements and others (himself, stationary object in the scene and other person), probabilistically.
- An adaptive segmentation approach was proposed to change the sliding window approach by having feedback from previous outputs of the model. This approach tries to deal with temporal variation of human activities.
- Exploring through human body motions and modeling the dependencies between people movements to understand some standard defined interpersonal behaviours.
- Exploring through frequency-domain features in LMA component levels to analyse the performed interpersonal behaviours and a social role "Leading role".
- We proposed a probabilistic hierarchical framework that connects all this thesis study in one global model. This global model shows that, this study is generalizable thanks to the LMA components and Bayesian network approach.

1.5 Publications

Most of the thesis is based on the following publications and achievements:

1.5.1 Peer-reviewed journal articles

- Kamrad Khoshhal and Jorge Dias - "Probabilistic Human Interaction Understanding - Exploring Relationship Between Human Body Motion and the Environmental Context". Pattern Recognition Letters. Special Issue on Scene Understandings and Behaviours Analysis. Elsevier, 2013.
- Luis Santos, Kamrad Khoshhal, Jorge Dias - "Trajectory-based Human Action Segmentation". Pattern Recognition. Elsevier, [submitted].
- Kamrad Khoshhal, Jorge Dias - "Body Motion-based Social Behaviour Understanding by a Probabilistic Framework". IEEE Transac. Affective computing, [submitted].

1.5.2 Peer-reviewed international conference papers

- K. Khoshhal, J. Dias - "Improved Semantic-based Human Interaction Understanding Using Context-based knowledge". 2013 IEEE International Conference on Systems, Man, and Cybernetics (IEEE SMC 2013), Manchester, United Kingdom, October, 2013.
- K. Khoshhal, P. Menezes, J. Dias - "Descriptive Human Action and Interaction Analysis in a Hierarchical Framework". In 2012 International Workshop on Human-Agent Interaction (iHAI 2012), in IROS2012, Vilamoura, Portugal, 2012.
- K. Khoshhal, J. Dias - "Interrelation Analysis for Interpersonal Behaviour Understanding in Social Context". In 10th IFAC Symposium on Robot Control (SYROCO 2012), Dubrovnik, Croatia, September, 2012.
- K. Khoshhal, L. Santos, H. Aliakbarpour, J. Dias - "Parameterizing Interpersonal Behaviour with Laban Movement Analysis - A Bayesian Approach". in 3rd International Workshop on Socially Intelligent Surveillance and Monitoring (SISM2012) in CVPR2012, in Providence, Rhode Island, USA, June, 2012.

- K. Khoshhal, Jorge Dias - "Probabilistic Human Activity Understanding Exploring Relationship Between Human Body Motion and the Environmental Context" - in Workshop on Recognition and Action for Scene Understanding (REACTS) as a satellite event of the 14th International Conference of Computer Analysis of Images and Patterns (CAIP) Spain, 1-2 September 2011.

- K. Khoshhal, H. Aliakbarpour, J. Quintas, M. Hofmann and J. Dias - "Probabilistic LMA-based Human Motion Analysis by Conjugating Frequency and Spatial based Features" - in Wiamis 2011, 12th international Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS), 13-15 April 2011.

- K. Khoshhal, H. Aliakbarpour, J. Quintas, K. Mekhnacha, J. Ros and J. Dias - "LMA-based Human Behaviour Analysis Using HMM" - In DoCEIS11 , 2nd Doctoral Conference on Computing, Electrical and Industrial Systems, 21-23 February 2011, Costa da Caparica, Lisbon, Portugal.

- K. Khoshhal, H. Aliakbarpour, J. Quintas, P. Drews, J. Dias - Probabilistic LMA-based Classification of Human Behaviour Understanding Using Power Spectrum Technique - Fusion 2010 - 13th International Conference on Information Fusion - EICC Edinburgh, UK, July 26-29 2010.

- J. Quintas, K. Khoshhal, H. Aliakbarpour, M. Hofmann, J. Dias - "Using Concurrent Hidden Markov Models to Analyze Human Behaviours in a Smart Home Environment" - in Wiamis 2011, 12th international Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS), 13-15 April 2011.

- H. Aliakbarpour, K. Khoshhal, J. Quintas, K. Mekhnacha, J. Ros, M. Andersson and J. Dias - "HMM-based Abnormal Behaviour Detection Using Heterogeneous Sensor Network" - In DoCEIS11 , 2nd Doctoral Conference on Computing, Electrical and Industrial Systems, 21-23 February 2011, Costa da Caparica, Lisbon, Portugal.

- H. Aliakbarpour, J. F. Ferreira, K. Khoshhal, J. Dias - A Novel Framework for Data Registration and Data Fusion in Presence of Multi-modal Sensors - in Proceedings of the DoCEIS 2010 - Doctoral Conference on Computing, Electrical and Industrial Systems - Lisbon, Portugal, Feb. 22-24, 2010. Springer - ISBN 978-3-642-11627-8

- H. Aliakbarpour, P. Núñez, J. Prado, K. Khoshhal and J. Dias - An Efficient Algorithm for Extrinsic Calibration between a 3D Laser Range Finder and a Stereo Camera for Surveillance - Proceedings of the ICAR 2009 - 14th International Conference on Advanced Robotics - Munich, Germany, June 22-26, 2009.

1.6 Dissertation outline

Chapter 2 provides the fundamental information about the used body motion-based feature extraction approaches in frequency and spatial domains, which is inspired by a human movement descriptor framework.

Chapter 3 explains the proposed models in a couple of sections; individual-based and scene context-based human behaviour analysis. Using a probabilistic hierarchical framework as a bottom-up strategy to analyse different level of human behaviours, is discussed. Bayseian-based approaches (Bayesian naive, Dynamic Bayesian network and Human Markov Model (HMM)) are used to model the different level of human behaviour analysis.

Chapter 4 proposes an approach to solve sliding window-based segmentation problem which can improve the human movement-based analysis results. This approach tries to adapt slide window in matter of window size and time shift, based on the previous outputs of the classifier.

In chapter 5 we explored through human body motions to analyse social-context based behaviours which we call interpersonal behaviours and social role using Pentland's definitions [Pen08].

The overall conclusion, discussions and future work are presented in chapter 6.

Chapter 2

Human motion-based feature extraction

2.1 Introduction

Humans can usually realize each other's actions by observing their body parts motion dynamics as a kind of body language. Thus, if a system could observe enough information, such as human body parts motion dynamics, and symbolize the different types of human movements, the different actions would be more clearly and easily understandable in the decision making process, instead of using the LLFs directly.

Understanding human action and behaviour by using a standard human motion description framework can be very useful for the related scientific applications, for instance; surveillance system, eldercare, HRI, Social behaviour analysis, etc. Nowadays if a scientist wants to recognize human actions and behaviours computationally, (s)he needs to have a lot of prerequisite knowledge such as data registration, feature extraction, etc. Thus, in this work we intend to use and formulate a human movement descriptor. Although there are several attempts to define a descriptor for human motions , but still there is a lack of a comprehensive framework for the community.

As mentioned before, one of the basic problems is to define some useful Low Level Features (LLF) and observable information that are dependent on the environment and type of the sensory data. The interest features depend on the type of data that we intend to obtain from humans. Most of the attempts rely on the human motions [PPNH07]. Feature extraction approaches can be categorized into 2D and 3D based analysis. Recently, researchers are interested in using 3D data as input to reconstruct human body in order to avoid limitations of 2D image based approaches. However, it is usually a high time-consuming process.

By progressing the computer vision techniques such as [AFKD10], and devices such as Kinect, we believe that the 3D position of human body can be obtained sufficiently. Thus the biggest limitation of 3D based analysis will be solved in various applications and 3D based human behaviour analysis will be more interesting in the different communities. In this study, we used wearable sensors to obtain human body parts 3D position to have the input data, for other steps, such as feature extraction, human movement analysis, human behaviour analysis, etc., which are our research concerns.

After the detection of human body parts, we need to detect features. Features are clear properties which assist us in modeling of relevant problems. In this study, features are categorized in frequency and spatial domains. To explore through the features of human body motion scientifically, a human movement analysis system which is named Laban movement Analysis (LMA), is used. LMA was built for describing human motions (specially choreography of dance). In the last decade some researchers attempted to formulate its parameters to be computational [Zha01, Ret08]. We intend to define a LMA-based descriptor for analyzing human motions from different aspects by some qualities like if a body part motion is happening suddenly or sustained, strongly or lightly, etc. Fig.2.1 presents the related levels of proposed human behaviour analysis framework in this chapter.

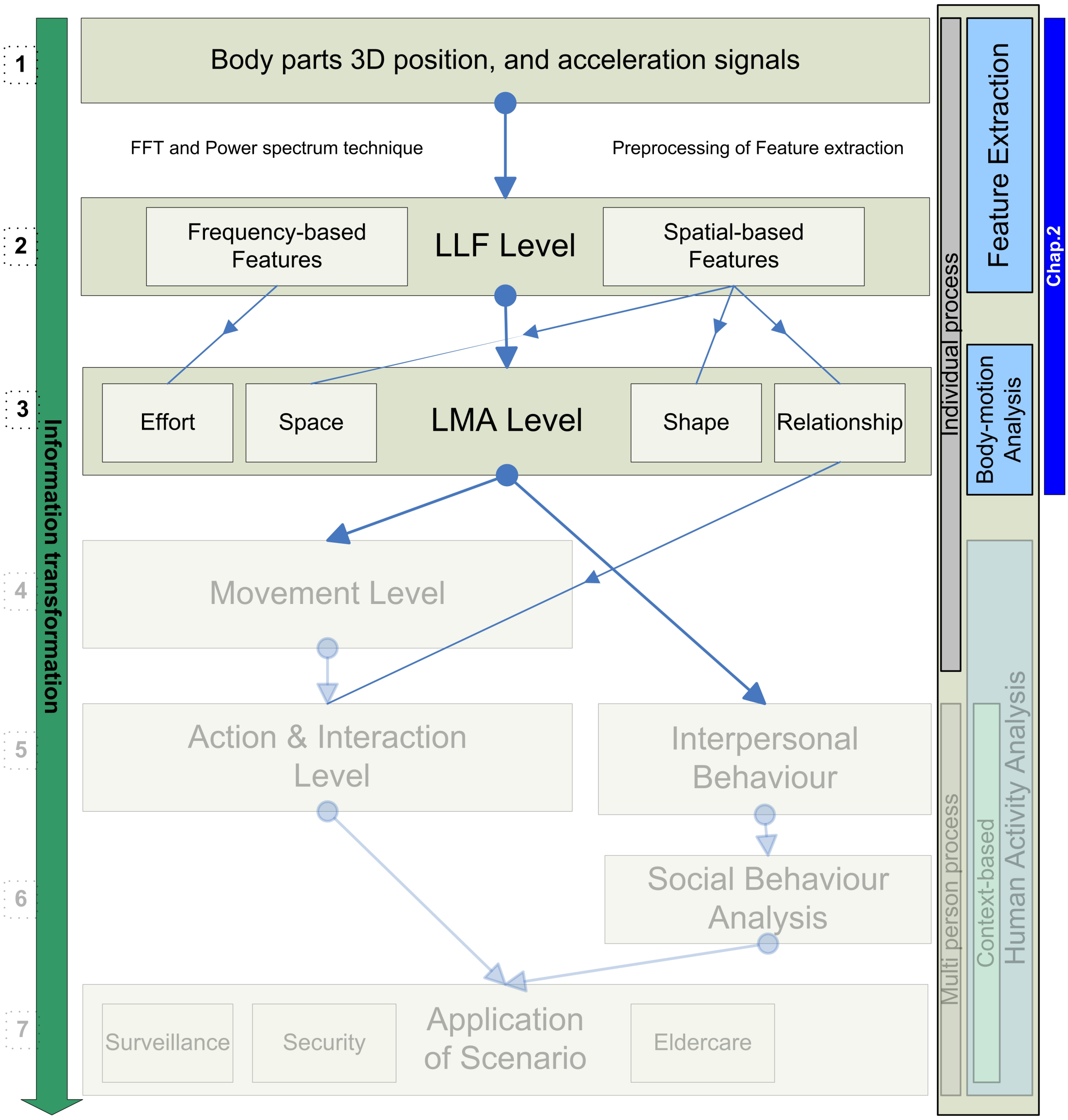


Figure 2.1: Human behaviour analysis in different levels with respect to this chapter

2.2 Related works

Movement understanding is the key to action recognition. Bobick in [Bob97] distinguishes between a motion (movement) and an action as follows: "a movement is the most atomic primitive, requiring no contextual or sequential knowledge to be recognized; an action is a larger scale event which typically includes interaction with the environment and causal relationships. Also, understanding an action implies producing a semantically rich description of the various action primitives and the relations between them".

There are several different categories of human movements that were attempted to understand. Each category of human movement can be identified by detecting the changes of the position in one or more involved parts of the human body. In this case, the concept of human movements is not important. Thus, researchers try to track one or more parts of the body in a sequence of captured data such as a sequence of images. Accurate human movement tracking is essential for the time behaviour analysis procedure. Legs, hands, head and face are the most important body parts to be tracked. For example, for gait recognition, leg motion is the relevant cue; for facial expression recognition, it's needed to focus just on the face; while in gesture recognition, several body parts need to be tracked, in particular hands and head.

There are several surveys about human motion analysis. The earliest related review was probably proposed by Aggarwal et al. [ACLS94], which covered different approaches, used in "articulated and elastic non-rigid motion". The approaches related to articulated motion, were analysed in using or not using a prior shape models. Cedars and Shah [CS95] provided a motion extraction method overview, in which human motion analysis was illustrated as " action recognition, recognition of body parts and body configuration estimation ".

Gavrila in [Gav99] presented different methodologies that categorized 2D and 3D approaches using or not using an explicit model of shape. It concluded the discussion with a couple of directions of related research through tracking in 3D space and action recognition. A relevant research by Pentland [Pen00] focused on; "person identification, surveillance/monitoring, 3D methods, and smart rooms/perceptual user interfaces", was provided to review the state of the art of "looking at people". The paper is not a survey about human motion analysis, but addressed various interesting issues in human motion analysis and its relevant applications. Moeslund and Granum presented a survey of computer-vision-based human motion capture [MG01]. Its focus was on a general overview based on the taxonomy of system functionalities, separated into four processes: "initialization, tracking, pose estimation and recognition". In addition, a number of common suppositions in the relevant research field, were recognized, and proposed different research directions for the problems. Wang et al. in [WHT03] provided a comprehensive survey of research on computer-vision-based human motion analysis. The article's emphasis was on three major issues related to a general human motion analysis system, namely "human detection, tracking and activity understanding".

Moeslund et al. published another survey that reviewed the trends in video-based human capture and analysis, as well as discussed open problems for future research to achieve automatic visual analysis of human movement [MbK06]. Poppe in [Pop07] described the characteristics of human motion analysis. He divided the analysis into a modelling and an estimation phase. He also discussed model-free approaches separately.

2.2.1 Specific subject related to human motion

Human motion is a very broad subject, and each researcher is studying a specific part, in order to solve sub problems in the area. We hereafter briefly describe some specific and popular works in this area.

2.2.1.1 Gait

The gait of a person is the manner of moving the body whilst walking [Hun03]. Thus, gait is not restricted to walking movement, it can also be usable to any means of movement on foot, such as running. The benefits of gait recognition technique over other biometric identification techniques forms, detailed by the following reasons in [Daw02]:

- Unobtrusive: "The gait of a person walking can be extracted without his/her knowledge and interfered without any cooperation in the information gathering stage, unlike fingerprinting or retina scans."
- Distance recognition: "The gait of an individual can be captured at a distance unlike other biometrics such as fingerprint recognition."

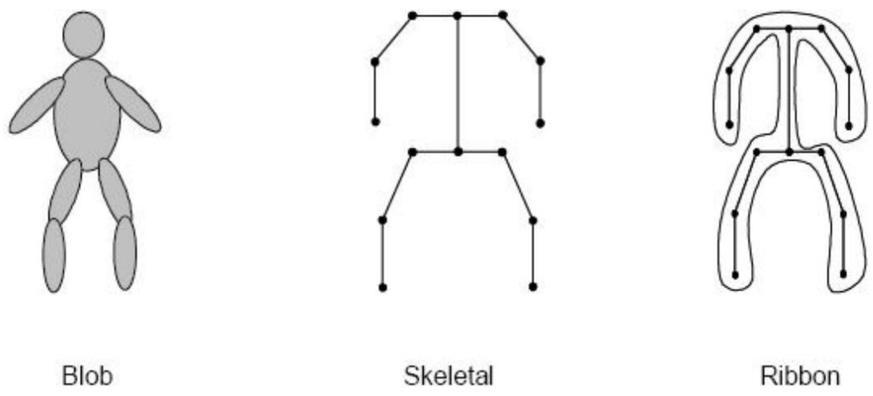


Figure 2.2: Different modeling techniques [Daw02].

- Reduced detail: "Gait recognition does not require images to be very high quality unlike other biometrics such as face recognition, which can be easily affected by low resolution images."

- Difficult to conceal: "The gait of an individual is difficult to disguise; in fact, by trying to do so the individual will probably appear more suspicious. With other biometric techniques such as face recognition, the individual's face can easily be altered or hidden."

Gait recognition techniques can be categorized into a couple of approaches: model-free and model-based. In the model-free approach, methods aren't linked to one object; as a holistic approach and, therefore, they can be used for animal gait too.

Model-based approaches use prior knowledge of the object for extracting features, which can be obtained from any captured data such as image. In Figure 2.2 presents the various representations, which were used by the relevant models. Model-based approaches can deal with occlusion and noise better than another. However, the disadvantage of implementing a model-based approach is its computational costs, due to the complex matching and searching that has to be performed.

2.2.1.2 Gesture

"People frequently use gestures to communicate; gestures are used about various issues, from pointing at a person to get their attention to conveying information about space and temporal characteristics" [Ken90], in which fingers, hands, arms, head, face, or body can be involved. S. Mitra and T. Acharya performed a survey on gesture recognition [MA07] and described a number of gesture recognition applications:

"

- Developing aids for the hearing impaired;
- Enabling very young children to interact with computers;
- Designing techniques for forensic identification;
- Recognizing sign language;
- Medically monitoring patients' emotional states or stress levels;
- Lie detection;
- Navigating and/or manipulating in virtual environments;
- Communicating in video conferencing;
- Distance learning/teleteaching assistance;
- Monitoring automobile drivers' alertness/drowsiness levels, etc.

"

2.2.1.3 Action analysis

Some research works are not only limited to a specific kind of human action such as gait or gesture recognition, they also address more general human activities, for example: opening or closing a door, picking up, putting down, running, walking, etc. [RYS02, YS05, YS07]. Rao et al. in [RYS02] proposed an approach to recognize human actions. They used some specific characteristics from trajectory of a part of actor body, which is involved in the actions. They illustrated that each of the actions trajectories have a number of useful characteristics which are proper features to recognize the actions (see Figure 2.3).

Leo et al. in [LDG$^+$04, LDS04] presented an approach for detecting some special activities which usually happen in archaeological sites (e.g. walking, probing the subsoil by a stick, damping the ground with a tank and picking-up some objects from the

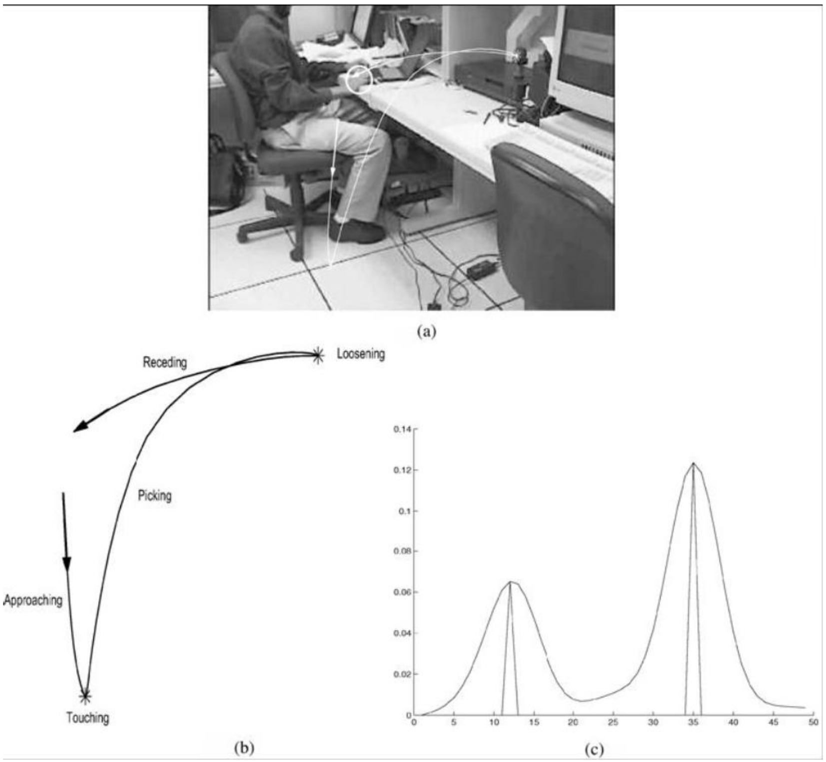


Figure 2.3: (a) The "picking up and putting down object" actions, the hand trajectory shown in white is super imposed on the last image; (b) action trajectory in terms of instants and intervals; and (c) corresponding " spatio-temporal curvature values and detected maximums (dynamic instants)" [46]

ground) as can be seen in Figure 2.4. A Discrete Hidden Markov Model (DHMM) was used for the purpose.

In [NFM05] a method is presented for recognizing some human activities in a shopping space (e.g. "entering", "exiting", "passing" and "browsing") as shown in Figure 2.5. They modelled trajectories of the human's blob by using "multiple dynamical models with a switching mechanism" method. Then, the activities were classified based on the trajectories of the human's blob.

Wu et al. in [WOQX05] presented two approaches to understand some abnormal activities such as; 'running people, bending down movement while most are walking or standing, a person carrying a long bar and a person waving hand in the crowd'. The first proposed method was using Principal Component Analysis (PCA) for feature selection and Support Vector Machine (SVM) for classification process. The second approach estimated optical flow to calculate each pixel's velocity to determine whether a human behaviour or activity is normal or abnormal.

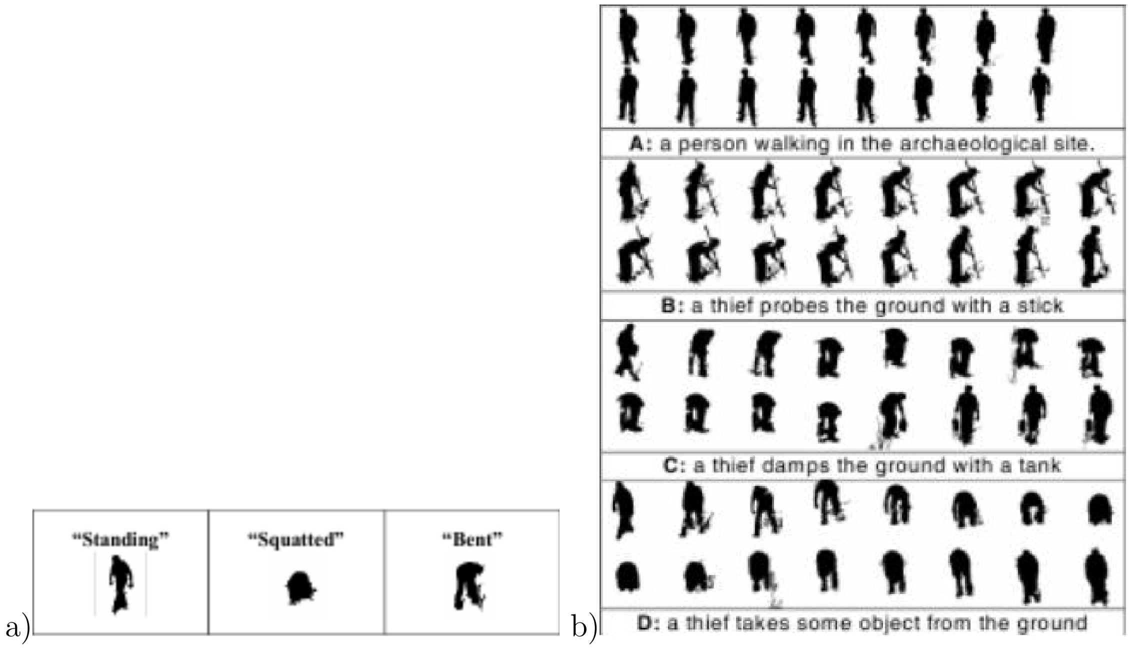


Figure 2.4: a) " Three fundamental postures classified in the archaeological site ", and b) some frames extracted from 4 of the 12 sequences (4 activities x 3 sequences) used to train the DHMM [LDG$^+$04].

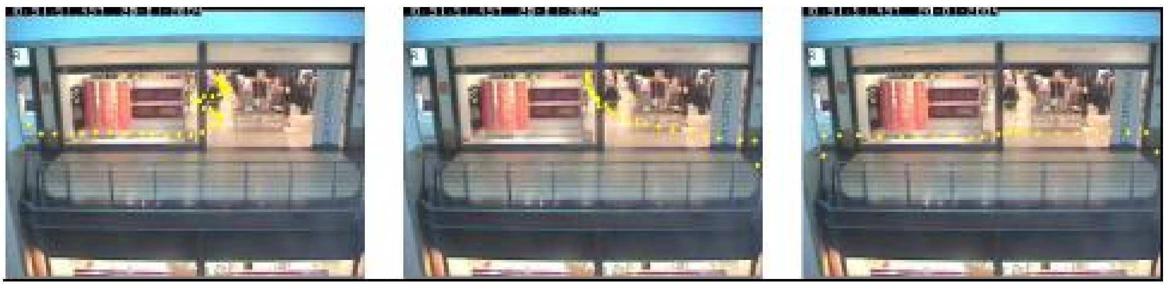


Figure 2.5: Examples of three different activities in a shopping centre (entering, exiting, passing) [NFM05].

Ragheb et al. in [RVRE08] described a mean power spectra-based method for classifying some human actions ['bend', 'jack' (jumping jack), 'jump', 'pjump' (jump on the same spot on two legs), 'run', 'side' (galloping sideways), 'skip', 'walk', 'wave1' (wave one hand), and 'wave2' (wave two hands)] from bounding boxes which contain the silhouettes of a human for a number of video frames representing a basic action.

Recently, researchers have been attempting to use advantages of human movement descriptors which provide useful knowledge about efficient parameters involving human motions. The human movement descriptors were invented mostly by choreographers, to interpret and annotate human movements (dance) by a number of symbols. Those symbols can be the targets to identify the proper human motion-based features.

2.3 Human movement descriptor

Several human descriptor systems were defined for annotating of human movement specially for choreography application. Beauchamp-Feuillet Notation was proposed by Pierre Beauchamp to record dances, in the 17th century. The basic elements of the Beauchamp-Feuillet Notation are the tract, symbols for the foot position, symbols for the step, signs for actions, turns and rhythm. Fig. 2.6 shows the bar of a French courante [Ret08].

Figure 2.6: Notation for bar of a French courante

The Benesh Movement Notation is another system of dance notation performed by Joan and Rudolf Benesh in the late 1940s. The system uses abstract symbols based on figurative representations of the human body. It uses a five line stave, which coincides with the head, shoulders, waist, knees and floor (from top to bottom). Same as a music score, it reads from left to right with bar lines to mark the passage of time. A number of signs are defined to notate the dimension and quality of the movement. Fig. 2.7 shows a sample of Benesh Movement Notation.

Dance Writing was first developed in 1966 by Valerie Sutton and extended to a greater body of work called Movement Writing. In this system stick figures are placed on a five-line staff where the movement proceeds from left to right. The lines of the staff represent levels of height (feet, knee, hip, shoulder).Fig. 2.8 shows the notation of a skateboard motion.

The Eshkol-Wachman Movement Notation (EWMN), is a system to annotate human or animal bodies positions and movement, invented by Noa Eshkol (1924-2007) and her student. EWMN was defined for movement notation, not a dance notation.

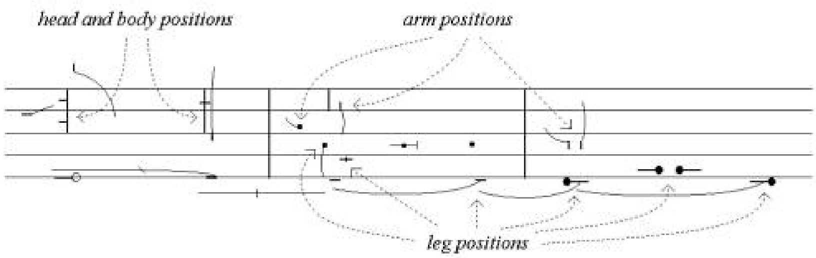


Figure 2.7: (Benesh Movement Notation) All information about body and limb positions is shown within the five-line stave

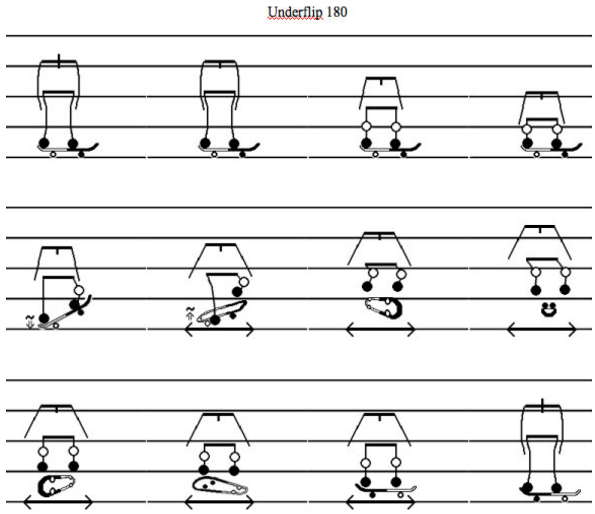


Figure 2.8: Movement writing : skateboard under-flip 180

Any form of human or animal movement without limitation of oneself to any particular style, can be described [Ret08]. The system uses a stick figure representation to describe the orientation of the different body limbs (Fig. 2.9 b)). The positions are expressed in a spherical coordinate system using a horizontal and a vertical component. It is often discretized in units of 45^o segments that are numbered from 0 to 7 (Figure 2.9 c)). The horizontal and vertical coordinates given by the sphere are written one above the other. Units of time are represented in grids from left to right, and limbs are written on a different line from top to bottom. The EWMN represents a good descriptor for spatial positions and the kinematic chains are not limited to the human body alone. The expressive content of movements are not addressed. A horizontally ruled notation page such as a sequence of spreadsheets which represents the body, is used to annotate movements.The page divided into columns by vertical lines to define units of time, and the symbols should be written from left to right, orderly. (see Figure 2.9 a))

Laban Movement Analysis (LMA) is a human movement descriptor, which consists

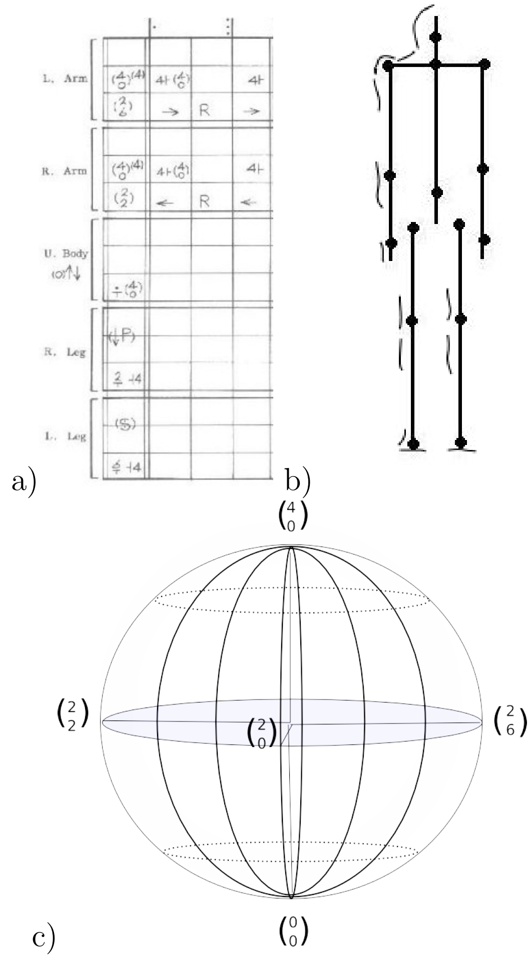


Figure 2.9: Eshkol-Wachman movement notation: a) Manuscript of movement b) Stick figure c) System of reference

five components (Body, Effort, Shape, Space, Relationship). Each component consists of several parameters that have some qualities to describe human movements. The LMA categorization and discretization inspired the computer scientists to model human movements computationally. In the following, we are going to introduce the LMA and the related state-of-the-art.

2.3.1 Laban movement analysis

Rudolph Laban (1879-1958) was a choreographer, and human movement theorist, whose work laid the foundations for LMA. One of his great contributions to dance was his 1928 publication of Kinetographie Laban, which is still used as one of the primary movement notation systems in dance. LMA is a system and language for observing, understanding, describing and notating various forms of human movement. LMA is used in several application as a tool by athletes, dancers, physical and occupational therapists [RDA08].

The general framework was described in 1980 by a student of Rudolf Laban, Irmgard Bartenieff in [BL80]. It was studied on dance and application to physical and mental therapy, widely [BL80], however it is not the same in the engineering domain. The group of Norman Badler was the first group, who started in 1993 to propose various computational models to formulate LMA components [BPW93]. For instance; a computational model of gesture acquisition and synthesis to learn motion qualities has been proposed in [ZB05]. Also recently but independently, neuro-scientist started to investigate the advantages of LMA to analyse certain effects on the movements of animals and humans. Foroud and Whishaw adapted LMA to capture the kinematic and non-kinematic aspects of movement in a reach-for-food task by human patients whose movements had been affected by stroke [FW06]. It was stated that LMA places emphasis on underlying motor patterns by notating how the body segments are moving, how they are supported or affected by other body parts, as well as whole body movement [RDA08]. Here we describe the LMA definition for the five components.

2.3.1.1 Relationship

Relationship describes modes of interaction with oneself, others, and the environment (e.g. facings, contact, and group forms). It is the only component which connects human movements to the environmental parameters. As Relationship appears to be

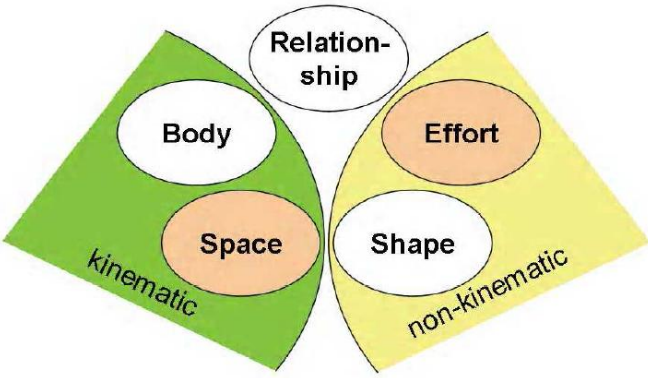


Figure 2.10: Five components of LMA; Body, Space, Shape, Effort and Relationship [RDA08]

one of the lesser explored components, some literature [FW06] only considers the remaining four major components [RDA08]. Several properties were defined to cover all possibilities of human relationships with others [RD13a], details are explained in the next chapter.

2.3.1.2 Body

Body indicates each body parts motion, related to the centre of body, the kinematics involved and the emerging locomotion. Body has various subcategories [RDA08]:

- Initiation of movement starting from specific body parts;
- Connection of different body parts to each other;
- Sequencing of movement between parts of the body;
- Patterns of body organization and connectivity, called " Patterns of Total Body Connectivity", "Developmental Movement Patterns", or "Neuromuscular Patterns".

2.3.1.3 Space

Space deals with the spatial range of the actor's Kinesphere (reach-space), and the exhibited form of the body trajectory. There is a category that describes and notates choices which refer specifically to space, paying attention to [RDA08]:

- Kinesphere: the area that the body is moving within and how the actor is paying attention to it.
- Spatial Intention: the directions or points in space that the actor is identifying or using.
- Geometrical observations of where the movement is being done, in terms of emphasis of directions, places in space, planar movement, etc.

The Space category is explored on Euclidean and non-Euclidean geometry.

2.3.1.4 Effort

Effort, or dynamics as sometimes named by Laban, concerns more subtle characteristics of the way a movement is performed with respect to inner intention. To be more visible, consider that the difference between pointing at someone whit the hand and punching someone in anger, in terms of body organization (both rely on extension of the arm), is small. The attention to the strength, control and the timing of the movement are the significant concerns of the Effort component. Effort contains four subcategories, each of them has two opposite polarities states(see Fig. 2.11).

- Space: Direct / Indirect
- Weight: Strong / Light
- Time: Sudden / Sustained
- Flow: Bound / Free

2.3.1.5 Shape

Shape concerns about non-kinematic deformation human body (as a blob) in the 3D space. The interpretation of Shape as a property of Body and Space probably is the reason for Irmgard Bartenieff to eliminate it from the major components of LMA. Foroud and Whishaw suggested [FW06] to group Body and Space as kinematic-based features describing changes in the spatial-temporal body relations, while Shape and

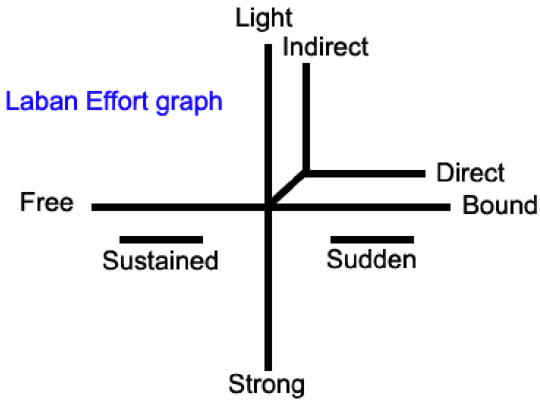


Figure 2.11: Laban effort graph

Shape	example	Plane
Enclosing /Spreading	Opening arms to embrace	Horizontal
Sinking / Rising	Reaching for something in a high shelf	Vertical
Retreating / Advancing	Avoiding a punch	Sagittal

Table 2.1: Shape qualities

Effort are part of the non-kinematic features contributing to the qualitative aspects of the movement [RDA08] as shown in Fig. 2.10.

Shape contains several qualities: Shape-Flow concerns with the mover's body shape within itself, the increasing or decreasing the body volume and the movement toward or away from the body centre. Rett and Dias [RD07] summarized the first three Shape qualities and express it in terms of spatial directions. By using a major direction we are able to express the Shape in the concept of Three Planes (vertical, horizontal, sagittal) (Tab. 2.1).

2.4 Feature extraction

We attempted to extract the proper features for human activity analysis with respect to the LMA components characteristics. Various works were proposed related to different human motion analysis, with respect to the body motion dynamics [CCZB00],human body shape deformations [STMea09],and with combination of both characteristics to analyse human gestures [Ret08], but they used only spatial-based features to obtain all the parameters. We propose to use frequency-based features to analyse Effort component which is related to the human body parts motion dynamics [KAQ+10, KAQ+11].

2.4.1 Frequency domain

Fast Fourier Transform (FFT) and Power Spectrum (PS) techniques are the well-known methods for feature extraction of a signal sequence data. FFT is an advantageous method for processing of frequency domain and analyzing spectrum, and PS which derives from FFT answers the question "how much of the signal is at a frequency"[Cro00]. In some research work these techniques were used in order to achieve proper features for various purposes related to human motion detection.

In our experiments, we analysed different signals (trajectory, velocity, angular velocity, acceleration and etc.) in frequency domain. Then a method is prepared to extract some features from acceleration signals of six parts of body (head, right hand, left hand, right foot, left foot and centre of the body) which are obtained by a motion tracker suit, its output can be seen in Fig.2.12. Acceleration signals of these parts of body are more representative of human motions. We used the advantages of the power spectrum technique to extract needed features in the LMA framework.

By having a signal data, its Fourier series should be calculated to estimate the PS of the signal [Cro00]. If f(t) is a finite-energy signal (such as acceleration signals), the Power spectrum $\Phi(\omega)$ of the signal can be achieved by:

$$\Phi(\omega) = \left| \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t)\, e^{-i\omega t} dt \right|^2 = \frac{F(\omega) F^*(\omega)}{2\pi} \tag{2.1}$$

where ω denotes the angular frequency and $F(\omega)$ is the continuous Fourier transform of $f(t)$, and $F^*(\omega)$ denotes its complex conjugate. If the signal is discrete with values

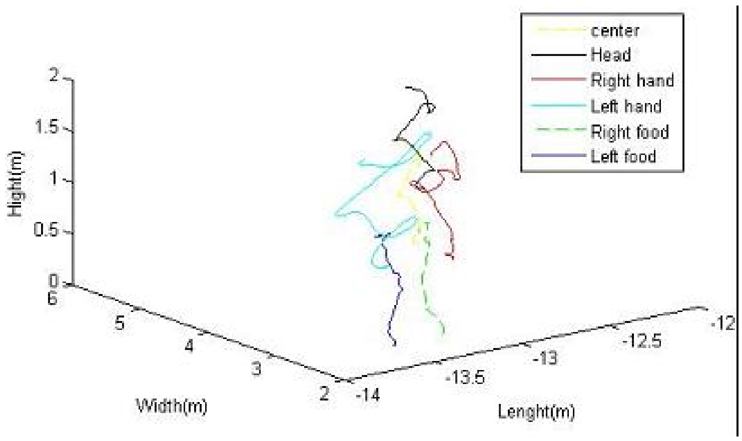


Figure 2.12: Some sequences position of the 6 parts of body for running action in 3D view.

f_n, over an infinite number of elements, we still have an energy spectral density:

$$\Phi(\omega) = \left| \frac{1}{\sqrt{2\pi}} \sum_{-\infty}^{\infty} f_n e^{-i\omega n} \right|^2 = \frac{F(\omega) F^*(\omega)}{2\pi} \tag{2.2}$$

Using Eq.2.3, 3D acceleration signals are calculated from the recorded data achieved by a motion tracker suit which provides a sequence acceleration for each part of the body in x, y, and z coordinate ($f_n(x)$, $f_n(y)$ and $f_n(z)$), separately.

$$f_n = \sqrt{(f_n(x))^2 + (f_n(y))^2 + (f_n(z))^2} \tag{2.3}$$

The PS signals of the 3D acceleration sequence data can be calculated by Eq.2.2, a sample of PS signal of an action.

In order to obtain features from FFT or PS signals, some researchers collected some coefficients (peak) of the extracted signals, like [SZJ+09] which collected first ten coefficients of each FFT result, however they just used one motion tracker which cannot present different human motions. In this study, however, we present a more comprehensive approach by dividing the frequency domain of PS signal of six parts of the body to several sub-domains (classes) to process more information and decrease the effect of possible noise. As [RVRE08] mentioned, the power of the PS signals for human motions are usually high in low frequency domains. Thus, the domain frequency is segmented in eleven sub-domains frequencies empirically, and in low frequency domain, the segmentation size is smaller than in the high frequency domain as can be seen in

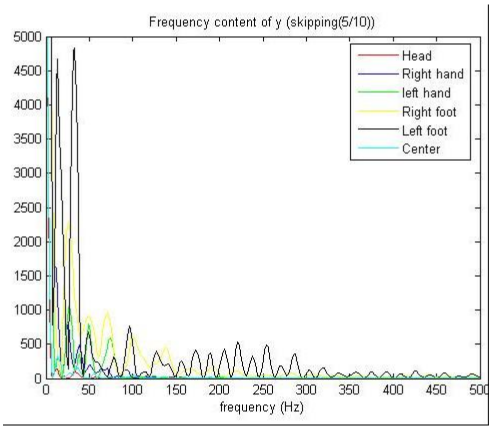


Figure 2.13: A sample of PS result of 3D acceleration signal of the 6 parts of body for skipping action

Fig.2.14. Moreover, the maximum of the content (power) of each frequency sub-domain is calculated. Thus, eleven features (peak) for six parts of body are defined to be used for classification of various actions [KAQ+10]:

$$F = Max\left\{{}^{Acc}f_i^{pb}\right\},$$

where the pb denotes set of body parts, Acc denotes the acceleration signal data and $Max\left\{{}^{Acc}f_i^{pb}\right\}$ denote the maximum content of each i subdomain frequency of acceleration signal for each pb. The set of pb and subdomain frequency are as follows:

pb={Head, Left hand, Right hand, Left foot, Right foot, Body center}

subdomain frequency ={(0 - 10), (11 - 20), (21 - 30), (31 - 40), (41 - 50), (51 -100), (101 - 150), (151 - 200), (201 - 300), (301 - 400), (401 - 512)}

2.4.2 Spatial domain

The frequency-based data, which is obtained from the acceleration signal, is the proper feature to distinguish between similar movements e.g. walking compared to running, or sitting compared to falling down. However, in some cases, they are not the same and have difference in terms of spatial property, e.g. sitting compared to rising, and running compared to falling down. Thus, there should be other features to distinguish

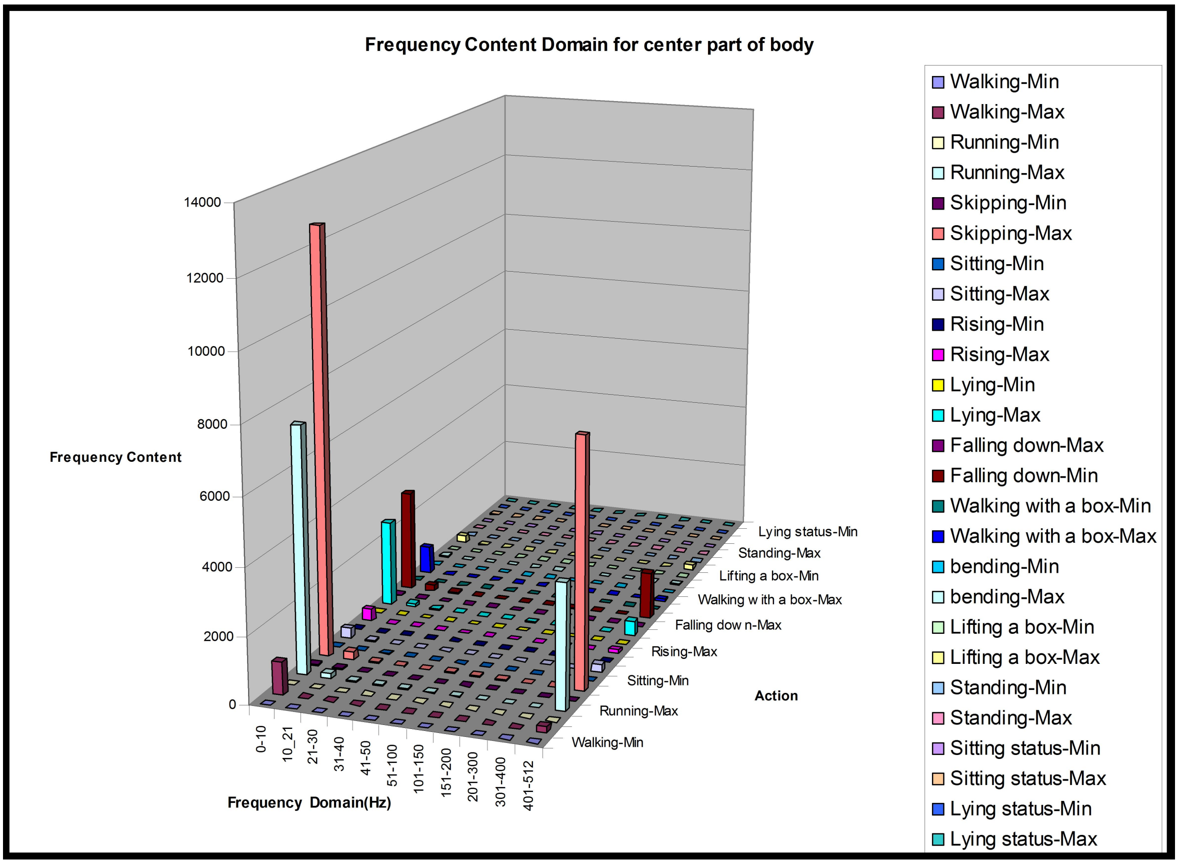


Figure 2.14: A histogram belongs to a body part, that is showing the definition of some frequency sub-domains on the power spectrum signals and minimum and maximum content of each frequency sub-domain for each action separately.

those movements, that is the spatial relation of two objects like head and feet during the different movements. For instance; in sitting and falling down movements, distance of those body parts (height of the human blob) is reducing and in a rising movement there is an opposite situation. However in standing or walking, there is no considerable change in this aspect. Therefore, the difference of distance between head and feet during a movement can be obtained:

$$\Delta D = \sum_{i=2}^{n} \left(\left(X^{i}_{obj1} + X^{i}_{obj2} \right) - \left(X^{i-1}_{obj1} + X^{i-1}_{obj2} \right) \right), \tag{2.4}$$

where X^{t}_{obj1} and X^{t}_{obj2} denote 3D position of the two objects $obj1$ and $obj2$ at frame i, respectively. n denotes the number of frames inside of the window signal, and ΔD denotes the difference distance between those objects during the window signal in the meter unit.

It should be mentioned that in what concerns human movements analysis in terms of spatial domain, there are plenty of works which attempt to formulate it with respect to Space and Shape components [ZB05, Ret08, SD10], but not to Relationship's component of LMA framework. Thus we explored the features in spatial domains with respect to the Relationship's component which will be described in the next chapter.

2.5 Conclusion

In this chapter, we explained Laban movement analysis (LMA) system and the related works which have been done by using descriptive methods such as LMA. Then, we proposed a novel approach to extract human motions-based features through the LMA framework by conjugating two different types of feature domains. Frequency-based features which are extracted by using the FFT and PS techniques, with respect to analysing the Effort component of LMA which relies on the quantification of energy during human motion. Spacial-based features that rely on the distance between objects were introduced in order to analyse the spatial components of LMA (Shape, Relationship). This will be the subject of the next chapter. The following chapters will use the mentioned features as inputs to analyse human behaviour in different contexts.

Chapter 3

Human behaviour analysis

3.1 Introduction

Human behaviour is highly connected to environmental constrains and context. For instance; between surveillance and elder-care applications, or between individual and group context, the relevant human behaviours are different. Thus, researchers used to consider constraints for environment and scenario, to be able to obtain interesting results. Depending on the problems and constraints, the proper models were proposed. For instance; Pentland and Liu in [PL95] discussed human behaviour modelling in a car driving scenario. They believed that it is useful to have a number of dynamic models to classify the driver's behaviour by comparing it with the models, for each kind of driving such as relaxed driving, tight driving, etc.

Remagnino and Jones [RJ01] used a Hidden Markov Model (HMM) approach to model parking lot environment behaviour. In [ORP00] Oliver et al. defined a number of sequences of human movements to understand the relevant people behaviours. The main constraints, in the mentioned works, are that there are limited to some specific human actions or behaviours. Most of them are not extendable to other levels of human activity analysis, or generalizable to other possible applications.

In this study, a bottom-up strategy is proposed. In the first step, the LMA parameters are classified and the second step, based on the obtained LMA parameters, human movements are recognized. In the last step, semantic-based human-object and human-human interaction are investigated (as a scene understanding application). Therefore, this chapter is divided into three sections; individual-based human action analysis, context based human behaviour analysis, and semantic-based scene understanding. In the first step, body motion-based features are explored to analyse human movements. In the second part, we attempted to explore through existing relationship between body movements of a person with him/herself, an object, and another person. Thus, the Relationship component of LMA, which is the less explored LMA component, is analysed and formulated. Semantic-based scene understanding section proposes an improved approach to understand human activities by using context-base knowledge.

Fig.3.1 presents the related levels of proposed human behaviour analysis framework in this chapter.

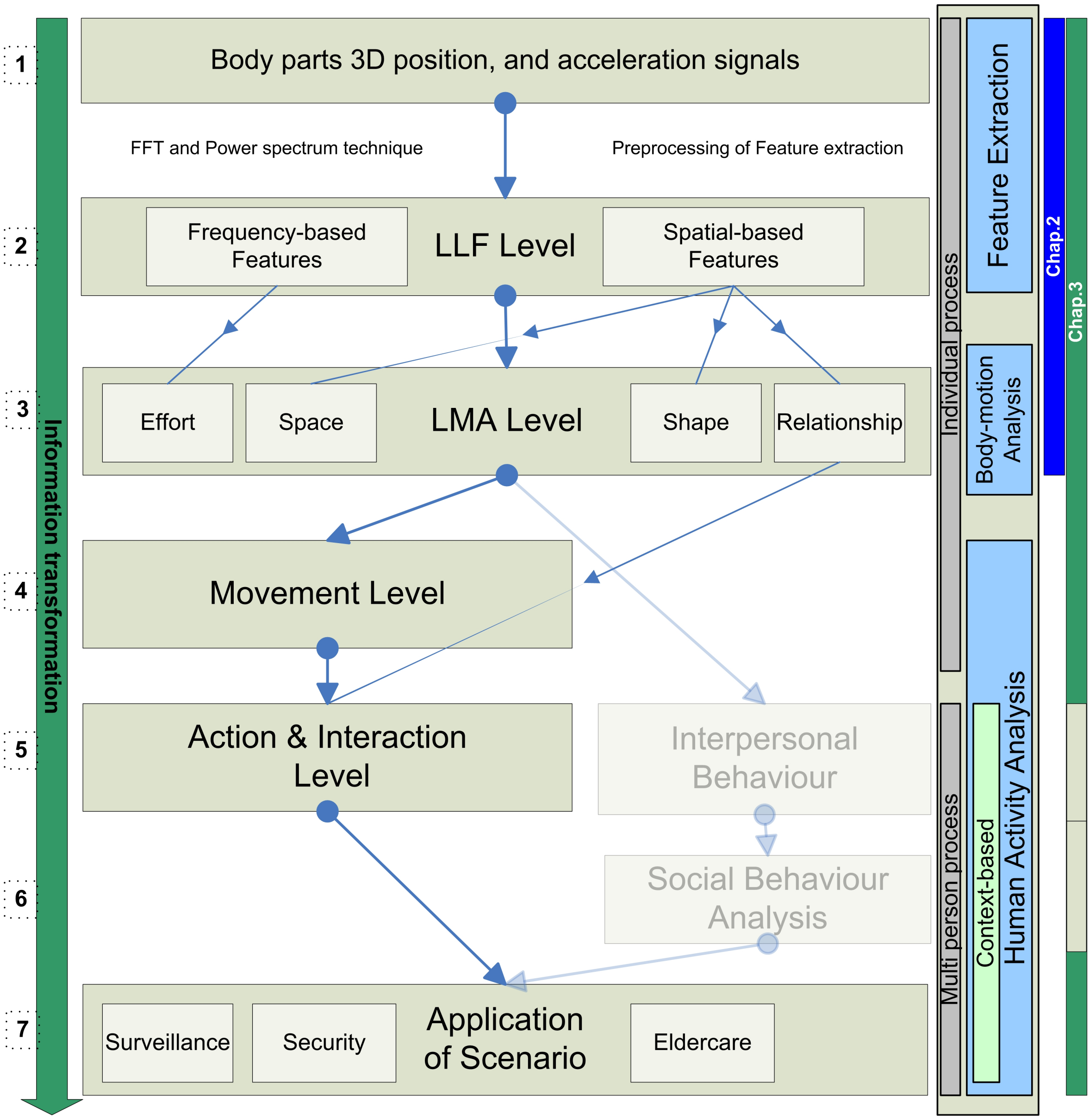


Figure 3.1: Human behaviour analysis in different levels with respect to this chapter

3.2 Individual-based human action analysis

Human action modeling is one of the popular challenges in artificial intelligence science. There are several applications related to the subject, for instance; surveillance (e.g. airport, bank, train station, etc.), virtual reality (e.g. interactive virtual worlds, virtual studios teleconferencing, etc.), motion analysis (e.g. choreography of dance, clinical studies of orthopedic patients, etc.) and Human-Robot Interaction (HRI).

To better understand the importance of the studies in the existing potential applications, we present an example for fall detection case. Falls are among the top causes of unintentional injury and death in the elder population. Several studies point out for the social impact of this problem, which reaches global scale [Rub06] [CRG+05]. Falls in older people are estimated to affect approximately 30% of those over 65 in an annual basis. Although most falls cause no injury, between 5-10% of the elderly who fall each year sustain serious injury, such as fracture, head trauma, or serious laceration. Approximately 15% of elderly people who fall and injure themselves require hospitalization. Hip fractures are the most serious fall-related injury (95% of hip fractures are due to falls). Among the elderly who fall and sustain a hip fracture, 20-30% die within one year of the fracture. As many as two thirds of the elderly with hip fracture never regain their pre-fracture activity status and one-third require nursing home placement. Consequently, the economical impact related with falls is expected to reach near $55 billion in the USA. World wide efforts are spent in order to work in the prevention of these situations either by studying methodologies for prevention as [PRO] or designing fall detection systems [WCL+08] which try to help in nursing tasks.

Given these facts, human behaviour analysis can contribute with a strong point both on the prevention and detection of this type of hazardous situations. Systems monitoring the elderly living space could analyse potential risks of falls occurring and identify potential causes for falling, consequently leading to correct adaptations in the living space. In terms of fall detection, it would be advantageous, for those situations where full monitoring is not possible, to have systems with the ability to warn caregivers about abnormal situations.

Human behaviour comes from the different human actions and reactions, which is usually shown by a person's body motion, voice and facial expression. Relevant researchers investigate all the possibilities to recognize human behaviours depending on their application limitations [PPNH06].

Nowadays, most of the related applications such as surveillance and elder-care, rely on human movements. Bobick [Bob97] presented a survey related to human movement, activity and action. As Bobick said, the movements are the lowest level of human motion which doesn't need any previous knowledge and contextual information to be identified, but to understand human action or behaviour we need to recognize a sequence of human movements or states, related to the corresponding environment and scenario.

A common approach to obtain human body parts poses to apply human behaviour analysis techniques, uses wearable motion trackers. Those trackers are more reliable than the vision-based trackers, however, they have their own drawbacks such as inapplicability in several real world applications, and have missed data in sudden human motions. We use a special suit (MVN®) which has several IMU's attached, to collect the interesting data such as positions and accelerations of human body parts, in 120 Hz resolution.

One of the basic problems is to collect proper Low Level Features (LLFs) and observable information which depend on the scenario and input data. To extract the features, the existing approaches can be categorized in 2D (image) and 3D. Recently, researchers are interested in using 3D data to reconstruct any object such as human body to avoid image based approach limitations,although a corresponding process is usually more time- consuming.

Leo et al. in [LDS04] introduced an approach for complex human activities detection from image sequences in outdoor environment such as archaeological sites (e.g. walking, probing the subsoil with a stick, damping the ground with a tank and picking-up some objects from the ground).

Arsic et al. in [AWSR05] proposed a real-time behaviour detection method which is video based. Their application is for passenger behaviour detection in public transports such as airplanes. They defined some special human behaviour such as aggressive, nervous, tired, kid and talk.

Nascimento et al. in [NFM05] described a method for recognizing human activities in a shopping space. They used human motion patterns which are achieved from a sequence of displacements of each human's blob center. They modeled trajectories of the human's blob by using a method which they called a multiple dynamical models with a switching mechanism. Finally, they estimated the identification of the models which connected with a trajectory.

Yue et al. in [YZC03] proposed an image-based method which used a human body parts segmentation approach for covering the weakness of visual hull method for concave regions reconstruction. The first result was a virtual silhouette image fit to the given viewing direction, which used human body part localization method. Body parts produced separately in virtual view from the corresponding input views and then assembled together. The last silhouette image was used for removing the separate or squeezed region in final view.

Researchers have been attempting to use the advantages of human movement descriptors which include useful knowledge about efficient parameters involved in human motions. LMA is a well-known relevant system, which in a number of literatures like [Zha01, RDA08, NMS02, CCZB00, AKA+02] attempted to formulate its parameters to interpret human movements with an intelligence system.

Rett & Dias in [RD07] presented a real-time system that analyzes human movements as a contribution to the field of Human-Machine Interaction (HMI), based on the concept of the LMA. A Bayesian model used for learning and classification. They presented the LMA as a concept to identify useful features of human movements to classify human gestures based on vision and motion tracker data.

In this step, we attempt to distinguish various human movements by estimating human body parts motion dynamics. For that purpose, the relevant LMA component is Effort, which deals with the dynamic qualities of the movement and the inner attitude towards using energy [RD07]. However, Rett [Ret08] proposed a method to use LMA components specially Effort to detect human gestures, but as it was mentioned in previous chapter, we proposed different techniques (in frequency domain) to solve the problem. By having the other relevant knowledge [Ret08], and examining different existing data (trajectory, velocity, acceleration, etc.), the acceleration signal is selected as proper data to extract the needed features.

There is a number of limited works which explore features through frequency domains. Khalid and Naftel in [KN05] proposed an approach which is based on DFT technique for clustering and classifying spatio-temporal object trajectories using an artificial neural network approach. They used the Fourier coefficient features to recognize similar motion patterns.

Shi et al. [SZJ+09] introduced an approach to recognize five different human movements using body acceleration signals. They used inertial data (accelerometers, gyroscopes) signals obtained from an inertial sensor, which was attached to a person. Fourier analysis was used to extract the needed feature from the human motion signal.

Cheng et al. in [CCK04] used PS technique as a feature extraction method to classify some periodic human motion from sports video signals. Ragheb et al. [RVRE08] proposed a method for human movement recognition from video streams. They used mean power spectra technique to extract interesting features from the bounding boxes, which contain the silhouette of a human for a number of video frames, representing a basic movement. The two last works used PS technique and 2D (image-based) analysis. However, 3D data is used in this study.

3.2.1 LMA-based feature extraction

In Chapter. 2, it was explained how the features were extracted in two different domains; frequency and spatial. It was mentioned that in this study, the Effort component is explored in frequency, and Shape component in spatial domain.

3.2.1.1 Shape Component Characterization

The Shape component describes deformation in the human body shape (as a blob) through his/her limbs movements in three planes; vertical, horizontal and sagittal. There are three quantities for discretizing of the human shape deformation on each of the three planes:

$$\begin{cases} Vertical\,Plane & \in & \{Sinknig, Still, Rising\} \\ \\ Horizontal\,Plane & \in & \{Enclosing, Still, Spreading\} \\ \\ Sagittal\,Plane & \in & \{Retreating, Still, Advancing\} \end{cases}$$

In this study, the human shape's deformation in vertical plane is modeled, which can be extended to the other planes easily. Difference in height of human head and feet related to the body centre is used as a feature to analyse the shape's changes in the vertical plane:

$$\triangle H = \sum_{t=2}^{n} \left(\left({}^{z}X^{t}_{head} + {}^{z}X^{t}_{feet} \right) - \left({}^{z}X^{t-1}_{head} + {}^{z}X^{t-1}_{feet} \right) \right),$$

where ${}^{z}X^{t}_{head}$ and ${}^{z}X^{t}_{feet}$ denote position of head and feet related to the body centre in z axis at time t, respectively. n denotes the number of frames inside of the sliding window signal, and $\triangle H$ denotes the difference in height of human body between the first and end of the sliding window signal in the centimeter unit. By having the result of the equation, qualities are estimated by discretizing the data, empirically:

$$Shape_{ver} = \begin{cases} Rising & \triangle H > 0.2 \\ Still & 0.2 \geqq \triangle H \geqq -0.2 \\ Sinking & \triangle H < -0.2 \end{cases} \tag{3.1}$$

3.2.1.2 Effort component characterization

One of the most important components of LMA is Effort, and we attempted to characterize it in frequency domain. As it was mentioned before, Effort is a component dealing with the dynamics of motion for understanding the more subtle characteristics about the way a movement is performed. The difference between punching someone in anger and reaching for a glass is slight in terms of the physical body organization as they both rely on extension of the arm. However, the control and timing of the motion in each of the cases are very different. Effort consists of four quantities, and each of them has two bi-polar states. Dependencies between the LLFs and the Effort property are discussed in the previous chapter and [KAQ+10].

In [KAQ+10], one of the Effort properties (Effort-Time), is characterized by frequency-based features, which estimated from 3D acceleration signals of body parts with respect to a global reference. The obtained results showed a number of false Effort-Time estimations which affected the final classification results. Thus, obtaining LMA-based human motion parameters with respect to a local reference (human centre as an origin of the coordinate system) provides more accurate data. It follows the LMA definitions, and causes less data redundancy, and consequently, resulting in an improvement in human motion description analysis. For clarification purposes, we describe the analysis of the walking movement: by analyzing the data relative to body centre, the computed energy of body parts signals related to the body-centre is reduced. We describe a normal human walking movement as having motion on feet and hands, not on head. If we consider human body parts poses with respect to a global reference, we will see some

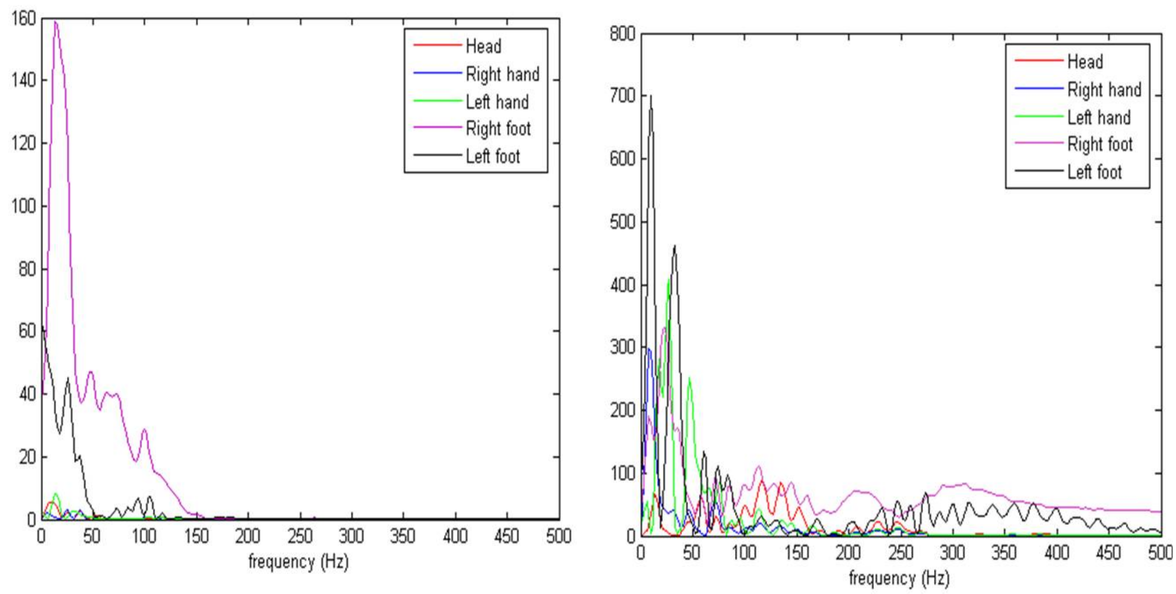


Figure 3.2: A couple of samples for corresponding PS result to the acceleration signals of body parts for (left) walking and (right) running movements.

energy on head motion, because of the existing locomotion, which can be reduced by using a local reference-based analysis. Consequently, local reference-based analysis will not distinguish between walking on the spot and normal walking.

To explore through frequency domain features, the advantages of PS technique is used. The PS (obtained by Fast Fourier Transform (FFT)) of the acceleration sequence data is calculated by equation (3.2) (as can be seen in Fig. 3.2):

$$\Phi_{bp}(\omega) = \left| \frac{1}{\sqrt{2\pi}} \sum_{-\infty}^{\infty} {}^{c}f_{bp} e^{-i\omega n} \right|^2 = \frac{F_{bp}(\omega) F_{bp}^{*}(\omega)}{2\pi} \tag{3.2}$$

where ω denotes the angular frequency and $F_{bp}(\omega)$ is the continuous Fourier transform of ${}^{c}f_{bp}$ which is body-centred frequency-based feature for bp's body part, and $F_{bp}^{*}(\omega)$ denotes its complex conjugate.

The sufficient features in the mentioned purpose, are obtained by the extracting PS signal of each selected body part acceleration signal for various movements, and then by collecting the first four coefficients (peak) of the extracted PS signals [KAQ+10]. Thus, the four coefficients for each of the five body parts, provide the input feature space for classification of various movements in Effort aspect. We named the coefficients as $Max\left\{{}^{c}f_{bp}^{i}\right\}$ which is the maximum content of each i sub-domain frequency of the PS

Definition:	Description
Description:	a declarative part
Specification:	presents preliminary knowledge
Pertinent variables:	relevant defined variables
Decomposition:	probability distribution
Parametric forms:	prior programming forms
Identification:	experimental dataset information
Question:	Computing a probability distribution

$$
Program \begin{cases} Description \begin{cases} Specification \begin{cases} Pertinent\,variables \\ Decomposition \\ Parametric\,forms \end{cases} \\ Identification\,based\,on\,Data \end{cases} \\ Question \end{cases}
$$

Figure 3.3: Structure of a Bayesian program and its brief descriptions [DBM03].

signal for each bp. The set of bp and sub-domain frequency are defined as follows:

$bp \in Head, Righthand, Lefthand, Rightfoot, Leftfoot,$

$Frequency\,domain\,class \in (0-10), (11-20), (21-30), (31-40).$

3.2.2 Bayesian programming

Bayesian Programming (BP) is a method to describe relevant variables, parameters and Bayesian rules to interpret the proposed model precisely. The BP was proposed by Lebeltel et al. with the details about this methodology and for a robotic application [LBDM04]. Diard et al. presented the capability of the BP methodology to interpret different probabilistic models [DBM03].

Figure 3.3 presents the BP formalism structure, which consists of two main parts;

- Description: it is a declarative part to specify a method to produce a joint distribution on a set of variables given a set of experimental data and preliminary knowledge. It divides into a couple of sections;
  - Specification: presents preliminary knowledge which constructed from following subsections;

      * Pertinent variables: the relevant variables on which the joint distribution is defined.
      * Decomposition: a probability distribution that gives the probability of each defined state of variables.
      * Forms: it assigns forms to each term of the selected product (which can be parametric forms, or recursive questions to other Bayesian programs). If there are free parameters in the parametric forms, they have to be assessed. They can be given by the programmer (a prior programming) or computed on the basis of a learning mechanism defined by the programmer and the experimental data [DBM03].
   - Identification: it explains about the dataset information.

- Question: it is for computing a probability distribution of the form $P(Searched\,|\,Known)$ using the previously defined descriptions [DBM03].

3.2.3 Bayesian-based modeling

Bayesian Network (BN) presents many advantages on using prior knowledge and modeling the dynamic dependencies between variables states. One of the key characteristics of BN is its ability to fuse different types of data in one model. This work developed a classification model based on BN, justified by this flexibility characteristic. Thus, applying Bayesian theory allows the preparation of a general LMA-based human motion analysis framework to fuse multi-modal data, e.g. data emerging from frequency and spatial domains.

Fig. 3.4 presents the BN with two levels in three abstraction layers. In the first level of the BN set, we provided six independent parallel BNs for the five body parts in Effort aspect and one quantity (vertical plane) in Shape aspect.

The set of LLFs emerge from $\left\{Max\left\{{}^{c}f^{i}_{bp}\right\}\right\}$ which has four states {No, Low, Medium and High}, are obtained using the same threshold definition as in [KAQ+10]. Effort-Time subcomponent $\{\,Ef.T^{bp}\}$ for a body part has a couple of states: Sustained and Sudden. The $\triangle H$ is discretized in three states {Up, Still, Down} by Eq. 3.1, corresponding to the Shape's states which have been explained before.

A one-second slide window which moves on signals by a half of second step is defined. It is a common rule to not lose information through the long signal data processing.

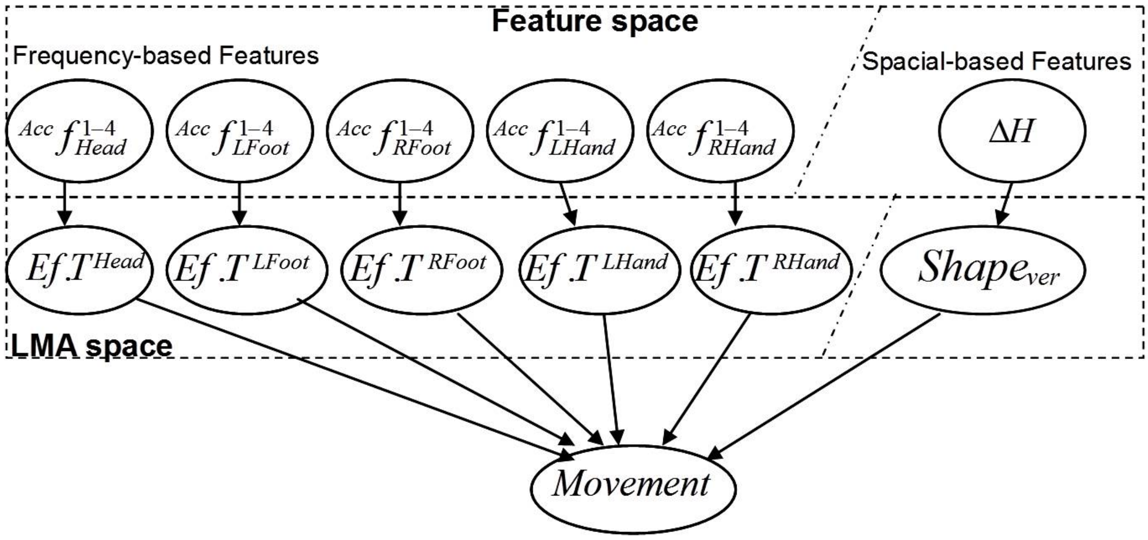


Figure 3.4: Global Bayesian network

To obtain a more reliable result, the training dataset is provided by using the most representative one second of each action sample.

After doing the learning process of the model, we are able to answer the Bayesian questions for the probability of the LMA parameters given the observed features $\{P(LMA|LLFs)\}$, and the probability of a movement given the LMA parameters $\{P(M|LMA)\}$. In [KAQ+10] we can see more detail about the proposed BN.

Fig. 3.5 presents the BP's corresponding of the proposed human movement model, with the details.

3.2.4 Experimental results

The experiments are done for various kinds of human movements (at least 10 times for each type of movement in different durations). A motion tracker suit is used to obtain human body parts position and acceleration signals. A one-second window-frame which shifts half a second, is defined on the signals, to collect the features from the sequence of data (Chapter 2).

A free Bayesian toolbox provided by Kervin Murphy and Berkeley [MB05] is used. The structure which is presented on Fig. 3.4 is implemented by using the toolbox.

Table 3.1 presents the results for walking, running and standing activities, with considerably high positive matches based only on the Effort.Time component, but in

Program / Description / Specification

Pertinent variables :

$Pb \in \{Head, Left\,hand, Right\,hand, Left\,foot, Right\,foot\}$

$f_{i=1:4}^{p\in\{pb\}} \in \{No, low, Medium, High\}$

$\triangle H \in \{Up, Still, Down\}$

$Ef_{time}^{p\in\{pb\}} \in \{Sustained, Sudden\}$

$Sh_v \in \{Sinking, Still, Rising\}$

$Movement \in \{Walk, Run, Sit, Rise, Fall, Stand\}$

Decomposition :

$$\left(P\left(Movement \prod_{p\in\{pb\}} Ef_{time}^{p} \prod_{p\in\{pb\}}^{i=1:4} f_i^p \triangle H\, Sh_v\right) = \right.$$
$$\prod_{p\in\{pb\}}^{i=1:4} \left(P\left(f_i^p\right)\right) \prod_{p\in\{pb\}} \left(P\left(\prod^{i=1:4}\left(f_i^p\right) \middle| Ef_{time}^p\right)\right)$$
$$\left. P(\triangle H)\, P(\triangle H\,|Sh_v)\, P\left(\prod_{p\in\{pb\}}\left(Ef_{time}^p\right) Sh_v\,|Movement\right)\right)$$

Parametric forms :

$P\left(f_{i=1:4}^{p\in\{pb\}}\right) : uniform$

$P\left(f_{i=1:4}^{p\in\{pb\}} \middle| Ef_{time}^{p\in\{pb\}}\right) : histogram(Matrix)$

$P(\triangle H) : uniform$

$P(\triangle H\,|Sh_v) : histogram(Matrix)$

$P\left(Ef_{time}^{p\in\{pb\}}\, Sh_v\,|Movement\right) : histogram(Matrix)$

Iden.

Ques. : $P\left(Movement \middle| \prod_{p\in\{pb\}}^{i=1:4}\left(f_i^p\right) \triangle H\right)$

Figure 3.5: Bayesian Program for the human movement model.

	Walk	Run	Sit&Bend	Rise	Fall	Stand	%
Walk	102	4	0	0	0	3	93.57
Run	1	67	0	0	0	0	98.53
Sit&Bend	0	0	6	0	0	20	23.07
Rise	0	0	0	0	0	10	0
Fall	17	4	8	0	5	4	13.16
Stand	0	0	0	0	0	105	100

Table 3.1: Classification result using body-centered based features in frequency domain(Effort)

	Walk	Run	Sit&Bend	Rise	Fall	Stand	%
Walk	56	0	0	0	0	0	100
Run	0	60	0	0	2	0	96.77
Sit&Bend	0	0	41	0	0	1	97.6
Rise	0	0	0	24	0	0	100
Fall	0	0	0	1	22	0	95.7
Stand	0	0	0	0	0	109	100

Table 3.2: Classification result using body-centered based features in both frequency and spatial domains(Effort and Shape)

the other activities which are more relevant in the Shape aspect, it completely failed [KAQ+10]. Table 3.2 presents the results of our idea by conjugating the Effort and Shape components.

The results demonstrate that the couple of components are sufficient to distinguish the performed human movements. Effort distinguishes between walking and running, and sitting and falling movements properly. As well as, Shape is the right component to distinguish between sitting and rising, and also between all activities which occur in standing state, like walking and running and standing, and others.

3.2.5 Conclusion

In this study, we presented a novel approach to understand human movements, based on the LMA concept by the conjugation of two different types of feature domains, and using a 2-step probabilistic approach. Frequency-based features, which were extracted by the PS of acceleration signals, are used to analyse the Effort-Time subcomponent of LMA which relies on the quantification of energy during human motion. For the Shape component which relies on the human shape changes on the vertical plane, the variation of human height is applied.

BNs are used in a couple of steps and in different domains (frequency and spatial) to obtain the probabilities of the LMA parameters and human movements. It is shown that, to analyse various human body movements, the dynamics of human body motions and the human body shape changes in the space are needed. Thus, having a combination of frequency-based features to analyse dynamics of human body movements, and special-based feature to analyse deformation of human body in the 3D space, is sufficient to estimate the performed human movements, as can be seen in the presented confusion tables.

3.3 Context-based human behaviour analysis

This section presents a new approach to analyse human activities based on existent relations between different human body motions properties and environmental parameters. Different kinds of human body motion-based features such as; position, velocity, acceleration, trajectory, etc. have been used to analyse human movements and activities. Those features are important but the relation between the features also is a key characteristic of analysing different types of human actions and interactions. These relations can be defined in three levels. In the first level, the relations come from between different motion parameters inside of a body (between body parts). Then the relations between a person and environmental objects are needed to analyse human activity in a scene. Finally, human-human interaction can be estimated by defining the relationship of human motion characteristics that belong to a couple of people. For instance; in a normal walking action, between legs and hands there is a harmony which is very important to complete the action. In another example,when clapping one's hands, the hands touching each other is the relation of these two body parts. It appears also between a body and an external object; such as reaching to or grasping a glass, and also between a person and another body, such as hand shaking, punching, etc. In this study, an approach is presented to analyse the different types of human activities based on the relationships.

The idea came from one of the LMA components, which provides a way to explain the relationship between a human body part motion and other objects (other body parts of her/himself, environment and other bodies). Those relationships can appear widely, and having a global framework that could be able to explain them as much as possible, is a big challenge that we attempted to prepare using the LMA system framework.

Fig. 3.6 presents a sequence of human movements, which can be analysed as a sequence of human interactions with a robot. That analysis is useful in different applications such as; smart-home, surveillance, human-robot and human-human interaction, etc.

In the previous parts of the study, several features, based on the LMA concept, are defined to analyse human movements. In [KAQ$^+$10] an approach is presented to use frequency-based features to estimate human movements based on some body parts acceleration signals. Then in [KAQ$^+$11], spatial-based feature also is used to improve

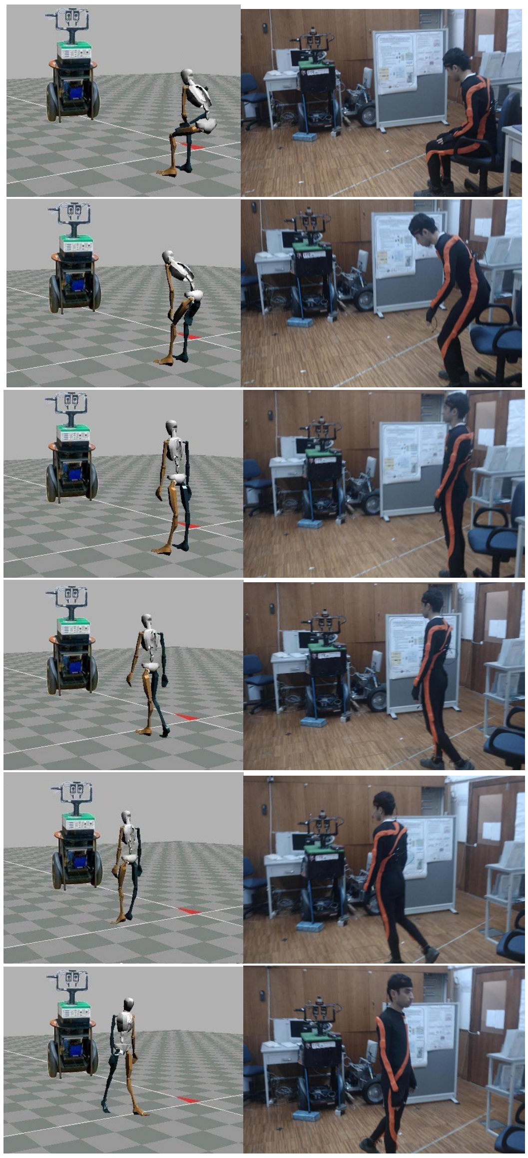


Figure 3.6: A sequence of human movements with respect to a robot. The images in the right side are showing a person which dressed a motion tracker suit (MVN), and the left side are presenting the corresponding skeleton (obtained by the motion tracker suit) of the person.

the previous estimation results. In this work, we attempt to define the relationship parameters using the previous obtained features, and also environmental parameters. To implement the idea, BN and Hidden Markov Model (HMM) as a Dynamic Bayesian Network (DBN), which are popular approaches in this area, are used. The capability of the system generalization, because of using LMA framework and Bayesian approach, is one of the main advantages of this study.

3.3.1 Previous related works

Analysing human motion is a prerequisite to understand human activities, such as human behaviour, human-robot and human-human interaction, etc. In Bobick's terminology, the movements are the lowest level of human motions which do not need any contextual and previous knowledge to be identified. To understand human action or behaviour we need to have a sequence of human movements or states with respect to the environment or scenario [Bob97].

There are many different kinds of human movements which are interesting for researchers to recognize in various applications. Each of them can be identified with position variations of one or more involved human body parts, without the necessity to know the underlying movement concept. Thus, researchers attempt to track body parts in a sequence of captured data such as image sequences, being able to estimate human dynamic characteristics.

An accurate human movement tracking is essential for a precise human interaction analysis. Feet, hands, head and face are the most important body parts to track in different relevant applications. Some examples can be given such as; gait recognition, where the dynamics of leg motions are important, facial expression recognition, the focus lies solely on the face, while in gesture recognition, all parts of the body can be involved, where hands and head stand out as the most important ones.

As Bobick [Bob97] says, to analyse human activities, we need to know the underlying movement concept. It means to understand human activities; the relations between human movements and environmental parameters need to be considered. There are infinite relations that can appear through human activities in any scene. Thus, researchers always define specific relationship properties to propose their methods. For instance; Rao et al. in [RYS02] presents a computational representation of human action to capture the changes using spatio-temporal curvature of 2-D trajectory of hands.

Then, in the experiment part, a limited number of activities such as picking up an object from the floor and putting it down on the desk were defined. Thus, we can infer that we need a general framework that is able to analyse those kinds of relation parameters in mid-level. LMA is the system that can be used as the mid-level features of human motions. To have this mid-level, several works have been done for various types of features, based on the LMA systems [FW06, KAQ+11, KAQ+10, RDA08, ZB05]. All of the previous works were not aware of the Relationship component parameters of LMA, but in Effort, Shape and Space.

As it was mentioned, a BN can model the dependencies between parameters. Those dependencies between parameters somehow play some of the relationship roles, but they are very dependent on the learning process. The point is, the relation parameters were not defined explicitly on the model, thus, it is very reliant on the learning process. Therefore, by considering those relationships properties, such as the input features of the model, more reliable results can be obtained. Finally, by connecting human body motions to the scene, which is one of the Relationship component purposes, the probabilistic model for analyzing human interactions is explored.

3.3.2 Features extraction

Human motions consist of a number of features which are defined and extracted in different domains. In the previous steps, frequency-based features [KAQ+10] and spatial-based features [KAQ+11] were extracted to analyse human movements. The results presented that by conjugating the different types of features, the classifier's efficiency significantly improves. Thus, we used those features to rely on existing relations between human body parts motions and environmental parameters. Based on the relationship definitions which are described in the next section, various relations can be defined. Therefore to perform an experiment and depend on the activities, we need to collect the proper features which can disclose differences in the activities perfectly.

3.3.2.1 Body motion based features

In this step, we estimate some general human activities, like walking, running, siting, rising, falling down and standing, which can be extended to more types of movements, without using environmental parameters. Thus, we attempt to define the Relationship parameters between body parts of a person.

Relationship characteristics are very wide, and play an important role in performing any activities, for instance; in walking types movements, usually there are harmonic motions in the hands and feet related to the body centre. Those harmonic motions can be estimated by frequency-based properties [CCK04]. The frequency-based features can be extracted by FFT and PS techniques which are the known approaches [RVRE08]. By exploring different collected signals (trajectory, velocity, acceleration) from body motions, the acceleration signals of human body parts related to the body centre are selected (based on the previous work [KAQ+10]), and then FFT and PS signals of the acceleration signals are extracted [KAQ+10] (see Fig. 3.7-top-left and Fig. 3.7-top-right).

As [RVRE08] mentioned, power of the PS signals for human motions is usually high in low frequency domains. Thus, based on the previous work [KAQ+10], the peaks of PS signals in the first four frequency sub-domains (1-10 HZ, 11-20 HZ, 21-30 HZ, 31-40 HZ) for different movements are collected as the LLFs. Other frequency domains data can be achieved, but in this kind of applications the selected domains are more representative and sufficient. Fig. 3.7-Left-down) presents a histogram of the frequency content for a specific body part for different movements, therefore, a number of histograms for different body parts is generated.

The frequency-based data, which is obtained from the acceleration signal, is the proper feature to distinguish similar human movements like walking and running, or sitting and falling down, but not for the others, which have difference in terms of spatial property, like sitting and rising, running and falling down. Therefore, other useful features to distinguish those movements are the relation of two objects such as head and feet in terms of the relation between those body parts during the different movements. For instance; in sitting and falling down movements, the distance of those body parts reduce and in a rising movement there is an opposite situation. However, in standing or walking, there is no considerable change in this aspect. Thus, the difference of distance between head and feet during a movement can be obtained:

$$\Delta D = \sum_{i=2}^{n} \left(\left(X^{i}_{obj1} + X^{i}_{obj2} \right) - \left(X^{i-1}_{obj1} + X^{i-1}_{obj2} \right) \right), \tag{3.3}$$

where X^{t}_{obj1} and X^{t}_{obj2} denote 3D position of the two objects $obj1$ and $obj2$ at frame i, respectively. n denotes the number of frames inside of a window signal, and ΔD denotes the difference distance between those objects during the window signal in

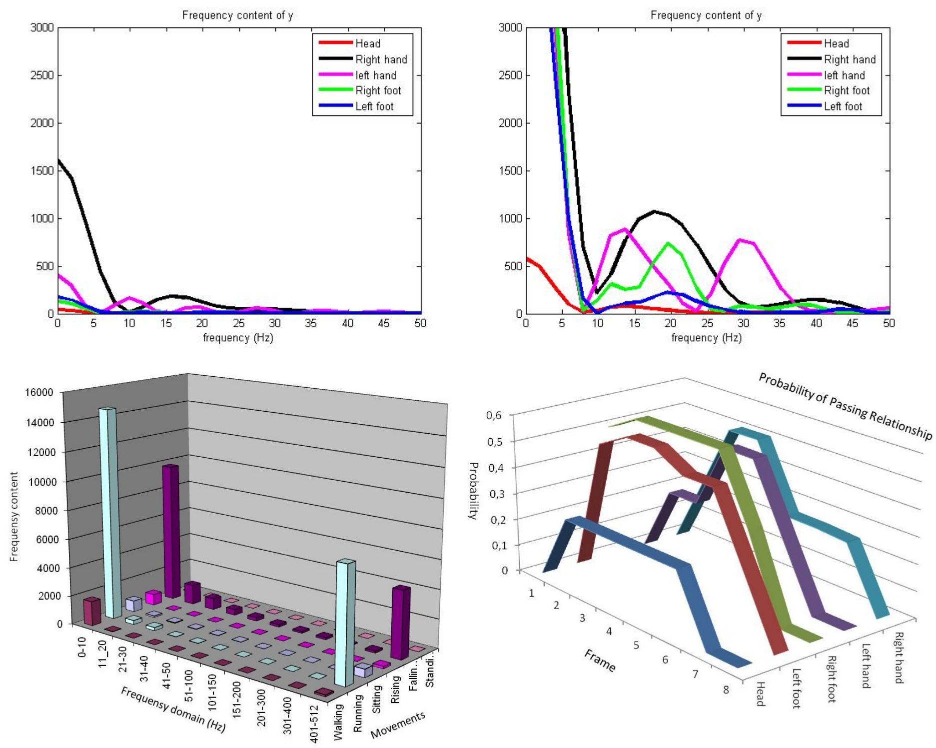


Figure 3.7: (Left-top) and (Right-top) diagram present PS results of some body parts acceleration signals for walking and running movements respectively. (Left-down) presents an histogram to show the frequency-based features of one body part for different movements in different frequency sub-domains. (Right-down) passing probabilities diagram for all body parts during a sequence of frames.

the unit of meter. The point is, when one of the objects passes through another during the window frame, the equation will calculate the difference distance if it is more in a reaching or spreading state.

3.3.2.2 Environmental based features

To implement this part, a couple of scenarios are proposed to present the idea more clearly;

- First scenario, which includes a static robot agent which people can interact with, is defined. Some relevant activities are performed like, reaching, spreading, and passing.

- Second scenario is about a couple of people interaction. In this scenario, there are not only previous activities, but also more complex ones (such as following, handshaking and pushing) which needs more input features to be estimated.

The features in this level are differences of distance between two objects during human activities. For the first scenario the difference of distance between a person and a robot using the Eq.3.3 is necessary. In the second one, we also need the distance between the hands of the people involved.

3.3.3 Relationship modeling

As it was mentioned before, the works of Norman Badlers group [BPW93, ZB05] mention five major components; Body, Effort, Space, Shape and Relationship. Each of those components describes human motion in different aspects. Several papers were published around these components, but there is no attempt to analyse human activities in the relationship component, which explains human body parts motion relations to other parts of the body, another body or environmental parameters [RDA08].

In the Relationship component of LMA there are several parameters that categorize different types of possible relations. As Hutchinson [Hut74] those mentioned parameters are called; Addressing, Nearness, Contact or Touch, Support or Carry, Enclosing or Surrounding, Toward and Away, and Facing. Each of those parameters can be in three situations; Passing, Retention and canceling of the relationship.

In the mentioned application, to analyse a person's simple movements and activities (interact with a static object (see Fig. 3.6), such as; reaching, passing and spreading, and a couple of persons activities such as; handshaking, following and pushing), some of those parameters, which are more representative, are used to present the approach. The approach allows us to generalize the system, however, for different activities, some of the Relationship's parameters are more sufficient. The performed relationship parameters are described as follows [Hut74]:

- Toward and Away: A performer may gesture toward or away from a part of his body, another person, an object, or a part of the place. For instance; Left hand of a person moves toward another's left shoulder, head moves down in a sitting movement, and approaches or moves away from your partner.

- Passing, transient relationship: Each of the relationships, addressing, nearness, touching, etc., may occur in passing, this is, the relationship may be established, momentarily sustained, and then relinquished. For instance; right hand passes near the left hand, one person passes near a robot.

- Retention of a relationship: when a relationship retains for more than one moment, it usually depends on what comes next, and when no obvious cancellation occurs, it is expected to remain. For instance; keeping the hands near each other, the box is to be kept in the hands, the object is kept near the person.

- Contact, Touch: When a body part is active in producing a touch or contact with another part, an object, or another person. For Instance; Hands touching opposite elbows, Hand-shaking between a couple of people.

3.3.3.1 Relationship component modelling

Several properties for any human activities can be defined in Relationship concept. As it was mentioned before and to simplify the system, some of those properties are modeled for the performed activities, as following subsections.

3.3.3.1.1 passing and retention relationship: These properties can be used between every two objects. In this model, the objects are the body parts. During each human activity, there are different motion signals between body parts. For instance, in a walking type movement, there is the same motion signal between opposite sides of hands and feet in the same moment. To analyse those signals, characteristics of passing and retention relationship are used between each body parts related to a reference point like body centre (if a body part is in the passing or retention relationship with the body centre). If the frequency-based quantities, which are extracted for each body parts movements, are more than a specific threshold, it means that there is a passing status related to the body centre, thus, it states a passing relationship, otherwise it states a retention one. Based on the training dataset, some thresholds are defined to discretize the frequency-based quantities [KAQ+10]:

$$f_{pb}^{i} = \begin{cases} No & Max\left\{f_{pb}^{i}\right\} \leqq 20 \\ Low & 20 < Max\left\{f_{pb}^{i}\right\} \leqq 150 \\ Medium & 150 < Max\left\{f_{pb}^{i}\right\} \leqq 1000 \\ High & Max\left\{f_{pb}^{i}\right\} > 1000 \end{cases}$$

where $Max\left\{f^{i}_{pb}\right\}$ denotes a frequency-based coefficient for a body part (pb) (Head, hands and feet), in i^{th} frequency sub-domain where

$$i = \begin{cases} 1 & 1 \leqq f \leqq 10HZ \\ 2 & 10 < f \leqq 20HZ \\ 3 & 20 < f \leqq 30HZ \\ 4 & 30 < f \leqq 40HZ \end{cases}.$$

Hence, for each body part, one variable with a couple of states is defined as following:

$$Pass_{pb} \in \{passing, retention\} \tag{3.4}$$

Fig. 3.7-(right-down) presents an example of the output of the $Pass_{pb}$ model for all body parts during a sequence of movements. That diagram shows the passing state probabilities of human body parts during a specific movement.

3.3.3.1.2 Toward and Away relationship: They can be defined in two different spaces; between body parts, and between a person and a robot or another person. By having the Eq.3.3 for the two objects , Toward qualities can be estimated by discretizing of the ΔD using a couple of thresholds, which are obtained by observing several experimental data set:

$$Toward = \begin{cases} N & \Delta D > 0.3\,m \\ S & 0.3 > \Delta D >= -0.3\,m \\ P & \Delta D < -0.3\,m \end{cases}, \tag{3.5}$$

where P, S and N denote Positive, Still and Negative qualities of Toward's property, respectively. These properties are used between head and feet as two body parts of one person (ΔD_b), and a person and a robot, or between a couple of people as two objects in a scene (ΔD_e). Thus, there are two types of Toward's variables with the three states. We applied different thresholds on the model (0.1 to 1.0 m), and the mentioned one provided the best result based on our collected data.

3.3.3.1.3 Contact or Touch relationship: Several activities will appear by contacting two objects, for instance; handshaking, grasping a glass, pushing, kicking a ball, etc. For modeling these variables, handshaking and pushing actions were selected. Contact qualities are estimated by using the Eq.3.3 for the two interested objects (e.g. two hands for handshaking), and discretizing of the equation output, using one threshold, which is obtained by observing the several experimental data sets:

$$Contact_{pb1-pb2} = \begin{cases} Connected & \Delta D_c \leqq 0.1\,m \\ Disconnected & \Delta D_c > 0.1\,m \end{cases} \tag{3.6}$$

where $pb1$ and $pb2$ denote hands of the two persons, and ΔD_C denotes the difference distance in contact space. There are two possible states; Connected and Disconnected, that discretized by a threshold 0.1 meter difference distance between two objects.

3.3.3.2 Human movement and activity model

By combining the Toward variable, which is defined between two body parts, and the pass variable, the human movements (see Set.3.7) are analysed.

$$Movement \in \{walking, running, sitting, rising, falling, standing\} \tag{3.7}$$

By using the other variables between person, robot or another person, and the Movement variables, the performed human activities are estimated. The activity states are:

$$Activity \in \{reaching, spreading, passing, handshaking, following, pushing, other\} \tag{3.8}$$

Toward's property between human and robot or another person, is the proper feature to analyse reaching and spreading activities. The passing activity can happen when the Toward variable states Still's state while a person is walking, but when both persons are walking, then the following activity is happening. For the handshaking action both persons are in a standing state and the Contact variable is in connected state, In pushing action, however, there is connected state also, but one of the persons will have a fall down movement in the end. There are more possibilities to define more activities by having those movements and the environmental parameter states. For modeling the idea, it was not supposed to implement a complex model, but to present an approach which can be easily generalized. Fig. 3.8-a) presents the idea in a scene.

There are three objects (two people and one robot) in the scene. As can be seen in the figure, depending on the situation of each of them and between each couple of objects, one specific activity can be estimated.

3.3.4 Modeling based on Bayesian framework

Bayesian approach is a popular and well known method to classify human motions and activities [RA08, RDA08]. A Bayesian Network (BN) is a suitable method to deal with variable dependencies and uncertain data, learning with a small bunch of data, and fusing different types of features. HMM (as a DBN) is another popular approach, which was used also for this kind of applications. HMM works when the defined scenarios are based on a sequence of states. In this study, both approaches are implemented for analysing human activities.

3.3.4.1 Bayesian network

In the model (see Fig. 3.8-b)), as it was mentioned before, there are both frequency and spatial based features in parallel for different properties. Frequency-based features for each body part (*pb*) are used for the Passing/retention relationship of the same body part:

$$P\left(Pass_{pb} \mid \prod_{i=1:4} Max\left\{f^{i}_{pb}\right\}\right) = \frac{P\left(Pass_{pb}\right)\prod_{i=1:4} P\left(Max\left\{f^{i}_{pb}\right\} | Pass_{pb}\right)}{\prod_{i=1:4} P\left(Max\left\{f^{i}_{pb}\right\}\right)} \tag{3.9}$$

The probability of Toward/Away relationship between body parts can be obtained by:

$$P\left(Toward_b \mid \Delta D_b\right) = \frac{P\left(Toward_b\right) P\left(\Delta D_b \mid Toward_b\right)}{P\left(\Delta D_b\right)} \tag{3.10}$$

and the probability of similar property but between a person and an external object (such as a robot or another person) can be achieved by:

$$P\left(Toward_e \mid \Delta D_e\right) = \frac{P\left(Toward_e\right) P\left(\Delta D_e \mid Toward_e\right)}{P\left(\Delta D_e\right)} \tag{3.11}$$

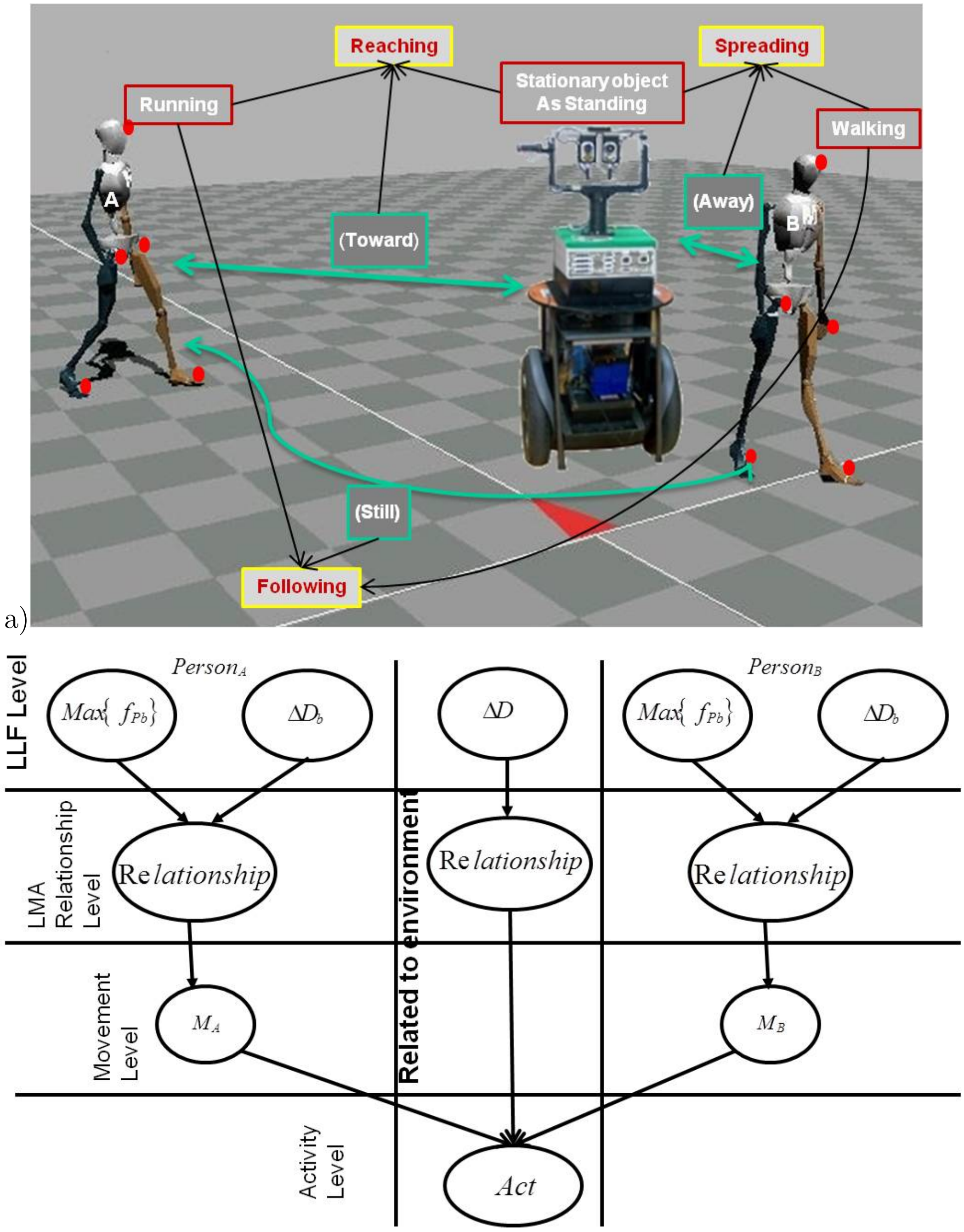


Figure 3.8: a) A scenario of mentioned system. b) The global Bayesian model for all the process of analysis in four layers, and two different spaces (related or not related to the environmental parameters)

and for the Contact property:

$$P(Contact \mid \Delta D_c) = \frac{P(Contact)\,P(\Delta D_c \mid Contact)}{P(\Delta D_c)} \tag{3.12}$$

There are two other levels of analysis; one of them is movement's level which is not related to the environment parameters. The other one is activity level which can be analysed by finding the connection between human movement and the scene information. The free-context based movement model is defined as:

$$P\left(M \mid Toward_b \prod_{pb=1:n} Pass_{pb}\right) = \frac{P(M)\,P(Toward_b|M) \prod_{pb=1:n} P\left(Pass_{pb}|M\right)}{P(Toward_b) \prod_{pb=1:n} P\left(Pass_{pb}\right)} \tag{3.13}$$

In the activity level, there are not only each human movement probabilities, but also the environmental parameters:

$$P(Act \mid Toward_e\, Contact\, M_A M_B) = \frac{P(Act)\,P(Toward_e|Act)\,P(Contact|Act)\,P(M_A|Act)P(M_B|Act)}{P(Toward_e)P(Contact)P(M_A)P(M_B)} \tag{3.14}$$

where M_A, M_B and Act denote the movement state of person A and B, and activity, respectively, and n denotes the number of body parts.

Fig. 3.8-b) presents the dependencies of all those different variables (LLFs, Relationship's component, movement, activity) in two different spaces (related and not related to the environmental parameters) in one model. Fig. 3.9 presents the BP's correspondent of the proposed human action and interaction model.

Fig. 3.10 presents two examples of the model results in different steps. First scenario is about pushing activity. Fig. 3.10-a) and Fig. 3.10-b) present both persons' movements classifier results, and Fig. 3.10-c) shows their trajectories in the scene. Fig. 3.10-d) presents the results of the activity classifier. Another scenario shows handshaking activity. Fig. 3.10-e) presents a sequence of three images of the related activity. Fig. 3.10-f) shows the model of the same scenario that was obtained by the motion tracker suit, and Fig. 3.10-g),Fig. 3.10-h),Fig. 3.10-i) and Fig. 3.10-j) present results movement classifier of person A and person B and the trajectories of

Program — Description — Specification

Pertinent variables :

$Pb \in \{Head, Left\,hand, Right\,hand, Left\,foot, Right\,foot\}$

$NP \in \{personA, personB\}, D_b^{j\in\{NP\}}$ and $D_e \in \{Neg., Still, Pos.\}, D_c \in \{Low, High\}$

$f_{i=1:4}^{p\in\{pb\}} \in \{No, low, Medium, High\}, Pass_{P\in\{Pb\}}^{j\in\{NP\}} \in \{passing, retention\}$

$Toward_b^{j\in\{NP\}}$ and $Toward_e \in \{Negative, Still, Possitive\}, Contact \in \{Connected, Disconnected\}$

$M_{\in\{NP\}} \in \{Walk, Run, Sit, Rise, Fall, Stand\}$

$Act \in \{Reach, Spread, pass, handshake, follow, push, other\}$

Decomposition :

$$P\left(Act\,Toward_e\,Contact \prod^{j\in\{NP\}}\left(Toward_b^j M_j\right)\prod_{P=\{Pb\}}^{j\in\{NP\}}\left(Pass_P^j\right)\prod^{j\in\{NP\}}\left(D_b^j\right)D_e\,D_c\prod_{P\in\{Pb\}}^{i=1:4,j\in\{NP\}}\left(f_{i,j}^P\right)\right) =$$

$$\prod^{j\in\{NP\}} P\left(Toward_b^j \prod_{P=\{Pb\}}\left(Pass_P^j\right)|M_j\right)$$

$$P(Toward_e|Act)\,P(Contact|Act)\,P(Contact|M_A)\,P(Contact|M_B)$$

$$P(D_e|Toward_e)\,P(D_c|Contact)\prod^{j\in\{NP\}} P\left(D_b^j\middle|Toward_b^j\right)\prod_{P\in\{Pb\}}^{j\in\{NP\}} P\left(\prod^{i=1:4}\left(f_{i,}^p\right)\middle|Pass_P^j\right)$$

$$\prod_{P\in\{Pb\}}^{i=1:4,j\in\{NP\}}\left(P\left(f_{i,j}^p\right)\right)\prod^{j\in\{NP\}}\left(P\left(D_b^j\right)\right)P(D_e)\,P(D_c)$$

Parametric forms :

$P\left(f_{i,j}^p\right)_{P\in\{Pb\}}^{i=1:4,j\in\{NP\}}, P\left(D_b^j\right)^{j\in\{NP\}}, P(D_e), P(D_c) : uniform$

$P(D_e|Toward_e), P(D_c|Contact), P\left(\prod^{i=1:4}\left(f_{i,j}^p\right)\middle|Pass_P^j\right)_{P\in\{Pb\}}^{j\in\{NP\}} : histogram(Matrix)$

$P(Toward_e|Act), P(Contact|Act), P(Contact|M_A), P(Contact|M_B) : histogram(Matrix)$

$P\left(D_b^j\middle|Toward_b^j\right)^{j\in\{NP\}}, P\left(Toward_b^j\prod_{P=\{Pb\}}\left(Pass_P^j\right)|M_j\right)^{j\in\{NP\}} : histogram(Matrix)$

Iden.

Ques. : $P\left(Act\middle|\prod_{P\in\{Pb\}}^{i=1:4,j\in\{NP\}}\left(f_{i,j}^p\right)\prod^{j\in\{NP\}}\left(D_b^j\right)D_e\,D_c\right)$

Figure 3.9: Bayesian Program for the human interaction model.

both persons in the scene, and finally, activity results, respectively.

3.3.4.2 Hidden Markov model

HMM as a DBN is a common approach to estimate human activities based on sequence states. In this work concurrent HMM was implemented, which is described more detail in [KAM+11]. A concurrent HMM is composed of several HMMs (see Fig. 3.11-a), and each one describing one class. The inputs of the model are all probabilities of both persons movement classes, and the probabilities of the relationship parameters, which are obtained by the performed BN in the Movement and Relationship layers. In every sequence of data, the activity class, which has most probability of its corresponding HMM, is the output of the model (see Fig. 3.11-b). This is performed by finding the HMM Act that maximizes $P(Act|O_{t-n}...O_t)$ where the O_t denotes observation data at time t .

3.3.5 Experiments

To obtain the input data (3D position and acceleration of body parts such as hands, feet, head, etc.) The body motion tracker (MVN suit) [KAQ+11] is used (Chapter 2). The suit uses a global reference in the scene as a starting point which is known. Thus, the distance between human body and other such as another person or a robot, during the activities, can be calculated.

Several different human movements and activities are performed. One second window which shifts half a second, during each record is defined to feature extraction process of all performed movements. Two series of data are achieved; 3D human body parts positions related to the body centre, and the person's 3D position related to a global reference which the person is supposed to interact with, in the scene. A free Bayesian toolbox provided by Kervin Murphy and Berkeley [MB05] is used to implement the BN model.

The BN estimates the high probability of the person movement and activity states in each second. Table.3.3 presents the classification results in movement level, as inputs of activity level of the model. This data is obtained from around 100 sequences of human movements with different duration (first half of the data is used in learning and others in classification process and then vice versa (the first half of data is used for the

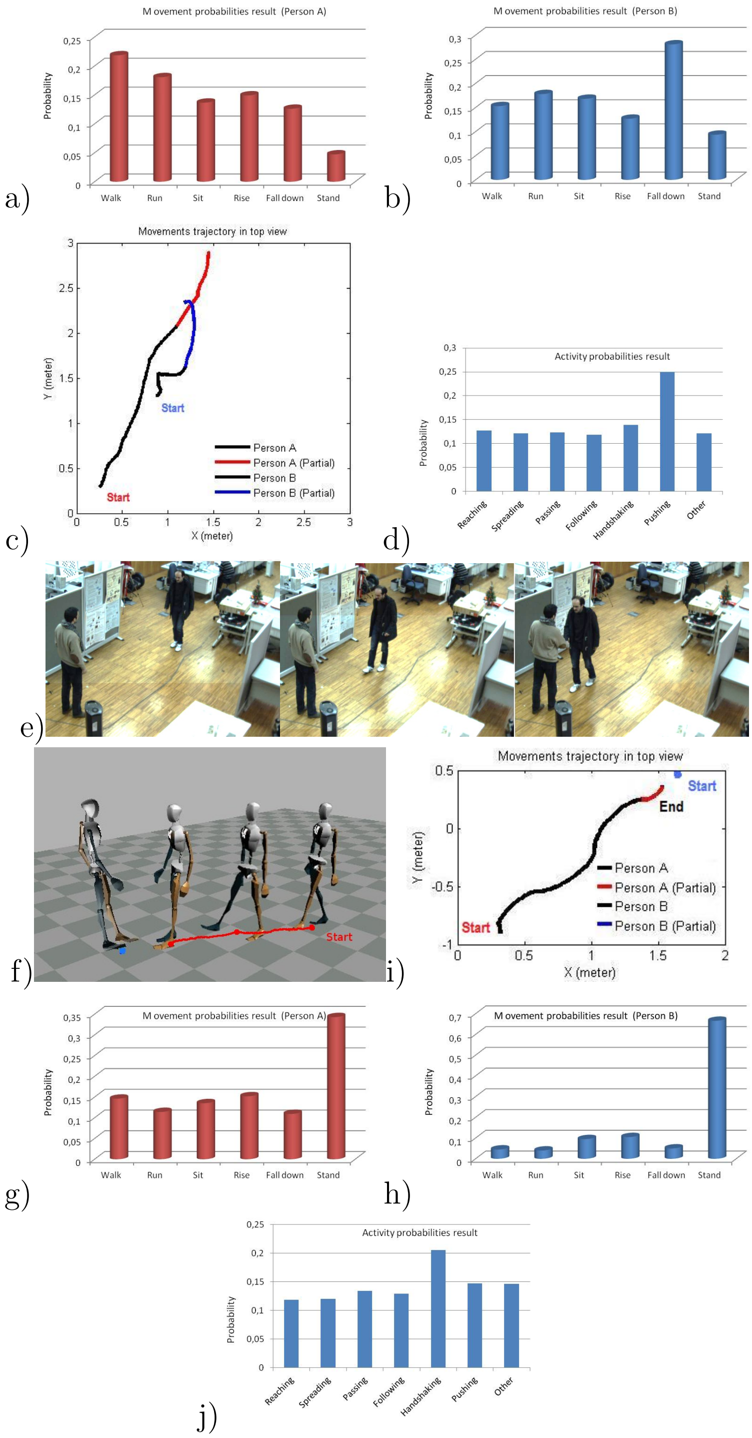


Figure 3.10: The diagrams present two samples; first is a pushing activity scenario, the persons trajectories were shown in (c), and movements classification results in (a and b) and activity in (d). Second scenario presents handshaking activity. e) shows a sequence of three images to present a normal handshaking activity. f) presents an example handshaking scenario which was obtained by the motion tracker suit. g) and h) shows the movement classifier results of both persons. i) shows the trajectories and j) shows the related activity classifier results.

a)

b)

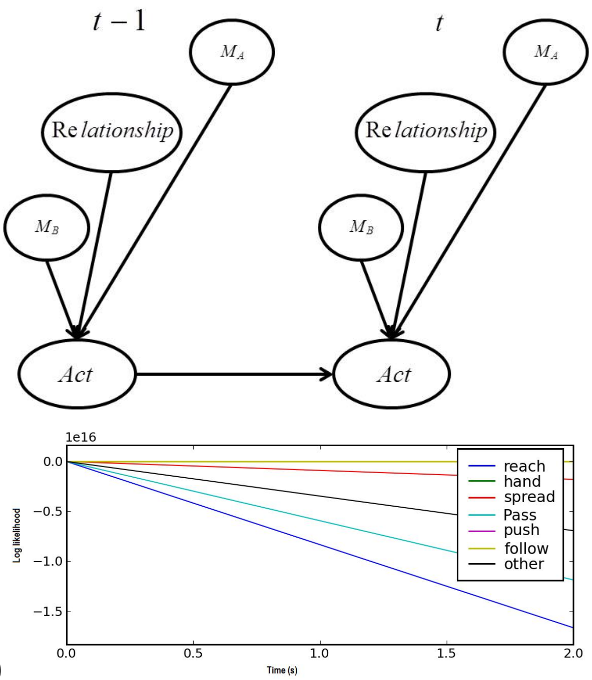


Figure 3.11: a) Structure of proposed HMM to classify human activity (Act) by having several inputs; movements type of person A (M_A) and person B (M_B), and Relationship parameters related to environment ($Relationship$) as the $Toward_e$and the $Contact$ components, at time t and $t-1$. For each class of Act we have one HMM of that. b) A diagram shows the log-likelihood of each class during the sequence of data. The log-likelihood data is the output of the model. Thus, the class which has more log-likelihood than the others will be the activity model estimation result.

classification process and the others for the learning).

To be able to have a comparison with the state-of-the-art, we attempt to find the related works in 3D based analysis which is the main characteristic of the work, however the experimental setup is not the same, and it's not a quite fair comparison. In [HMNP11], 3D data-based classification results for different numbers of human movements can be seen. The overall result for classification of six human movements was maximum 89.58%. In this aspect our model provided better performance with 96.45% accuracy. in [TVC08] very accurate classification results (overall 98.78%) were estimated, but their performed movements are quite different in spatial aspect. For instance, there is no running movement in their movement classes which can be easily confused with walking movement for this kind of models. Probably that is the reason that their walking movement class always estimated 100%. As can be seen in [HMNP11], their less accurate results were around between walking and running movements.

	Walk	Run	Sit	Rise	Fall	Stand	%
Walk	63					2	96.92
Run	1	72			2		96.00
Sit			46		2		95.83
Rise				34		1	97.14
Fall		1		1	26		92.85
Stand						155	100

Table 3.3: Human movement classification result

To make the multi-layer classifier results more visible, Fig. 3.12 and Fig. 3.13 show probability diagrams of both classifiers of movement and activity in a sequence of steps. In the first sample of Fig. 3.12 (first column), one person is in walking and another in running movement state. These two persons will meet in the end and shake hands, but in the 6th frame we see the pushing result activity, because of the person who was in the running state in the previous step, that usually happens more in pushing than handshaking activity in a normal scenario. In the second column, one person runs and pushes the other. In Fig. 3.13 and in the first column presents a scenario which shows a person who walks and passes near another person who is in standing state. Finally, the second column shows two persons in the running state, but one of them is faster than the other, thus, in the activity results show the person is going away from the other one.

Table.3.4 presents the result of the model for human activities level using the BN model. The result shows that our model for handshaking activity is not as accurate as others. The reason is that the related features are not sufficient enough to distinguish between handshaking and pushing activities.

To implement the HMM approach, an interface which was described in [KAM$^+$11], is used. The input data in this case consists of several sequences of the observations which consist of both persons movement states and relationship parameters probabilities. For each class, several sequences data for leaning and for classification process, are collected.

Table.3.5 presents the result of the model for human activity level using the HMM in the last layer. The result shows that the model is more reliable than the previous one, with less false detections. It can be improved by using other relationship parameters [Hut74]. The advantage of the HMM approach is using a sequence of observations, which is relevant to the activity definition [Bob97]. The results also proved the Bobick's

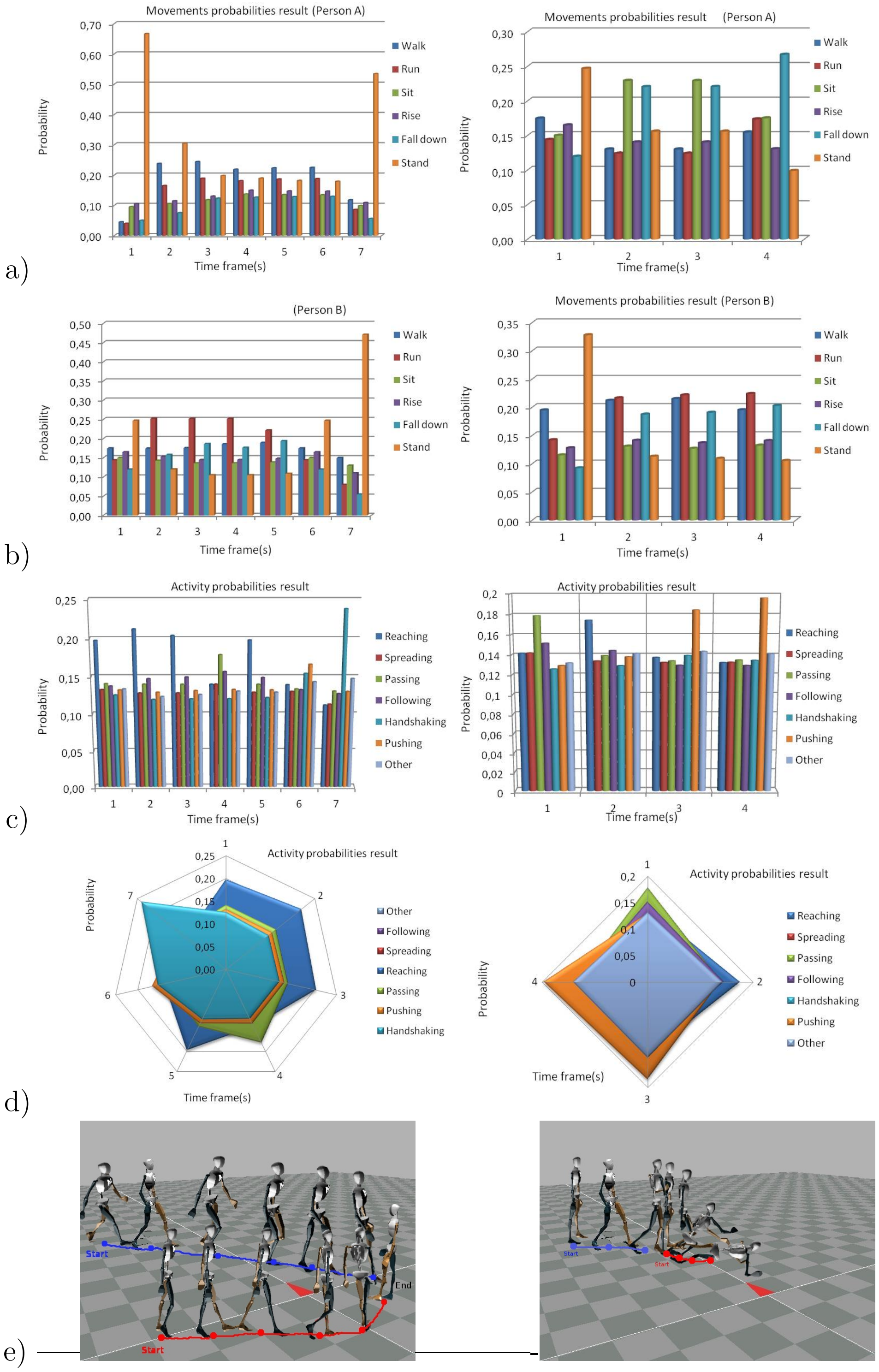


Figure 3.12: Two sequences steps of changing the human activity states from two human movements in a scene. a) and b) present the both human movements classifier results, c) and d) show the activity classifier results in different schema, and e) presents scenarios (Handshaking and Pushing) in different steps.

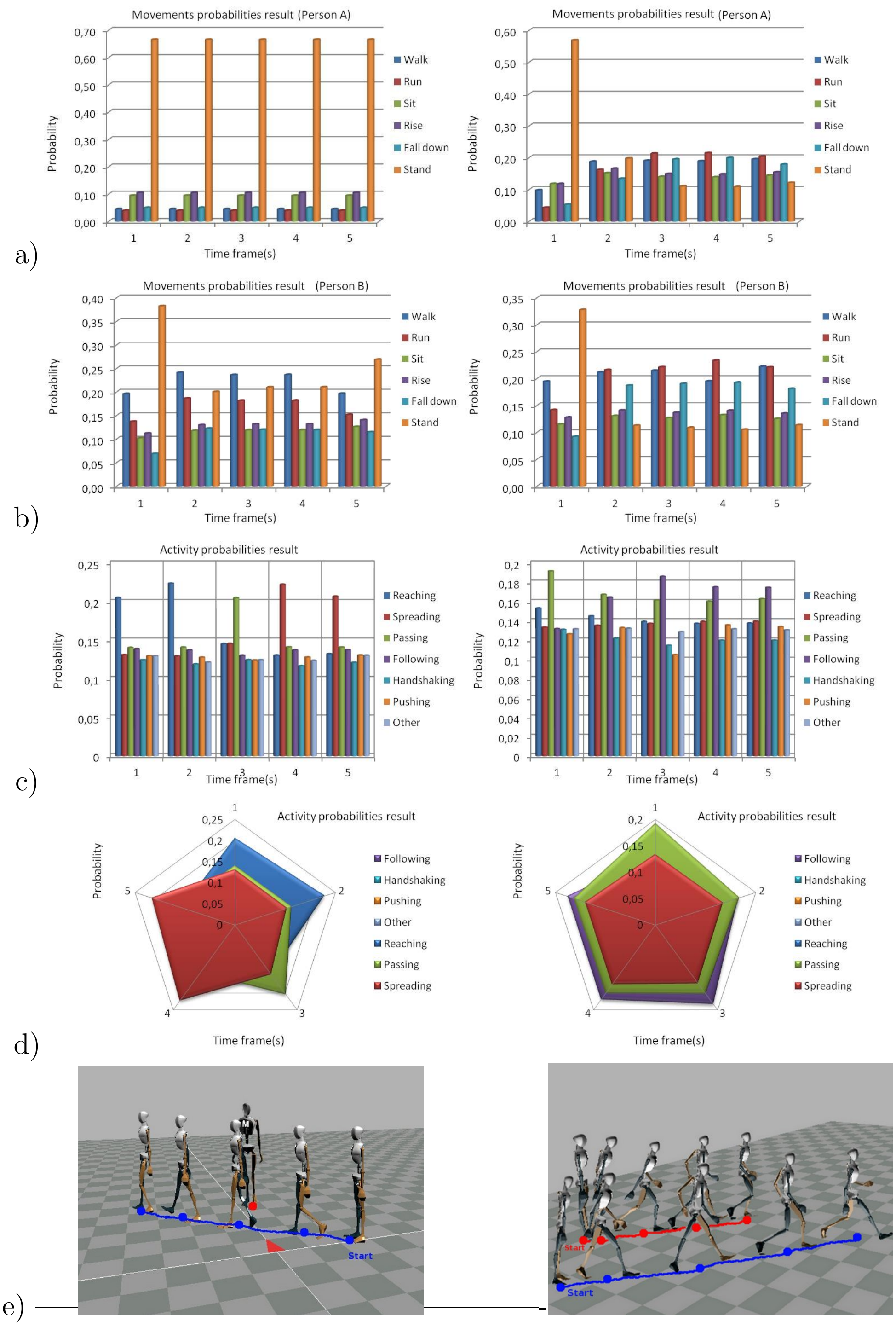


Figure 3.13: Two sequences steps of changing the human activity states from two human movements in a scene. a) and b) present the both human movements classifier results, c) and d) show the activity classifier results in different schema, and e) presents scenarios (Passing and following) in different steps.

	Rch	Spd	Pas	Flw	Hsk	Psh	Oth	%
Rch	133	0	2	0	0	0	7	93.66
Spd	0	141	6	0	0	0	5	92.76
Pas	8	1	127	9	0	0	3	85.81
Flw	1	1	6	50	0	0	3	81.96
Hsk	3	0	0	0	53	11	1	77.94
Psh	0	2	0	0	3	68	3	89.47
Oth	1	1	4	1	1	2	211	95.48

Table 3.4: Human activity classification result using the BN which can be seen in Fig. 3.8-b)

	Rch	Spd	Pas	Flw	Hsk	Psh	Oth	%
Rch	45	0	1	0	1	1	0	93.75
Spd	0	45	1	0	0	2	0	93.75
Pas	3	0	43	0	1	1	0	89.58
Flw	0	0	0	45	0	1	0	97.83
Hsk	0	1	0	0	42	5	0	87.50
Psh	0	0	1	1	2	44	0	91.67
Oth	0	0	0	0	1	1	55	94.49

Table 3.5: Human activity classification result using the HMM as can be seen in Fig. 3.11

terminology.

A general discussion and comparison between the existing methods for human activities understanding was presented in [Pop10]. In this work, we presented the two most popular methods (BN and HMM) in the related area, to show their results in the mentioned framework also. Each has its own advantages, for instance; however HMM needs previous state knowledge, but it gives better results than the BN approach. We believe that HMM shows better performance for complex activities.

By having the obtained information by the models, it is possible to analyse more complex human activities, like rubbery (when the rubber performs first reaching and then spreading activity in running movement state's), fighting (reaching in running or falling down movement's states), etc. It means, these relationship parameters can assist us to analyse even more complex human activities in different concepts and applications.

3.4 Improved semantic-based scene understanding

Human activity analysis can be categorized as context-free based and context based. In context-free based approaches the model is independent of scene parameters, and just relies on the features belonging to the person. However, in reality, context-based features play a very important role in analysing human activities. For instance, when a person is going to reach a chair, we will realize that the person is going to sit on the chair, not to sleep in it.

As Delaitre et al, described in [DFL+12], since object detection is a widely studied topic in computer vision, analysing the relation between human movements and the existent objects around, can produce valuable information for human daily activities. For instance, people have learned the (most probable) normal activities when the person is reaching to a chair, thus, people have a probabilities set of activities depending on the objects in the scene.

The question which level of human movement information might be useful, and then how can a general framework be defined for analysing any possible human-object interactions. For the mentioned aspect, the lowest level information such as body parts motions and the higher ones such as human interactions can be useful. A complex model is needed to deal with the mentioned different information. Thus, a hierarchical framework is used to reduce the complexity of the model [AR11] and to provide multi-level of human activity analysis [RD13b].

Probability distributions of the relationship between human motions and human-object information, can be obtained by giving the possible activities and the interested objects in the scene. LMA system, which consists of several components, is used to define proper human motions (Effort, Shape) [KAQ+11, KAQ+10] and human-scene relations (Relationship) [RD13a, Hut74] variables. Gupta et al. in [GKD09] tackled the problem, based on the 2D images. Therefore, they were focused more on the computer vision problems for the mentioned applications, and used human hand trajectory information to analyse human-object interactions (reaching and manipulation). Their mentioned Bayesian model cannot deal easily with the extension of the work for other activities. Thus, we propose a hierarchical model to deal with the problem. To avoid the limitation of the 2D-based analysis, the motion tracker suit (MVN) with several inertial sensor attached on the different body parts, is used. The suit provides 3D pose of human body parts with maximum 120 frames per second resolution. However,

<table>
<tr><th>Category</th><th>LMA Component</th><th>Variable</th><th>States</th><th>Feature's Domain</th></tr>
<tr><td rowspan="5">Body parts motion</td><td rowspan="2">Effort</td><td rowspan="2">Time</td><td>Sudden /</td><td rowspan="2">Frequency</td></tr>
<tr><td>Sustained</td></tr>
<tr><td rowspan="3">Shape</td><td rowspan="3">Vertical</td><td>Rising/</td><td rowspan="3">Spatial</td></tr>
<tr><td>Still/</td></tr>
<tr><td>Sinking</td></tr>
<tr><td rowspan="7">H-O relationship</td><td rowspan="7">Relationship</td><td rowspan="3">Toward (H-O1)</td><td>Toward/</td><td rowspan="7">Spatial</td></tr>
<tr><td>Still/</td></tr>
<tr><td>Away</td></tr>
<tr><td rowspan="3">Toward (H-O2)</td><td>Toward/</td></tr>
<tr><td>Still/</td></tr>
<tr><td>Away</td></tr>
<tr><td>Contact (H-O1)</td><td>Connected/ Disconnected</td></tr>
</table>

Table 3.6: Different LMA components in a couple of categories which are defined for human interaction analysis. H-O1 denotes Human-Object1. In the experiment, O2 is used as another person, and O1 as a chair.

there are several works using 3D-based human movement analysis with high accuracy [MS10, CGV+11], and also in 3D virtual applications [ENVRZ09], but only focused on simple human movements classification.

3.4.1 Hierarchy-based human activity analysis

Table.3.6 presents the defined LMA parameters based on the three components (Effort, Shape and Relationship) for this study.

Based on the mentioned features, various human-object interactions can be analysed. For this study, depending on the interested objects, a set of interactions is defined. Sitting and standing up actions for human-chair interaction, pushing and handshaking actions for human-human interaction purpose, and reaching and spreading in general are defined. Table.3.7 shows the defined classes for both movement and interaction variables.

3.4.2 Bayesian network modeling

As it was explained before, BN is a well-known approach to model an hierarchical-based analysis [RD13a], because of its flexibility and capability of fusing different types of

Variable	Classes
Movement	Walking, Running, Falling down, Sitting, Rising, Standing
Interaction	Reaching, Spreading, Sitting on the chair, Standing up, Handshaking, Pushing, Other

Table 3.7: Human movement and interaction classes. Other means any activities which are not belong to the defined classes.

features, and deal with uncertainty, decision making problem and prediction process. The Bayesian graphical model for the mentioned system can be seen in Fig. 3.14, which presents the dependencies between the different levels of information.

In each level of the BN, the probability of defined variables are modeled by Bayesian rule formulation. In the highest level, we intend to estimate the probability of each human interaction state, given the movement states probabilities of both person a and person $o1$, and the relation between person a and the two defined objects ($o1$ which is other person and $o2$ which is a chair). Thus the Eq. 3.15 presents the mentioned Bayesian rule, which can see the dependencies.

$$P\left(I_a|M_a,M_{o1},R_{a-o1},R_{a-o2}\right) = \frac{P(I_a)P(M_a|I_a)P(M_{o1}|I_a)P(R_{a-o1}|I_a)P(R_{a-o2}|I_a)}{P(M_a)P(M_{o1})P(R_{a-o1})P(R_{a-o2})} \tag{3.15}$$

where, I_a,M_a and R_{a-o1} denote person a's Interaction, Movement and Relationship with respect to object $o1$, variables respectively.$P\left(M_a|I_a\right)$ denotes the estimation of Movement's states of person a probabilities given probability of its I_a states.

In the Eq. 3.15 the variables which are located in lower levels, need to be solved. For instance; Equation 3.16 which proposed in [KAQ+11], was used to model human movement.

$$P\left(M_a|E_a^{bp}S_a^v\right) = \frac{P\left(M_a\right)P\left(E_a^{bp}|M_a\right)P\left(S_a^v|M_a\right)}{P\left(E_a^{bp}\right)P\left(S_a^v\right)} \tag{3.16}$$

where $E_a^{bp}$and S_a^v denote Effort component of LMA for bp's body part of person a, and Shape component of LMA for person a in vertical plane, respectively.

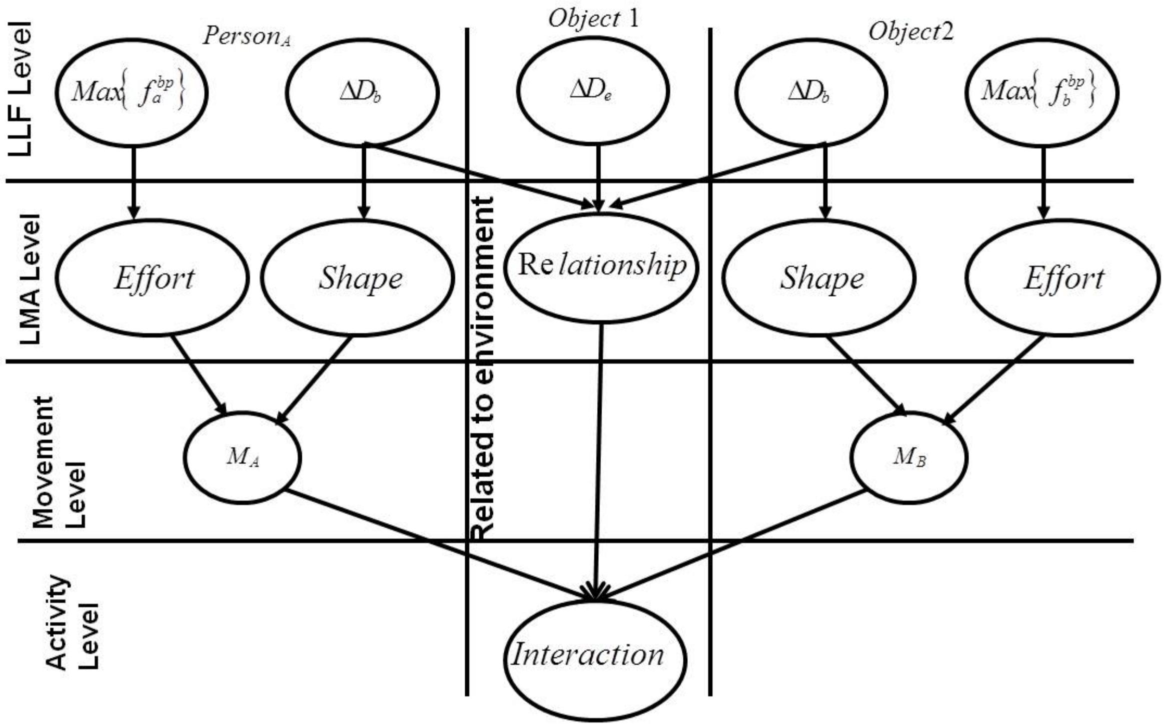


Figure 3.14: A hierarchical framework for scene understanding. LLF's level contains frequency-based features ($Max\left\{f^{bp}\right\}$) and spatial-based ones (ΔD_b for body parts motion's category and ΔD_e for human-object relationship). LMA's Level contains Effort and Shape components of the people and Relationship component. In the Movement level, we have M_a and M_b which denote movement classes belong to person a and b given their related Effort and Shape components, respectively. Finally Interaction variable are estimated given both person a and b movement classes and the existent Relationship states.

bp is the index of body parts which are used (hands, feet and head). $P\left(E_a^{bp}\right)$, $P\left(S_a^v\right)$and $P\left(R_{a-o1}\right)$ denote probability of Effort, Shape and Relationship components of LMA, respectively. The mentioned LMA components probabilities are estimated by [KAQ$^+$11, RD13a], given frequency and spatial based features.

3.4.3 Learning process

To obtain conditional probability of each variable in different levels of analysis, learning process is needed. Maximum likelihood is a well-known approach in the learning process. For each class of movement, several data by the motion tracker suit was collected. The interested features were extracted for different body parts (feet, hands and head). To obtain the conditional probability for a variable in each level, histogram-based approach is used. Fig. 3.15, presents a sample of the obtained histogram of the frequency

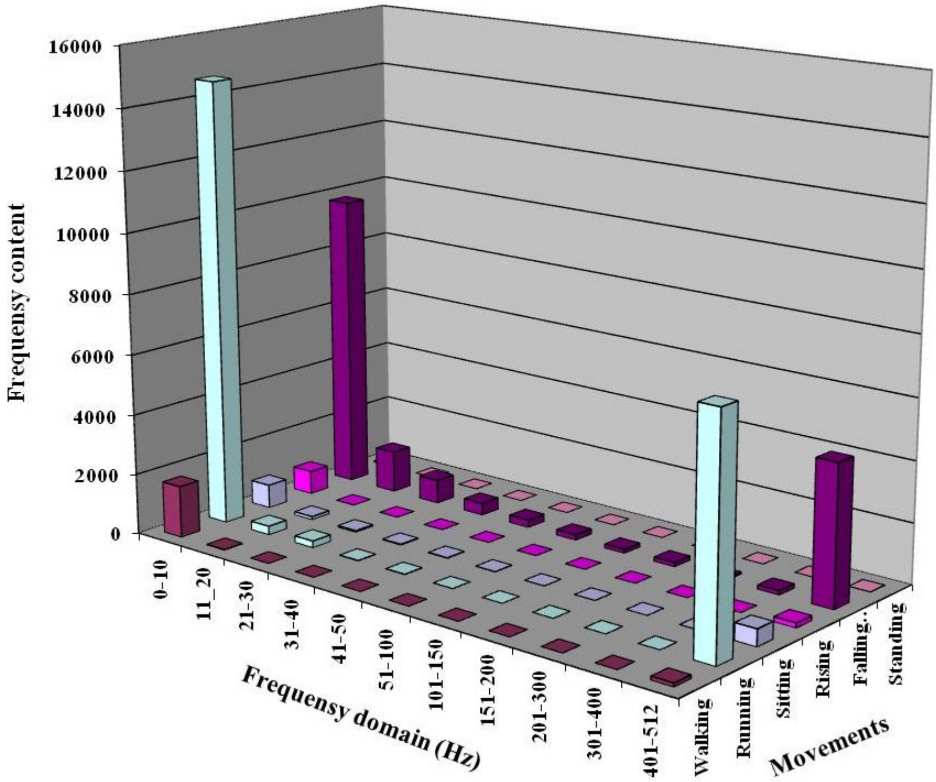


Figure 3.15: Histogram of the frequency-based features of a specific body part for different movements in different frequency sub-domains [RD13a].

content in different sub-domains for different types of human movements [RD13a].

Then, after applying the learning process in the all levels, we obtain the $P\left(\Delta D_e \mid R_{o1-o2}\right)$ to estimate Eq. 3.11, $P\left(E_a^{bp}|M_a\right)$ and $P(S_a^v|M_a)$ to estimate Eq. 3.16, and $P(M_a|I_a)$, $P(M_{o1}|I_a)$, $P\left(R_{a-o1}|I_a\right)$ and $P\left(R_{a-o2}|I_a\right)$ to estimate Eq. 3.15.

3.4.4 Experiments

The motion tracker gives body parts positions depending on a global reference which is obtained in the sensors calibration step. Based on that global reference the other interesting objects positions in the scene, are estimated (see Fig. 3.16). For the feature extraction step the sliding window approach is performed. The length of window has been defined in duration of one second, that the window shifts by half of the window's length. The frame rate of the system is 120 Hz. Ten different sequences (each sequence contains more than 1000 frames) are collected for each type of human movements, which performed inside of the different defined actions and interactions.

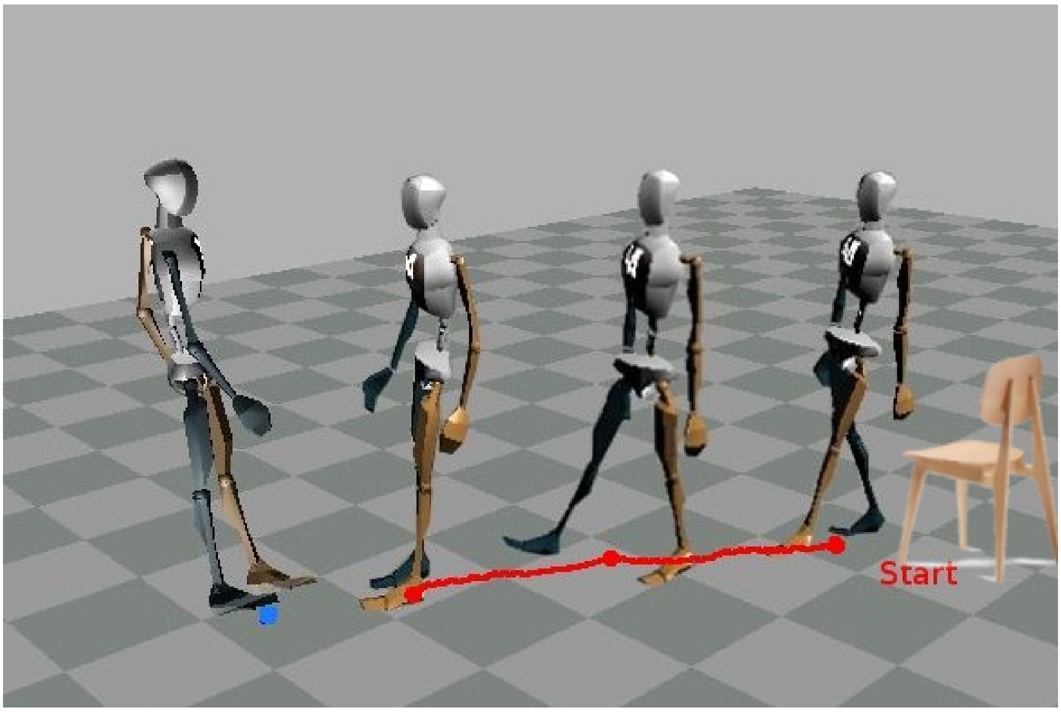


Figure 3.16: The scene with a sample of people activities, provided by the motion tracker suit (MVN®)

Table 3.8 presents the classification results of two window-slide sequences. As can be seen, different levels of information were presented semantically and probabilistically.

The overall results can be seen on the Table 3.9. The obtained results prove that the context-based knowledge improves the accuracy of the model (from 92.22% to 96.80%) by reducing the false detections which were presented on the previous section [RD13a]. As can be seen, between human-chair interactions and human-human ones there is no false detection, however, there are still some false detections between those similar context-based activities and especially between most of the classes with the Other's class. Most of those false detections happen in the boundary of between two classes, because of the use of sliding window-based segmentation approach. When a slide window occurs in the boundary, the new class of activity will be considered in the ground truth, though the window slide which had more signal belongs to the previous one.

Frame	Level	States with their probability
180		
	LMA.Effort.Time-Head	Sudden:45%, Sustained:55%
	LMA.Effort.Time-LFoot	Sudden:63%, Sustained:37%
	LMA.Effort.Time-RFoot	Sudden:61%, Sustained:39%
	LMA.Effort.Time-RHand	Sudden:55%, Sustained:45%
	LMA.Effort.Time-LHand	Sudden:55%, Sustained:45%
	LMA.Shape.Vertical	Sinking:15%,Still:75%,Rising:10%
	LMA.Relationship.T/A H-O1	Toward:21%, Away:79%
	LMA.Relationship.T/A H-O2	Toward:68%, Away:32%
	LMA.Relationship.Contact H-O1	Connected:8%, Disconnected:92%
	Movement	Walking:45%, Standing:12%, Running:22%, Sitting:8%,Rising:7%,Falling:6%
	Action and Interaction	Reaching:48%, Spreading:6%, Sitting on the chair:14%Standing up:10%, Handshaking:4%,Pushing:6%,other:12%
240		
	LMA.Effort.Time-Head	Sudden:35%, Sustained:65%
	LMA.Effort.Time-LFoot	Sudden:29%, Sustained:71%
	LMA.Effort.Time-RFoot	Sudden:25%, Sustained:75%
	LMA.Effort.Time-RHand	Sudden:54%, Sustained:46%
	LMA.Effort.Time-LHand	Sudden:25%, Sustained:75%
	LMA.Shape.Vertical	Sinking:25%,Still:60%,Rising:15%
	LMA.Relationship.T/A H-O1	Toward:45%, Away:55%
	LMA.Relationship.T/A H-O2	Toward:55%, Away:45%
	LMA.Relationship.Contact H-O1	Connected:87%, Disconnected:13%
	Movement	Walking:16%, Standing:40%, Running:10%, Sitting:8%,Rising:16%,Falling:10%
	Action and Interaction	Reaching:8%, Spreading:10%, Sitting on the chair:10%,Standing up:6%, Handshaking:37%,Pushing:17%,other:12%

Table 3.8: A table with different levels of body-motion based information belong to a person who walks to reach another person and do handshaking (in a couple of sequences). The frame number shows the last frame number of the sliding-window (The first and second step contains the frames [60-180] and [120-240], respectively).

	Reach	Spread	Sit	Stand	Hand-shak	Push	Other
Reach	97,78%	1,22%	0,00%	0,00%	0,00%	0,00%	2,22%
Spread	0,00%	95,74%	0,00%	0,00%	0,00%	0,00%	4,26%
Sit	0,00%	0,00%	96,67%	0,00%	0,00%	0,00%	3,33%
Stand	0,00%	0,00%	2,70%	97,30%	0,00%	0,00%	0,00%
Hand-shak	0,00%	0,00%	0,00%	0,00%	97,83%	2,17%	0,00%
Push	0,00%	0,00%	0,00%	0,00%	2,27%	97,73%	0,00%
Other	1,09%	2,17%	1,09%	1,09%	0,00%	0,00%	94,57%

Table 3.9: Classification results

3.5 Conclusion

In this chapter, to deal with different levels of human movement analysis, a semantic-based hierarchical framework is proposed. LMA parameters are used to provide a standard description on body motion analysis level, which we call LMA level. There are three components in the LMA level, Effort (deal with temporal domain), Shape (deal with spatial domain) and Relationship (deal with the context-based information), which are formulated in this study. Bayesian-based approaches are used to model the multi-layer framework. This study is divided in three steps; individual human movement analysis, context-based human behaviour analysis, and semantic-based scene understanding. Output of the first step is used for the other ones output. In the second step, we attempted to formulate Relationship component to analyse different types of human actions and interactions. In the end, the semantic and context-based knowledge with respect to the Relationship component is used for modelling process to improve the accuracy of estimations. In the output of the model, a probabilistic-based descriptor for human activities is presented in the different levels. Based on the results, it is proved that the context of a scene, where the humans interact, can highly avoid the possible false detections, however, there are still a few drawbacks which happen in the boundaries of two classes of activities.

Chapter 4

Enhanced human movement segmentation

4.1 Introduction

Action recognition is an active research topic within the scientific community, with several applications, which include human-machine interfaces, intelligent video surveillance, video indexing and analysis, to name just a few. The action segmentation problem is a key issue in action recognition and may be divided in two stages: (1) Learning and (2) Classification. The learning stage often involves a data preprocessing step to find alternative, discriminant representations for different properties of the input signal. In this work, we consider a data driven probabilistic representation for the action model, which is learned from a set of training data. This action model is posteriorly used to identify to which action class each observable feature belongs.

A popular applied method in this problem is the sliding window approach. The window is used to progress sequentially through the input signal, creating data segments from which features are extracted. This method is popular because of its direct integration with the majority of classification algorithms. However, fixed parameter values are a significant cause of classifier under-performance: slow convergence and/or borderline decisions. Choosing the ideal parameter values is not a trivial task and an optimal selection may differ for different performers and/or actions. Thus in this study, which is a joint work with Luis Santos, we present a dynamically adaptive sliding window, where classification entropy is used to adjust the window length and time shift parameters at every step. Fig.4.1 presents that all levels of the proposed human behaviour analysis framework can be related to this chapter.

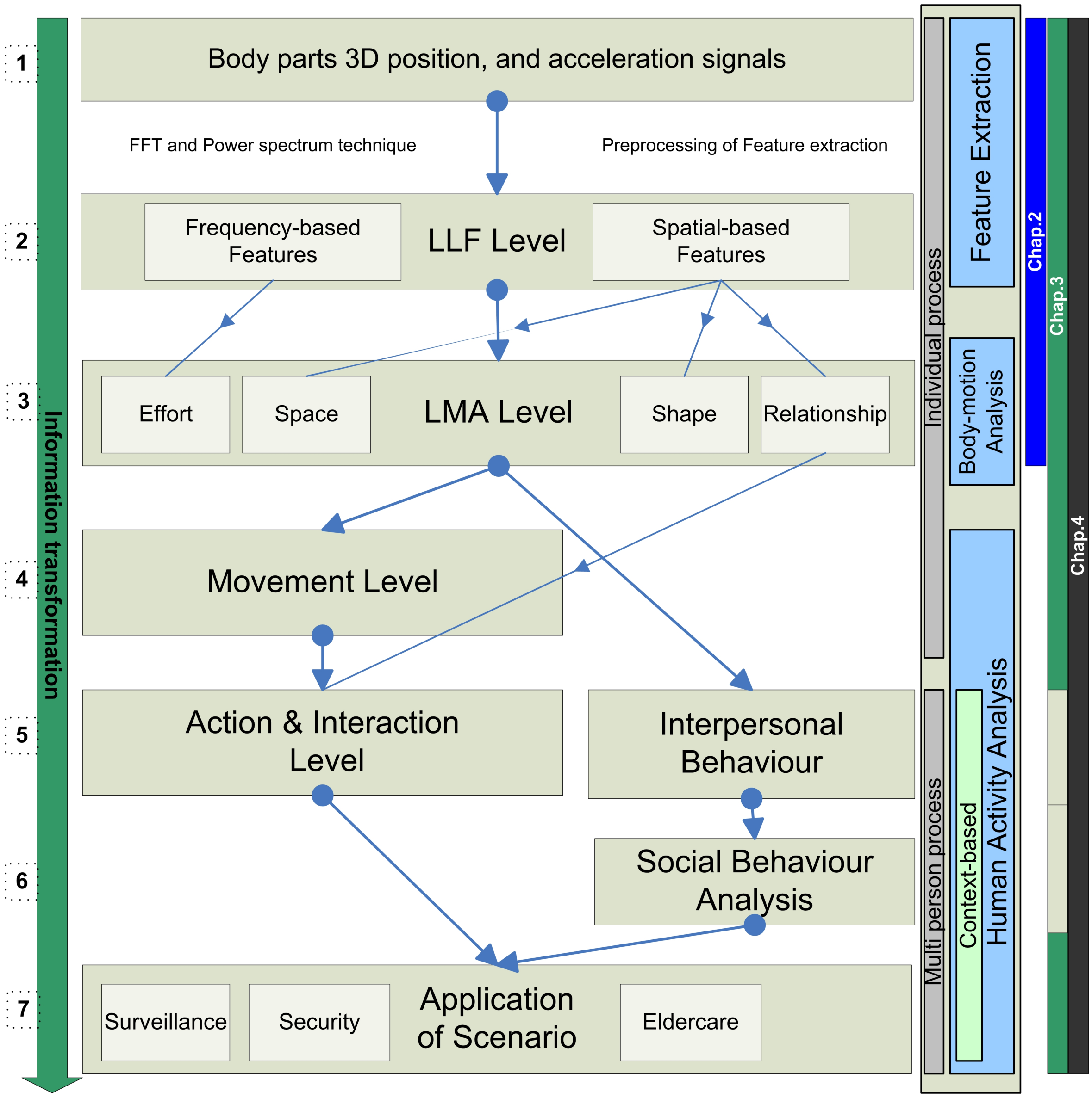


Figure 4.1: Human behaviour analysis in different levels with respect to this chapter

4.1.1 Action segmentation issues

A human action instance differs from person to person. Factors like rigidly defined performance instructions, mobility restrictions introduced by the experimental set-up, cultural or anatomical characteristics are known to introduce variability. The majority of action models usually rely on a set of assumptions, which somehow jeopardize generalization and present classification challenges. In our work, we are addressing the following problems:

- Frameworks can present high classification accuracy and the majority of the correct decisions are of low confidence. This fact is specially true as the number of different actions grows.
- The time it takes for a model to make a decision is highly dependent on the generated features, whereas decision anticipation is of great interest for an accurate temporal segmentation.

Approaches within action segmentation somehow try to address these factors. In this research, we are focused on extending our previous work using a fixed length sliding window approach [RSAD12a, LS09], improving our segmentation solution to cope with the variability issue. A survey on action segmentation [WRB11] identifies other works which also use fixed length sliding windows [ZMI01, ZSV04, FC05, KSH05, KSH07], where some are augmented with multiple concurrent classifiers using windows of different lengths at the expense of increasing computational cost. Sliding windows can also be integrated with Dynamic Time Warping [DP93, ML98], or Grammars [BI98, WB99]. Approaches which allow adjusting the sliding window parameters for action segmentation are seldom explored [LOB11].

4.1.2 Other related works on action segmentation

A recent survey by Weinland et al. [WRB11], has identified three major action segmentation categories: Sliding Window , Boundary detection and Grammar Concatenation. The already reviewed Sliding windows are used to divide a motion sequence into multiple overlapping segments, which are bounded by the window limits. The information within the window, may or may not be processed for alternative representations. Each candidate segment (or equivalent representation) is then used for sequential classification. The success of this approach strongly depends on the discriminant abilities of the generated representations. As mentioned this technique is easily integrated with the majority of static and dynamic classifiers. The major drawbacks of this technique are computational burden, and the need of multiple window sizes to overcome the variability problem. Boundary detection methods generally identify discontinuities or local extrema in observed motion signals. The boundaries usually define an implicit basic action taxonomy, without however depending on specific class definitions. A branch of works identify boundary at the cost of the dynamics of the observed signal, such as [VKD08, LLF00]. Others depend on geometric property changes observed through techniques like Principal Component Analysis [BSP+04] or piecewise arc fitting models [AM09, BN10]. A related research addresses the segmentation problem from the subspace separation perspective, exploring the so called Agglomerative Lossy Compression [RTVM10]. In [RA00], the authors apply Singular Value Decomposition (SVD) to a long sequence of optical flow images in order to detect trajectories discontinuities within SVD component trajectories. Ogale et al. [OKGfA04] also explore optical flow of body silhouettes, performing segmentation by detecting minima and maxima values of the absolute value sequence. A method using features from visual hulls is developed in [WRB06]. This category of approaches is very sensitive to noise and other related errors (e.g. camera perspectives). Additionally, it allows generic segmentation, but is not particularly suitable for labelling purposes. The focus is on boundary identification rather than interpretation of intermediate data. Lastly, Weinland et al. [WRB11] identify Grammars as another category. The common approach is to model state transitions between actions, where Hidden Markov Models (HMM) are a popular approach. Multiple methods can be used to generate features. Some examples are curvature scale space and centroid distance function [Bea06], joint angles alone [LN06, PVW07], or together with velocity profiles [KI95], dynamic system representations [PR00, THB00, Kah03] and geometrical property encoding [OSK02]. These are applied to segment and label action sequences, at the expense of computing a minimum-cost path through the model using techniques like Viterbi path, Conditional Random Fields or Markov Models. However, these methods rely on the comprehensiveness of state grammars, which may jeopardize the model effectiveness and the generalization purpose, if large amount of training data is not available.

4.1.3 Definitions and problem statement

We define motion instance as human corporal activity, which is composed by a sequence of different basic actions. Let motion instance Ω be a sequence of 3-D Cartesian coordinates defining a discrete trajectory of random duration T, for a body part such that:

$$\Omega = \begin{bmatrix} Y_1 \\ \vdots \\ Y_T \end{bmatrix}, Y \in \mathbb{R}^3 \text{ and } T \in \mathbb{N} \tag{4.1}$$

In the processing stage, Ω is divided into multiple, overlapping segments δ, generated upon using a sliding window of length ω_t samples and each δ is separated in time by a time shift Δ_t, such that:

$$\delta_t \subset \Omega : \delta_t = \begin{bmatrix} Y_{t-\omega_t} \\ \vdots \\ Y_t \end{bmatrix}, \omega_t < T \tag{4.2}$$

To avoid using the raw segment data, each δ_t is transformed into a representative feature vector $\mathcal{V}$, of lower dimension, for which a transformation function exists, such that $\delta \mapsto \mathcal{V} : \{f_1, \cdots, f_i\} \in \mathcal{V} = g(\delta)$. Our framework uses two different variable classes for analysing motion instances. One corresponds to a set of motion descriptors defined upon Laban Movement Analysis [KAQ+10], which is referred to as $c_n \in \mathcal{C}$. The other emerges a combination of variables c_n, and defines the action space $\Lambda = \{\alpha_1, \cdots, \alpha_a\}$ Consider a sequence of N action segments α, where each α_N is a non-overlapping sub-set of Ω. A single label c_n is assigned to each segment α_j during a supervised learning approach. The challenge is devising an association process to learn the action model, envisioning its generalization capabilities. The model is posteriorly used in a classification process, from which the temporal segmentation of Ω is derived.

$$\alpha_j = \begin{bmatrix} Y \\ \vdots \end{bmatrix} \xleftarrow{c_n} \mathcal{C}, \alpha_j \in \Omega \tag{4.3}$$

Consider a new action α, for which applying a sliding window approach generates multiple segments δ. Most misclassified samples have their errors emerging from the incorrect selection of the fixed window parameters. Therefore, we hypothesize that adapting these parameters at each step will improve classification, thus coping with the variability of different performances for the same action. In fact, rather than selecting a method to optimize the fixed window parameters, the challenge is to formulate a model, which iteratively readjusts the length and the time shift based on entropy feedback and knowledge of previous parameter definition. Table 4.1 summarizes the relevant variables, which are used throughout this study.

Problem - Given an activity sequence Ω, find the current window length ω that best fits the current segment δ, minimizes the classification entropy h over $c_n \in \mathcal{C}$.

$$w_{t+1} = f(h_t, w_t) \Rightarrow min(h_t) \tag{4.4}$$

Additionally, when uncertainty is high (e.g. on class transition), adjust the time step so the classifier can adapt to changes without diverging to misclassified samples.

$$\Delta_{t+1} = f(h_t, w_t, \Delta_t) \Rightarrow min(h_t) \text{ and } \downarrow errors \tag{4.5}$$

Consider sequences to be subject to noise and instance variability for the same actions performed at different instants of time.

$$\Omega' = \Omega + \eta \tag{4.6}$$

where η is a source of additive white noise.

4.1.4 Our approach

In our work, we are addressing temporal action segmentation of body part trajectories generated upon random human activity performances, as an extended solution to our fixed sliding window classifiers in action recognition [RSAD12a, LS09]. To acquire 3-D trajectories from different body parts, we are using a Motion capture (Mo-Cap) device, which is synchronized with a video sequence I of activity performances. Feature vectors are computed upon application of a Discrete Fast Fourier Transform (DFFT)

Table 4.1: Summary of relevant variables.

Variable	Set	Space
f	$\mathcal{V}$	Low-level Features
c_n	$\mathcal{C}$	Laban Descriptors
α	Λ	Action

to the acceleration signals generated from the acquired body part trajectories. This feature approach has been previously applied with success in human motion analysis problems [RSAD12b]. To learn the action model, we apply a mixture model based approach, a popular methodology in action segmentation and recognition, for which we have past experience [RSAD12a, LS09]. The sliding window approach requires the learning process to be supervised, as it plays a crucial role for the success or failure of the model [WRB11]. The learned conditional models are integrated in a Dynamic Bayesian Network classifier, which using Bayesian inference, is used to segment an activity sequence using a maximum likelihood approach.

In our experimental set-up, two different parameters are adapted, both independently and simultaneously. One strategy adapts the window length ω and is referred to, using the acronyms Adapt-ω_(ω_{min}, ω_{max}) or Fix-ω, considering whether we are using the adaptive or fixed approach respectively. The other is concerning the time shift Δ_t. The acronyms for this approach are Adapt-Δ or Fix-Δ for adaptive and fixed strategies. Acronyms are then combined, so to allow identifying the applied strategies. Our proposed adaptive sliding window methodology (illustrated in Figure 4.2), is presented as an improvement to classic fixed sliding window classification methods which:

- shows increased classification confidence;
- increases the classifier speed therefore anticipating the decision;
- dynamically adapts to different sources of performance variability.

Figure 4.3 encompasses the proposed concept illustration, of the adaptive parameter based on entropy feedback and knowledge of previous parameters.

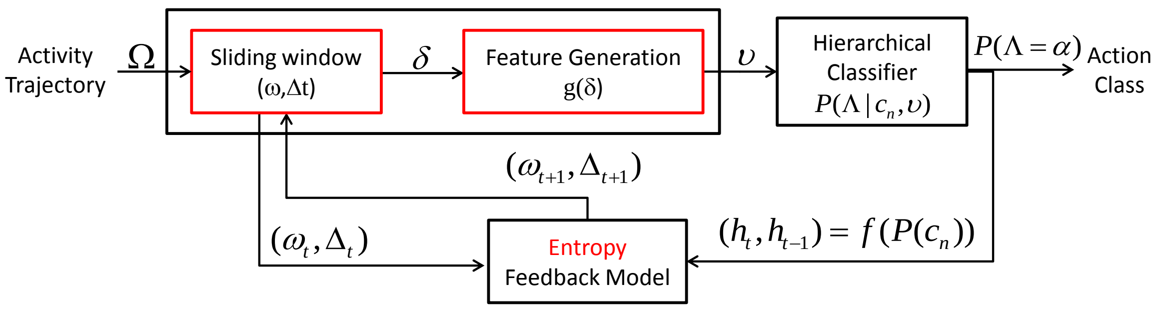


Figure 4.2: Simplified Block Diagram over-viewing our proposed approach.

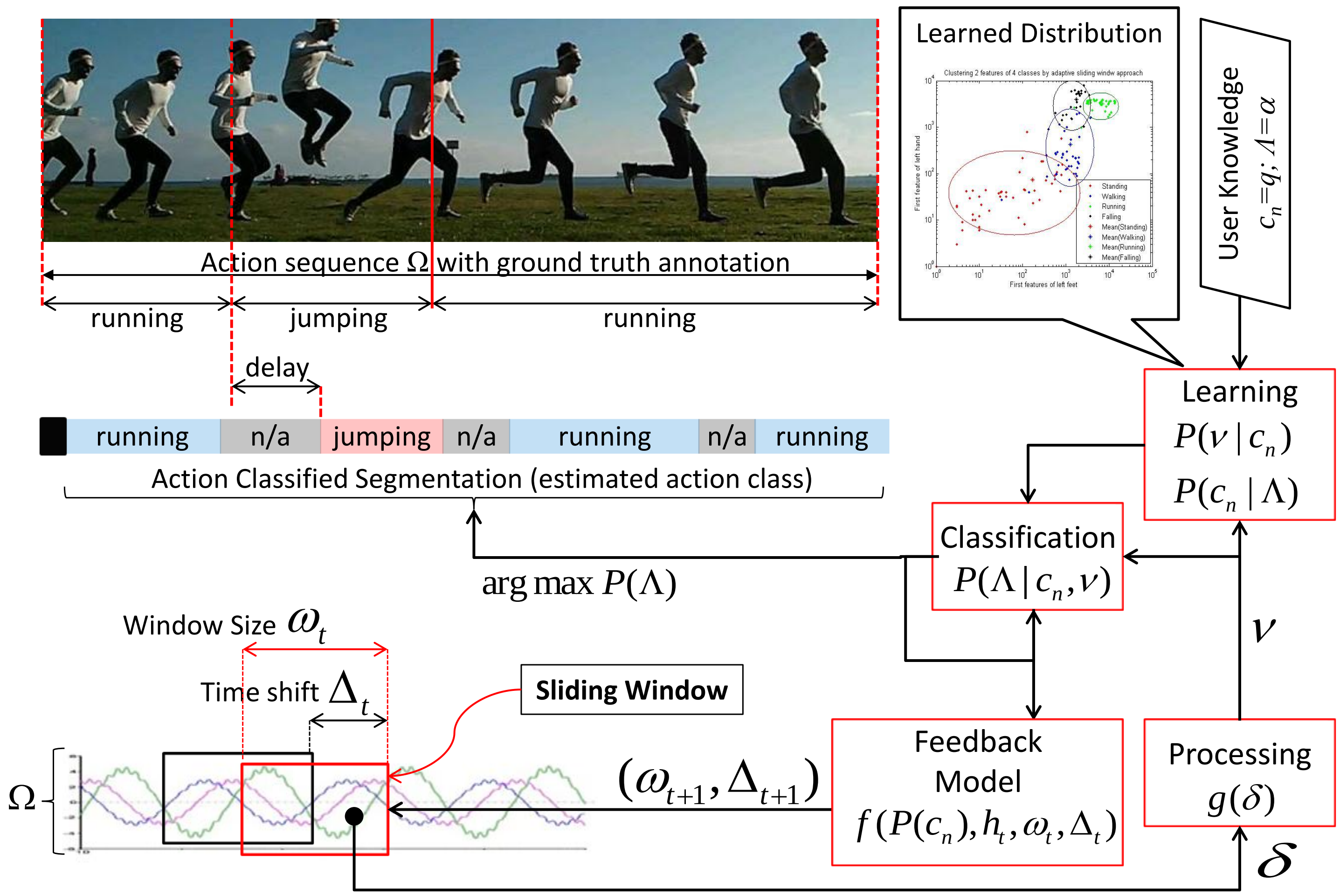


Figure 4.3: Scheme of the proposed concept along with the block diagram which formally describes our framework. An activity is segmented using a sliding window, whose parameters are adaptive based on entropy feedback. We learn body Laban and Action model, which are manually annotated within a supervised learning approach. To segment an activity in different actions, we select the most probably action Λ from our hierarchical classifier.

Fig.1.1 in the first chapter, presents the related levels of proposed human behaviour analysis framework in this chapter.

4.2 Learning the action model

In this section, the trajectory feature generation process is presented and also how different window size values influence the resulting probability distributions, upon application of the learning strategy.

4.2.1 Preprocessing

Our work emerges as an improved classification strategy to our previously developed research in action recognition, where features are represented in the frequency domain. An acceleration time series is computed from the Cartesian trajectories. Then, the Discrete Fast Fourier Transform (DFFT) and signal Power Spectrum (PS) are applied. Let the segment δ be bounded by a sliding window of length l, such that:

$$\delta = \begin{bmatrix} Y_1 \\ \vdots \\ Y_l \end{bmatrix}, Y \in \mathbb{R}^3 \tag{4.7}$$

Given the segment trajectory δ we compute acceleration $a_t = \frac{\Delta v}{\Delta t}$, where $v_t = \frac{\Delta Y}{\Delta t}$. The generated acceleration sequence $a(t) = a_1, \ldots, a_t$ will be decomposed using DFFT algorithm, generating the list of coefficients x of a finite combination of complex sinusoids, ordered by their frequency.

$$a(t) = \sum_{n=0}^{l-1} x_n e^{\frac{-i2\pi kn}{l}} \tag{4.8}$$

We can then calculate the PS of the acceleration signal, knowing that $a(t)$ is a finite energy signal, as:

$$\Phi(\omega) = \left| \frac{1}{\sqrt{2\pi}} \sum_{\infty}^{-\infty} a(t) e^{i\omega t} \right|^2 \tag{4.9}$$

The continuous approach can be generalized to discrete, for which are able to compute the energy spectral density. Feature variables are generated upon dividing the PS coefficient value ranges into four distinct classes as depicted in (4.10).

$$\mathcal{V} = \{no, low, medium, high\} \tag{4.10}$$

Further details on the presented feature generation process can be found in [RSAD12a, KAQ+10].

4.2.2 Learning

The learning method follows a Mixture Model approach, in which feature vectors are clustered according to a class c_n they belong, through supervised learning methodology (which has been conducted offline). The mixture obeys the following Gaussian decomposition:

$$P(\mathcal{V}|\mathcal{C}) = \sum_{i=1}^{n} \phi_i \; g(c_i|\mu_i, \sigma_i) \tag{4.11}$$

where class c_i is represented by an average vector μ_i and a covariance matrix σ_i. To evaluate the action model, we assess class variance which is an indicator of dispersion, and separability criteria for measuring inter-class distances. Variance σ_i appears directly from the formulation of equation (4.11). A popular measure is the Fisher's Discriminant (FD) [Fis36], which is applied to measure the separability between two classes. Rao [Rao48] generalized the FD to more than two classes, an extended formulation to find the subspace containing all class variability. First we define the class scatter as:

$$S_c = \frac{1}{n_i} \sum_{i=1}^{n} (x_i - \mu_c)(x_i - \mu_c)^T \tag{4.12}$$

where n_i is the number of samples for a given class i. From the class scatter, we can compute the within class scatter $S_W = \frac{n_i}{n} \sum_{i=1}^{c} S_i$, with n the total number of samples. Considering the Gaussian Mixture Model defined in (4.11), the between class variability can be defined for each class as:

$$S_B = \sum_{i=1}^{c} \frac{n_i}{n} (\mu_i - \mu)(\mu_i - \mu)^T \tag{4.13}$$

where μ is the mean of class means and n_i the samples for class i. The class separability will be given by

$$J = \frac{det(V^T S_B V)}{det(V^T S_W V)} \tag{4.14}$$

Vector V is computed by solving the eigenvalue problem $S_B V = \lambda S_W V$, where V is the eigenvector corresponding to the largest eigenvalue.

4.2.3 Experimental learning results

We now demonstrate how different lengths have direct impact in the supervised learning process. The presented results aim to show that selected values for length ω have con-

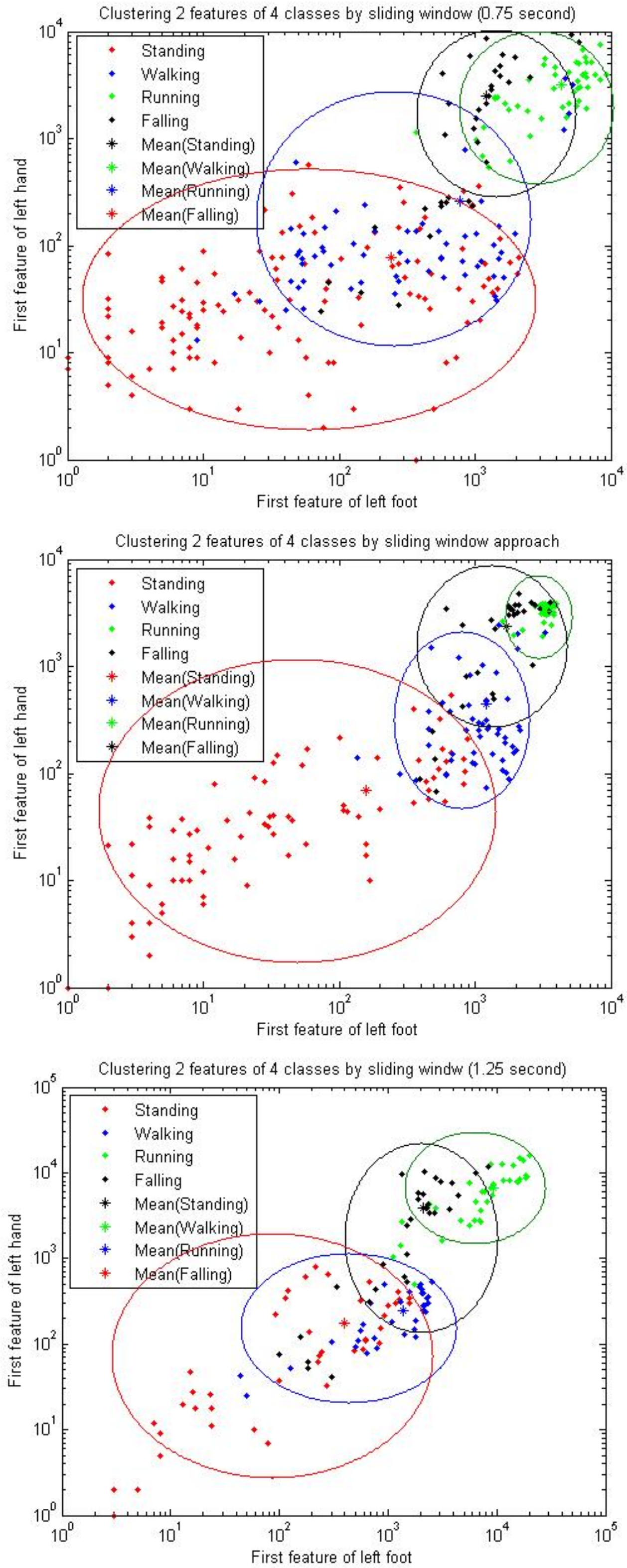


Figure 4.4: Class clusters for 4 different actions: standing, walking, running and falling. The images represent clusters computed upon a fixed window approach of: 0.75 seconds; 1.00 seconds; 1.25 seconds.

sequences which are reflected in the action model, both visually and through adequate metrics. This impact naturally propagates to the classification algorithm, and thus, in the entropy. The class clusters for the 2 dimensions of the feature vector are presented in Figure 4.4 using three different fixed window sizes. When using $\omega = 0.75$ *seconds*, we observe an overlap between class pairs standing-walking and running-falling. In the 1 *second* case, class running is completely inside falling, whereas with $l = 1.25$ *seconds*, there are multiple overlapping regions. Let us recall that the DFFT is being applied to the acceleration signal, therefore falling and running fall in the high acceleration signals while standing is mostly a static activity and walking is situated in between. Most importantly, we can visually verify that changes in the length of the window size, are reflected in the class learning process. We extend our analysis using the quantitative metrics presented in Table 4.2. Variables S_i represent scatter measures for each class. S_B refers to interclass average covariance, which can be interpreted as a dispersion measure, since it reflects the weighted distance from the class centres to their average value. The bigger the value of J, the better defined and separated are the learned class distributions. It is visible that, small changes on ω have high impact on class dispersion.

Table 4.2: Generalized linear discriminant analysis coefficients for the cases presented in Figure 4.4

		Window size (ω)		
		0.75 s	1.00 s	1.25 s
Scatter Values	S_1 $[s_x, s_y]$	[1187,2657]	[904,1493]	[1859,3632]
	S_2 $[s_x, s_y]$	[2273,1671]	[533,572]	[5731,4138]
	S_3 $[s_x, s_y]$	[443,114]	[245,93]	[465,202]
	S_4 $[s_x, s_y]$	[1185,643]	[609,548]	[794,148]
	J	1.191	29.594	12.279

Table 4.3: Summary of implicit signal rules. N/R = Not relevant.

dH	h	dW	w	d^2H	h	$\hat{w}$
+	Worst	+	Increasing	N/R	N/R	(-) Shrink
0	Stable	+	Increasing	+	Increasing Tendency	(-) Smaller Shrinkage
0	Stable	+	Increasing	-	Decreasing Tendency	(+) Smaller Growth
-	Good	+	Increasing	N/R	N/R	(+) Growth
+	Worst	-	Decreasing	N/R	N/R	(+) Growth
0	Stable	-	Decreasing	+	Increasing Tendency	(+) Smaller Growth
0	Stable	-	Decreasing	-	Decreasing Tendency	(-) Smaller Shrinkage
-	Good	-	Decreasing	N/R	N/R	(-) Shrinkage

4.3 Action classification

Our framework aims to segment actions in different abstraction symbolic levels, by means of a Bayesian classifier. Those levels are:

- Laban Movement Analysis: a set of activity invariant descriptors based on the LMA's component Relationship.

- Action: a variable whose states represent different movements as a combination of Laban variables.

The input signal in our experimental set-up are 3-D Cartesian trajectories of body parts and stationary object positions, acquired using a Motion Capture device, the inertial measuring unit based MVN suit from XSENS and Polhemus Liberty magnetic tracker. The generated spatial and frequency-based features are used as evidence in a Bayesian classifier model towards action segmentation. We have used our public domain database.

4.3.1 Action classification model

The action model is a hierarchical framework, in which inference occurs sequentially. To learn the model two strategies are assumed. To associate Laban variables to the frequency based features, we use Gaussian distributions. While learning the action model, a statistical approach is applied, where occurrences of c_n are accounted and normalized, into generating histogram probabilistic distributions. The first layer of the action model is parametrized as:

$$P(laban|\mathcal{V}) = P(laban)\frac{\prod_{q=1}^{i} P(\mathcal{V}_q|laban)}{\prod_{q=1}^{i} P(\mathcal{V}_q)} \tag{4.15}$$

We will be focusing our attention at this level, because it is where the window parameters will have most impact. In fact, the Laban model is learned based on the data bounded by the window. The entropy used to feedback the window's parameters is computed from the output $P(laban|\mathcal{V})$. The action variable states are inferred as a combination of previously estimated laban variables. An action is inferred based on:

$$P(action|laban) = P(action)\frac{\prod_{q=1}^{n} P(laban_q|action)}{\prod_{q=1}^{n} P(laban_q)} \tag{4.16}$$

The estimation occurs using Bayesian inference algorithms, where a Maximum A Posteriori (MAP) approach is applied, which is done using numerical approach, given that our formulation poses a closed-form solution. The most probable state for a variable θ upon knowledge from observations x is given by:

$$\hat{\Theta}(x) = arg_{\Theta} max P(\theta)P(x|\theta) \tag{4.17}$$

The variable states for each abstraction level which present the maximum probability value, are selected as the ones describing the corresponding segment δ, thus segmenting a sequence Ω.

4.3.2 Adaptive sliding window

Since human movements occur variably temporally and spatially, the classification inference algorithms which apply fixed parameter sliding windows have difficulty in selecting optimal parameters. Actually, an optimal segmentation of human movements in different situations by fixed sliding approaches might not be obtained. Thus, to solve the mentioned classic sliding window approaches drawback, we propose an adaptive method, which continuously adjusts the window parameters, is proposed. In this study, as it was mentioned, we explore the effects of a couple of important parameters of the sliding window approaches; window size and time shift, in the proposed adaptive approach.

Consider that for a distribution $p = \{x_1, \cdots, x_n\}$, the entropy can be estimated by $h = -\sum_{i=1}^{n} P(x_i) log P(x_i)$. Bear in mind, entropy is a normalized value, $h \in [0, 1]$.

4.3.2.1 Window Size

Table 4.3 explains the proposed adaptive window size approach. To represent the idea more clearly, we assume the different cases as following (h_t, $w_t$denote Entropy value and Window size at time t, respectively.);

- The entropy value is decreasing; $h_{t-1} > h_t$, it means the model has become less uncertain at time t, which was obtained from previous window sizes w_{t-1}. Then the scale direction which is obtained from the first order backward difference, does not need to be changed, and the window size follows the same direction.

- The entropy value is increasing; $h_{t-1} < h_t$, it means the model has become less certain at time t, which was obtained from previous window sizes w_{t-1}. Then the scale direction which is obtained from the first order backward difference, needs to be changed, and the window size needs to follow in the opposite direction.

- The first backward difference is zero, then the first order backward difference is replaced by its second order, which represents the direction tendency. Equivalent to analysing the second derivative for a continuous time series, we assume that upwards concavity represents a tendency to increase and vice-versa. However the scaling factor needs to be constrained, depending on using the first and second order difference.

Formulation

The basic definition for adapting window length can be defined as; $w_t = (1+\alpha)w_{t-1}$, and the variable $\alpha = [a_{min}, a_{max}]$ a scaling factor such that:

$$w_{min} \leq w_t \leq w_{max} \tag{4.18}$$

where w_d, w_{min} and w_{max} denotes default, minimum and maximum window size, respectively. The scaling direction according to the aforementioned rationale , is formulated mathematically as $-\frac{dH}{dt}\frac{dW}{dt}$, where the $\frac{dH}{dt}$ and $\frac{dW}{dt}$ denote first order difference of entropy and window size (H and W denote Entropy and Window size in a time series, respectively.). Where $\frac{dH}{dt} = 0$, this argument is replaced by the second order backward difference $\frac{d^2H}{dt^2}$.

However, the second order difference is considered a weak indicator. Therefore, we propose a couple of constraints (a, b) that provides the impact of $\frac{dH}{dt}$ more than $\frac{d^2H}{dt^2}$, in the model. Thus, we obtain:

$$-\frac{dW}{dt}\left(a*\frac{dH}{dt}+b*\frac{d^2H}{dt^2}\right) \tag{4.19}$$

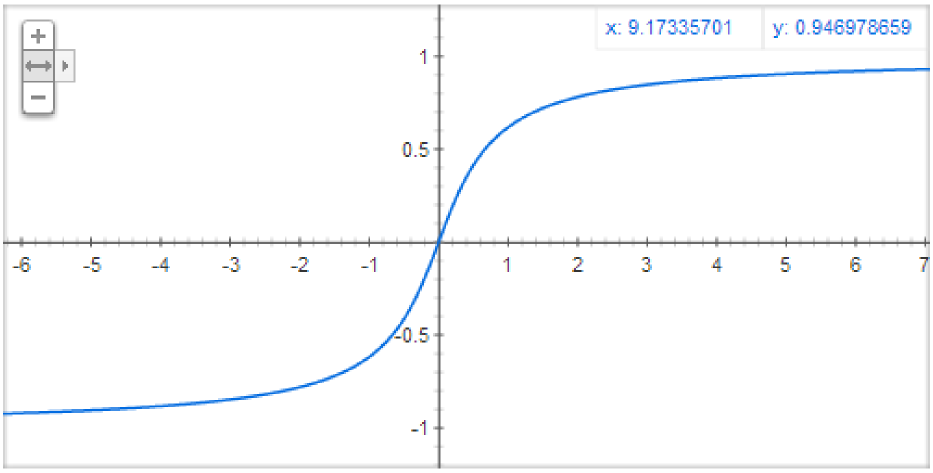


Figure 4.5: Envelope function for the growth percentage. When $x \rightarrow \infty$ then $y \rightarrow 100\%$

Another specific case, which should be considered, is where $\frac{dW}{dt} = 0$, in which the result of equation 4.19 leads to zero. The solution is making scaling direction converge to the default window size, Then:

$$(w_d - w)\left| a * \frac{dH}{dt} + b * \frac{d^2H}{dt^2} \right| \tag{4.20}$$

Then with respect to the derivation of window size, the scaling direction $\vec{a}$ can then be summarized as:

$$\vec{\alpha} = \begin{cases} -\frac{dW}{dt}\left(a * \frac{dH}{dt} + b * \frac{d^2H}{dt^2}\right) & , \frac{dW}{dt} \neq 0 \\ (w_d - w)\left| a * \frac{dH}{dt} + b * \frac{d^2H}{dt^2} \right| & , \frac{dW}{dt} = 0 \end{cases} \tag{4.21}$$

The obtained scalling factor answers that how much should the window grow or shrink needs to be address ($w_t = (1+\alpha)w_{t-1}$). This factor should be proportional to the margins between the current and maximum/minimum values for window size (w_{max} and w_{min}). In addition, the selected function should be symmetric to the origin, meaning that the factor α should share the same signal as $\vec{\alpha}$. Thus, equation 4.22 was proposed:

$$\alpha = \frac{\sqrt{(1+4\vec{\alpha}^2)} - 1}{2\vec{\alpha}k} \tag{4.22}$$

where k is an inverse proportional factor which may limit growth (default $k = 1$). Figure 4.5 illustrates equation 4.22 for a clearer visualization.

Approach	Description
Adapt1-Δ	During movement transition, apply short time shifts
Adapt2-Δ	During movement transition, apply long time shifts
Adapt3-Δ	The movement transition dedicated by first entropy deviation

Table 4.4: Time shift approaches

To avoid scaling the window size beyond the defined limits in the equation (4.18), the following formulation is proposed:

$$w_t = \begin{cases} w_{t-1} + \alpha|w_{max} - w_{t-1}| & if\ \vec{\alpha} > 0 \\ w_{t-1} + \alpha|w_{min} - w_{t-1}| & if\ \vec{\alpha} < 0 \end{cases} \tag{4.23}$$

which means that we are growing only a percentage of what is left within the window limits, assuring the window will never grow beyond them.

4.3.2.2 Time Shift

The other relevant parameter in sliding window approaches, is time shift, which concerns about segment overlap and the time between each classification. Selecting an appropriate value might present itself as an easier task than with the size parameter. However, as previously stated, we hypothesize that adjusting the time shift can optimize the segmentation process, speeding up the classifier and reducing the redundancy and adjusting segment overlap accordingly. Let us consider the time shift Δ limits as $\underbrace{\frac{f}{n}}_{\Delta_{min}} < \Delta < \underbrace{\frac{(n-1)*f}{n}}_{\Delta_{max}}$, which f denotes the acquisition frequency and n denotes a constaint.

During the exploration of different strategies to adapt time shif based on the values of the entropy, three different approaches are selected, which are summarized in Table 4.4, and introduced as following:

1. When entropy is high, we want to apply short time shifts. This approach aims at an exhaustive exploration of the data, by augmenting the number of analysed samples per second. Although we recognize that increasing the number of samples in degenerate data samples will naturally increase the number of miss-classified samples, we expect true positive results to be in greater number, resulting in a better overall accuracy ratio. The proposed formulation for this first approach

(Adapt1-Δ), is as follows:

$$\triangle_{t+1} = \frac{w_t - (h_t * w_t)}{f} \tag{4.24}$$

where f stands for sampling frequency, h_t the entropy at instant t and w_t the current window size measured in samples.

2. During action class state transitions, entropy values tend to be higher. In this case, we hypothesize that forwarding the window to a time period where the new action is already well defined can reduce the number of false positive results. Hence, we want to extend the time shift to its maximum value, thus yielding a minimum successive window overlap. Therefore, we propose the following formulation, which reflects our idea (Adapt2-Δ):

$$\Delta_{t+1} = \frac{w_t - ((1 - h_t) * w_t)}{f} \tag{4.25}$$

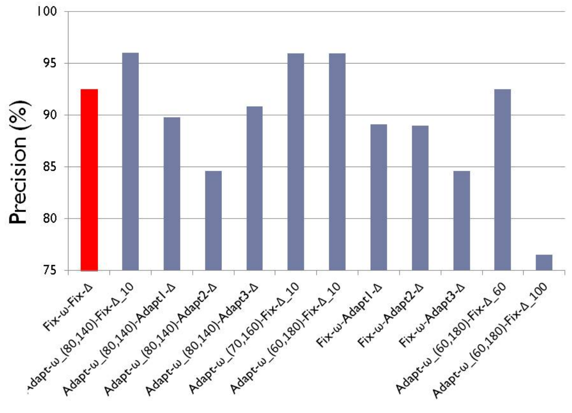
3. Another approach to detect class state transitions is proposed by addressing entropy when it becomes a volatile signal, i.e. it experiences big differences in consecutive computed values, which is reflected in its first derivative, as is illustrated in Figure 4.7. It means the high deviation in entropy signal is a good indicator for action transitions. In the low deviation entropy value the second approach was used. Thus equation (4.25), integrating the first order backward difference for the entropy signal, is proposed as:

$$\Delta_{t+1} = \begin{cases} \frac{w_t - ((1 - \bigtriangledown H) * w_t)}{f} & , \bigtriangledown H \geq thr \\ \frac{w_t - ((1 - h_t) * w_t)}{f} & , \bigtriangledown H < thr \end{cases} \tag{4.26}$$

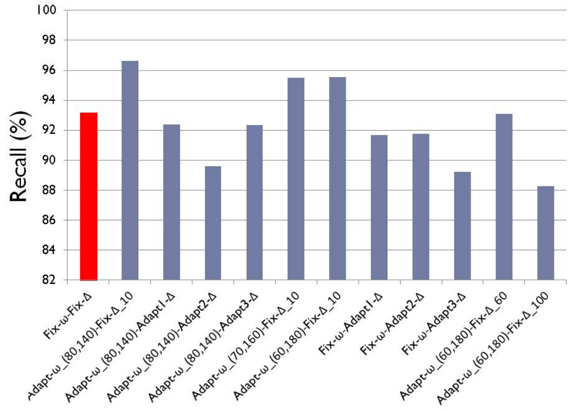
where $\bigtriangledown H = h_t - h_{t-1}$ corresponds to the first order backward difference, and thr to a pre-defined numerical threshold.

4.3.3 Classification results

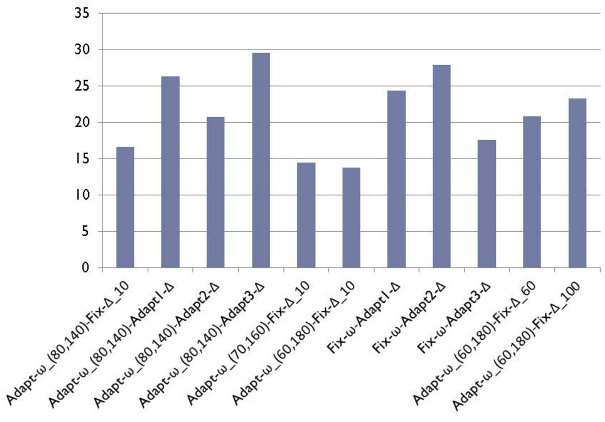
To evaluate the effects of the two mentioned parameters (Window size and time shift), with respect to the proposed approaches, the classification results of different combinations of the approaches are presented on the Figure 4.3.3-a) and Figure 4.3.3-b) by two different measurements (Precision and recall). Precision measures the number of



a)



b)



c)

d)

	W	Ru	S	Ri	F	No
Walking	1	0	0	0	0	0
Running	0.16	0.84	0	0	0	0
Sitting	0	0	0.97	0	0.03	0
Rising	0	0	0	0.98	0	0.02
Falling	0	0	0.03	0	0.97	0
No Move	0	0	0	0	0	1

Figure 4.6: (a) Precision and (b) Recall measures for the different scenarios (enumerated in Table4.5), (c)classification confidence improvement ration (%) with respect to fixed approach and (d)per-frame classification accuracy for the best approach Adapt-10-Tsfix-80-140.

Table 4.5: Different combinations of proposed approaches (ω_t and Δ_t denote window slide size and time shift). The description is defined as (type of approach, type of ω_t approach (if it is adaptive=min ω_t-max ω_t), Δ_t approach (if it is adaptive=which one of them/fixed=the amount of Δ_t)).

Acronym	Description
Fix-ω-Fix-Δ	Fixed (our previous approach), fixed ω_t and Δ_t
Adapt-ω_(80,140)-Fix-Δ_10	Adaptive, Adaptive ω_t (80-140), Fixed Δ_t (10)
Adapt-ω_(80,140)-Adapt1-Δ	Adaptive, Adaptive ω_t (80-140), Adaptive Δ_t
Adapt-ω_(80,140)-Adapt2-Δ	Adaptive, Adaptive ω_t (80-140), Adaptive Δ_t
Adapt-ω_(80,140)-Adapt3-Δ	Adaptive, Adaptive ω_t (80-140), Adaptive Δ_t
Adapt-ω_(70,160)-Fix-Δ_10	Adaptive, Adaptive ω_t (70-160), Fixed Δ_t (10)
Adapt-ω_(60,180)-Fix-Δ_10	Adaptive, Adaptive ω_t (60-180), Fixed Δ_t (10)
Fix-ω-Adapt1-Δ	Adaptive, fixed ω_t , Adaptive Δ_t
Fix-ω-Adapt2-Δ	Adaptive, fixed ω_t , Adaptive Δ_t
Fix-ω-Adapt3-Δ	Adaptive, fixed ω_t , Adaptive Δ_t
Adapt-ω_(60,180)-Fix-Δ_60	Adaptive, Adaptive ω_t (60-180), Fixed Δ_t (60)
Adapt-ω_(60,180)-Fix-Δ_100	Adaptive, Adaptive ω_t (60-180), Fixed Δ_t (100)

correctly classified samples, i.e. the model accuracy, and is given by:

$$precision = \frac{true\ positive}{true\ positive + false\ positive} \tag{4.27}$$

Precision is mostly used together with Recall, which represents the number of relevant classifications within all the results yielding a given class, such that:

$$recall = \frac{true\ positive}{true\ positive + false\ negative} \tag{4.28}$$

The different adaptive parameter approach combinations are listed on Table 4.5. As can be seen on the diagrams, using adaptive Window size with fixed time shift, improves the classification results, however, the different thresholds for shrinking and extending of the window slide did not affect the results impressively. On the other hand, The size of time shift affected the results significantly. As it was mentioned before, the mentioned approaches try to adjust the model to be more confident when making classification decisions. Thus, Figure 4.3.3-c) presents the improvement in classification results confidence, which constitutes a relevant achievement. The vertical axis values represent the ratio between the number of accurate samples in each of the adaptive approach, when compared with the fixed strategy. As can be seen, all of the approaches are successful in improving model confidence. Approaches using adaptive time shift, especially those which use the second and third time shift approaches, exhibit better improvement in confidence ratio. The left axis represents the augmented confidence, in percentage values, with respect to the fixed window approach. In addition to the presented results, Table 4.3.3-d) shows the confusion table with the per-frame classification amongst all available classes. As visible, the adaptive sliding window size has show impressive accurate results, with an overall ratio of 95%, which is an improvement with respect to our previous fixed approach [RSAD12a], which is depicted in a red bar in Figures 4.3.3-a) and 4.3.3-b). We conclude the section with Figure 4.8, where we can see an action sequence, the ground truth annotation and the corresponding delay and classified classes.

4.3.4 Action anticipation

One other relevant factor is the convergence speed. In this subsection we analyse the amount of missed frames in the classification process, i.e. the number of frames between the ground truth annotation and the actual model classification. This effect is specially felt on action transitions, where the model needs to re adjust the classified state from one action class to a different one. Figure 4.7 illustrates the differences between using fixed and the adaptive time shift approach. It aims to demonstrate that we can anticipate the convergence to the correct class with respect to ground truth annotation. The Bayesian nature of the classifier, will show some resistance to this change, due to the effect of the prior probability, which naturally delays the state transition. The lowest image of Figure 4.7, presents that most of the approaches are improved the convergence speed particularly the approaches belong to adaptive window size, with fixed time shift. We can see there are approaches which reduce the delay in almost 70% with respect to fixed with approach, whereas our best approach in terms of precision and recall, also reveals itself to be the best in terms of speed improvement. In terms of segmentation accuracy, it means that segments will be labelled much more accurately, due to the fact that model classification decisions tend to be closer to their ground truth markers.

4.3.5 Result discussion

Based on the presented results, we can observe that the proposed approaches can improve the classification accuracy, confidence, and the speed with which the model achieve its decision with respect to an action class. The two tampered parameters are

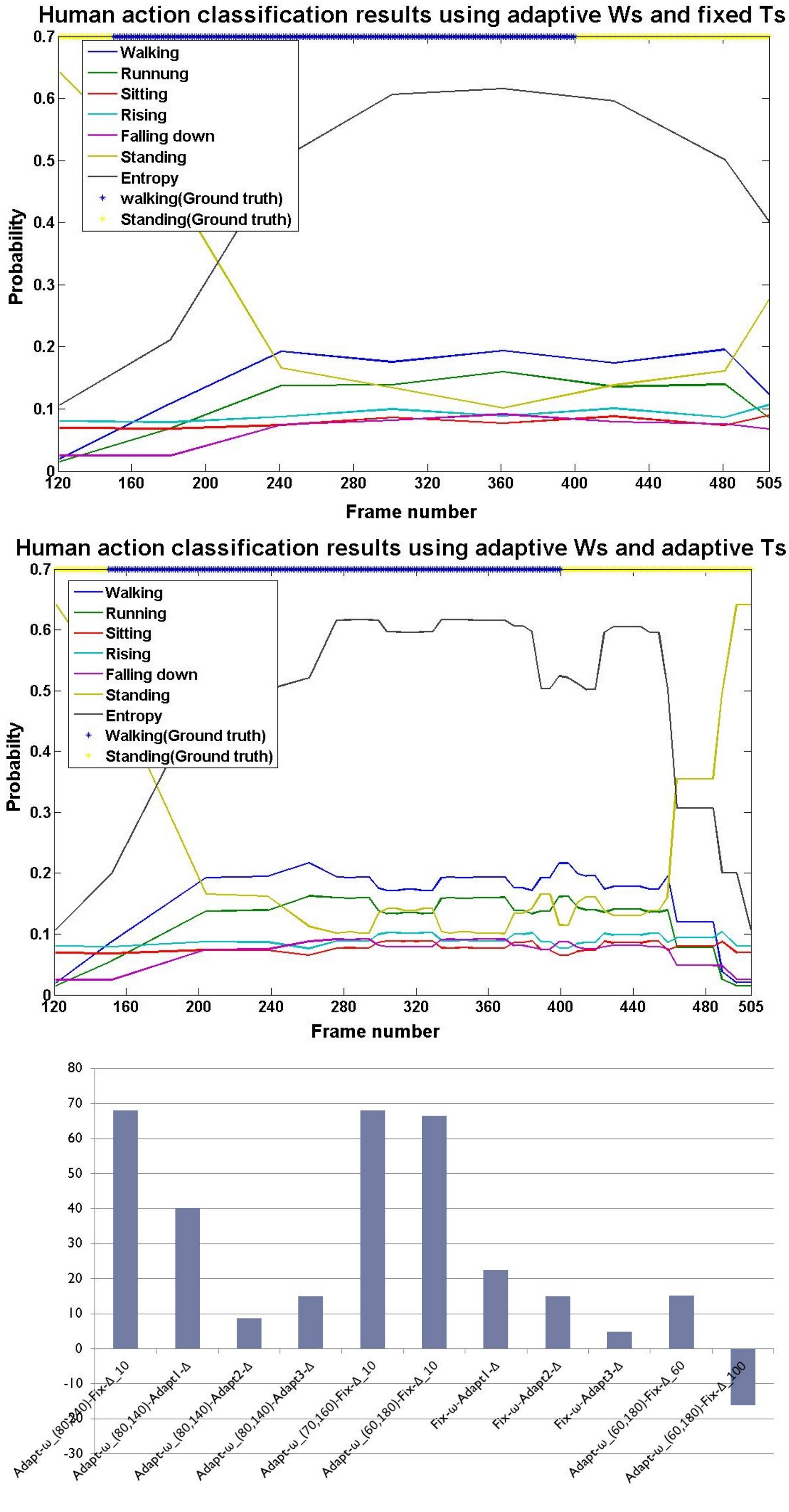


Figure 4.7: A sample results of adaptive sliding window approach using fixed and adaptive time shift approach (The colored top bar of the frames show the ground truth, the black line shows the entropy signal). Convergence speed Improvement (Percentage) with adaptive approaches when compared to fixed approach delays.

effective in some of the measurements which are categorized on Table 4.6.

As it was expected, the adaptive slide window approach shows high impact to improve the all mentioned important classification process outputs. However it shows less impact than the adaptive time shift approaches on the results confidence. On the other hand, adaptive time shift approaches have more capabilities to deal with low confidence data. Thus using adaptive time shift approaches can be complementary of the adaptive slide window approach to increase the results confidence.

Table 4.6 shows that shorter time shift affects the classification results and the action anticipation speed. However, it should be mentioned that by having shorter time shift the computational process increases. Therefore, there is a trade off between amount of time shift and the computational process, which is where the adaptive time shift approach can play a relevant role.

Table 4.6: Compare the effect of different parameters with different approaches

Measure	Description
Classification results	adaptive ω_s with fixed shorter Δ_t , ω_s's thresholds have less effect
Results confidence	adaptive Δ_t, especially 2nd and 3rd Δ_t approaches
Action anticipation speed	adaptive ω_s with fixed shorter Δ_t, ω_s's thresholds have less effect

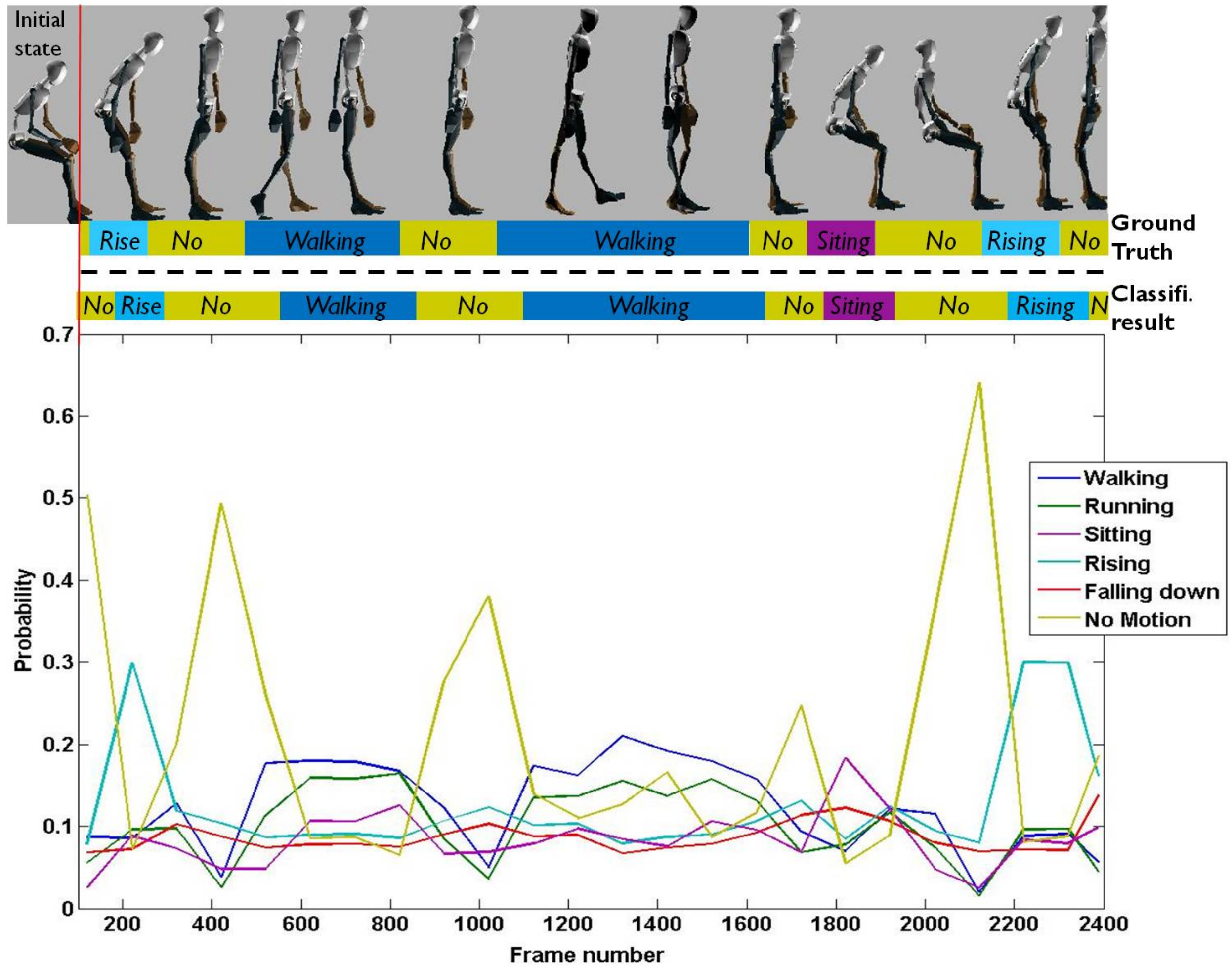


Figure 4.8: A sample of the proposed adaptive sliding window approach output on a sequence of human actions. In the bottom, the diagram shows the output of the proposed model which presents the probability of all classes on each frame (Frame rate is 120 Hz). First bar (from top) shows a sequence of actions which was performed by an actor with respect to the frame number. Second and third bars show the ground truth and the classification results with respect to the frame number, respectively.

4.4 Conclusions and future work

In this study, we propose a solution to action classification, an adaptive approach to continuously adjust the two key parameters in sliding windows: size and time shift. We have demonstrated that changes in these parameters have a high impact in the model learning. We have posed this as an entropy minimization problem, formulating a feedback model, based on entropy and previous sliding window parameters, which allowed the window to continuously adapt itself to the classification process. We have tested numerous scenarios, using different values for the limits of each parameter, and successfully demonstrated our approach to improve results, verified through adequate classification metrics: precision, recall, accuracy and convergence time (measured in frames). Moreover, our formulation is generalizable, i.e. it can be applicable to abstract classification frameworks, as long as they are based on the sliding window paradigm and values for entropy and window parameters are available.

Our future work encompasses the extension of our research to an accurate selection of window parameter limits. We expect to obtain generalizable limit selection, which can be applied in general classification problem in which a set of variables are known.

Chapter 5

Social context-based analysis

5.1 Introduction

A recent research trend is trying to analyse complex human activity, which usually appears under social contexts, which is named Interpersonal Behaviour (IB). This work proposes a model to parametrize a set of IB characteristics using a body movement descriptor, Laban Movement Analysis (LMA) [Hut74], by formulating the model under a Bayesian Network (BN) formalism (Fig. 5.1).

In society, people interact and influence each other, generating a number of complex dynamic processes (such as social roles, relationships, etc.) which happen when interacting. Human communication analysis has been studied by psychologists for decades, and they believe that there is a meaningful connection between nonverbal signals and social interactions [VSP09]. In [For10], studying of groups was named group dynamics and it is related to psychology,sociology, and communication studies. In the field, a group is commonly defined as more than one individual who is connected with others by social relationships.

As discussed before, LMA provides a language and vocabulary for interpreting body movement, which is useful to extract features from complex human movements such as interpersonal activities [LLRE07]. To analyse group dynamics, Pentland [Pen08] presents several definitions for IBs which allow to enhance the existent connections between psychology and artificial intelligence science. Certainly there are many groups investigating group dynamics in psychology, but the Pentland's recent investigation is probably the first noticeable work which attempts to analyse the IBs, relating both

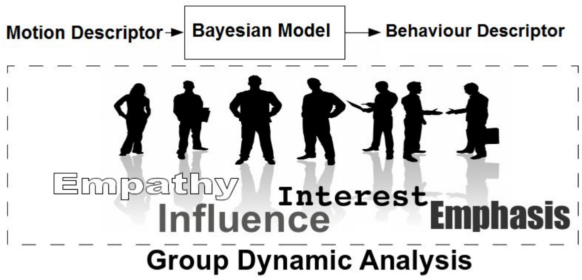


Figure 5.1: Proposed approach for interpersonal behaviour understanding

sciences.

In this study, two different approaches were proposed. The first one explored through LMA components to find the existent dependencies between people body motions for each defined IBs. Then we explored through LMA parameters in frequency domain to find more reliable features to analyse those IBs and one Social Role (SR) which is "Leading". Fig.5.2 presents the related levels of proposed human behaviour analysis framework in this chapter.

Fig.5.2 presents the related levels of proposed human behaviour analysis framework in this chapter.

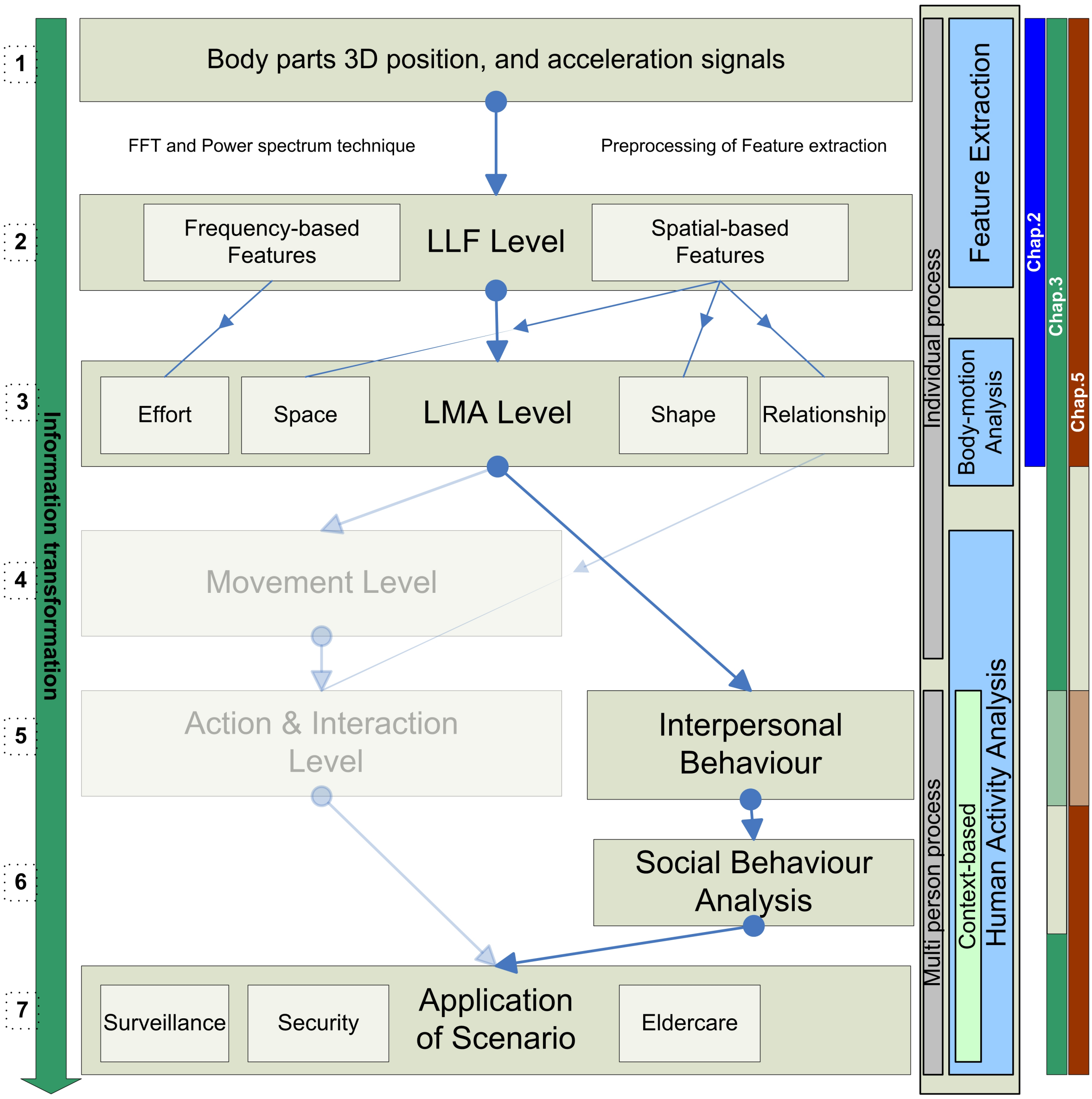


Figure 5.2: Human behaviour analysis in different levels with respect to this chapter

5.2 Related work

Nowadays human behaviour analysis is a big challenge in different fields, specially in the social aspect. Social signals which come out of a group, are very important in social science and surveillance systems, in many applications like decision making and analyzing social behaviour, which is the core of social intelligence [VPB09]. For the first time, Pentland's group [Pen08] proposed social signal processing for different applications like salary negotiations and hiring interviews. A few groups are researching about social behavior analysis using different types of sensor data, in different scenarios such as small group interactions, roles recognition (in broad cast material and small scale meeting), and user interest sensing in computer characters. In [VPB09], a couple of works related to each of these categories can be found.

The behaviour or social signal can appear from different kind of features which can be generally categorized in three parts; facial expression, voice and body motion [PPNH06]. Analyzing human motion is a prerequisite for understanding human activities, such as human-robot and human-human interaction. Analysis of human activities can be investigated in different levels. Bobick in [Bob97] presented a survey about the different levels of human motion definitions, such as human movement, activity and action. In Bobick's terminology, the movements are the lowest level of human motions which do not need any contextual or previous knowledge to be identified, but to understand human action or behaviour we need to have a sequence of human movements, related to the scenario. Several surveys about human motion-based analysis issue such as [ACLS94, CS95, Gav99, Pen00, MbK06, WHT03, Pop07] were published. As can be realized from those previous works, analyzing IBs based on human body motion is less explored.

James Borg states that human communication consists of 93 percent of body language and paralinguistic cues, while only 7 percent of communication consists of words themselves [Bor10]. Others assert that between 60 and 70 percent of all meaning is derived from nonverbal behavior [Eng06]. Thus body language analysis became very interesting for social and computer scientists.

Recently, some groups are investigating in social behaviour analysis deeply. For instance; Zancanaro et al. in [ZLP06], attempted to detect group functional roles through face to face interactions. To understand the person role, they used his/her speech and fidgeting activities which come from voice detectors and Motion History

Images (MHI). This work presents that to analyse some of IBs for a person, the features of other people are useful. As Zancanaro et al. concluded in their paper, using a state-based model (like HMM) may eventually lead to better results, and they had a plan to add more features starting from vocal energy, 3D postures and focus of attention to analyze group roles.

Human voice and visual body energy based features were used by Dong et al. in [DLC+07], to analyse human social roles under different approaches (SVM and influence model). The authors suggestions were hierarchical training and adding more features to improve their work. Sun et al. in [SNTP11], used just visual motion energy-based features to analyse an IB (mimicry). In the work it also was mentioned that kinematic-based features are needed to have better results. Thus, they "will focus on the optical flow fields in motion parts of a body, computation of kinematic features (e.g., divergence, velocity, symmetric flow fields etc.)" in their future work.

Jayagopi et al. in [JHYGP09], explained their study about the effect of different audiovisual nonverbal cues for modelling dominance patterns in a group conversation scenario. They realized that; "It was particularly interesting to observe that reasonable performance was achievable in the most dominant case without having to listen to the conversations at all.". However, it was concluded that audio-based features gives better results. It should be mentioned that they just analysed 2D based features in video side.

BN presents many advantages on using prior knowledge and modeling the dynamic dependencies between parameters of object states. In related fields, this approach is popular and researchers have been keen on applying it, such as Rett [Ret08], who applied a general BN framework for analyzing human motions, and Ryoo and Aggarwal [RA06], who presented a framework for human action modeling by using BN for analyzing human activities. There are several reasons to use a Bayesian approach in our application: Bayes theorem is valid in all common interpretations of probability, can represent and solve decision problems under uncertainty, it is a common approach to predict, an explicit approach to follow states, does not need a large amounts of data to be trained, and it is able to fuse different types of data in one model [KAQ+11].

Pentland's group in [Pen08] presented an interesting work to analyze IBs in different context like classroom, casino, etc. , defined several features as Honest signals and measured them by a mechanism, namely, Sociometer. Then, a number of social roles was defined by combination of those signals. Recently, a few works have been proposed in this direction, for example; simple body motion-based and speech-based features are used in [DLC+07], and silhouette motion-based features are used in [SNTP11], for the mentioned purpose. A brief survey about social signal analysis was published in [VPB09].

The state-of-the-art shows that there are several works which have been done in simple human motions activities and behaviours, but there is still a big gap between body motions and IBs context applications. This kind of applications, when you just rely on body motions, are very valuable in many applications such as surveillance systems, but less explored. Thus, in this work, based on the pentland's definitions to estimate the different social signals or behaviors, an approach is proposed to obtain them by using just body motion-based features, which is less explored and can be achieved by LMA descriptor [Hut74, ZB05]. The main contribution of this work is presenting a new approach to parameterize IBs using LMA components which can bridge the gap between human motion signals and the complex human behaviours.

5.3 LMA-based IBs analysis

This section focuses just on 3D body motion-based signals as a part of nonverbal signals to analyse IBs. The related works rely more on voice and facial expression features, but the contribution of this work is to explore through different features which only belong to body parts motions to analyse IBs. Based on the interested IBs definitions, the parameters in Laban components and IB level with their corresponding states are analysed and modelled using a BN technique.

5.3.1 Variable space

In order to parameterize the IBs, we should firstly define some interesting parameters or variables. The proposed model in this work is divided into two abstraction layers. Each of those layers has its own set of variables. These variables are inspired by both Pentland's and Laban's work. The model aims to use body motion information to infer IBs. Thus, we use features obtained by LMA in Feature Space, to estimate IB variables, defined by Pentland's definitions in the Behavior Space.

5.3.1.1 LMA components

Composing the Feature Space are Laban Components, which constitute the observations driving the model. LMA is a framework to describe, interpret and analyze human motions using five different components. Each component deals with different aspects of human motion [Ret08, Hut74, BPW93]. All variables defined in feature space are inside the five component set (Fig. 5.3-Left): Effort, Space, Shape, Body and Relationship.

Effort describes the dynamics of body motion [Ret08], and is divided into four qualities: Time, Space, Weight and Flow. Each of them has a bipolar state. For instance; Effort.time presents if the body part motion is in sustained (like touching carefully movement) or sadden (like punching movement) state, and Effort.space describes if the motion is in direct (like hand pointing) or indirect (like bye bye) state, etc.

Space interprets the trajectory of each body part in a 3D space [Ret08]. Researchers were discretized the direction of body motions with some states depending on their applications. Shape describes deformation of a body as a blob in three plans; sagittal,

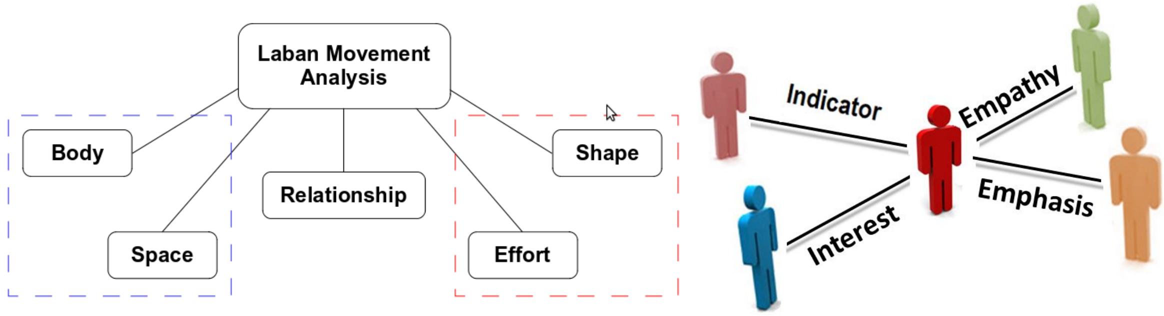


Figure 5.3: Left) LMA framework with five components [BPW93]. Right) Interpersonal behaviours which were explored in [Pen08]

LMA components	States
Space.Head	Forward, Backward
Space.Hands	Forward, Backward, Up, Down, Right, Left
Effort.Time	Sudden, Sustained
Effort.Space	Direct, Indirect
Shape.Sagittal	Advancing, Retreating
Shape.Vertical	Rising, Sinking
Shape.Horizontal	Spreading, Enclosing

Table 5.1: LMA parameters

vertical and horizontal [SD10], and it consists of some states in each plan; like if the whole body is rising or sinking in vertical, advancing or retreating in sagittal, and spreading and enclosing in the horizontal plan. Since we are using hands and head poses in this work, the mentioned blob will consist of the space between hands and head. Body shows body part relative state to body centre [BL80]. Relationship appears as the less studied component and presents the relation between body parts and environmental parameters or others [Hut74].

Depending on the objective, researchers rarely use all LMA components. To quote some examples, [KAQ+11] uses Shape and Effort for human action recognition, whilst Rett or Zhao [Ret08, ZB05] use Space and Effort to classify and analyze human gestures. Given the Pentland's descriptions of IB, the selected Laban Components for Feature Space are:

$$FeatureSpace \in \{Effort, Space, Shape\} \tag{5.1}$$

Table.5.1 presents all defined LMA parameters based on the three components for this work.

5.3.1.2 Interpersonal behaviour

The last decade brought multiple works using LMA-based computational systems to characterize different phenomena in different applications: human-robot interaction [Ret08], human gesture analysis [ZB05], rehabilitation [FW06] and human movement understanding [KAQ+11].

All over-mentioned works somehow address human gesture classification in single person perspective. This work goes one step further, using LMA concepts to characterize human behavior rather than gesture, in context of social interaction. To undertake such task, the Pentland's definitions are used to categorize IBs, which are behavior (Honest) signals present in all social interactions. Thus, this work defines the set of IB variables as (Fig. 5.3-Right)); Indicator, Empathy, Interest and Emphasis. Each of the IBs variables has two states, which are defined as follows:

$$Indicator \in \{influenced, influent\} \tag{5.2}$$

The set.5.2 presents a variable which consists of two possible states, influenced and influent. As Pentland's describes [Pen08], within a group, there is tendentially someone who tries to have an edge over the remaining. This edge is described as a person's capability to aggregate others around the same line of thought, or more generally, to be the emerge as group leader. Hence, we call it Indicator variable.

$$Empathy \in \{uncoordinated, mimicry\} \tag{5.3}$$

Mimicry is a state, which is related to Empathy behaviour, and as Pentland mentioned in [Pen08], more empathetic people are more likely to mimic their conversational partners. Therefore, the Empathy variable has two states (Set.5.3), mimicry state if there is imitation motions, otherwise uncoordinated state.

$$Interest \in \{passive, active\} \tag{5.4}$$

The Interest variable represents whether a person is engaged to the situation or outside context. This behaviour is characterized by, what Pentland describes, level of activities. Thus, we defined two states, passive and active, for this variable (Set.5.4).

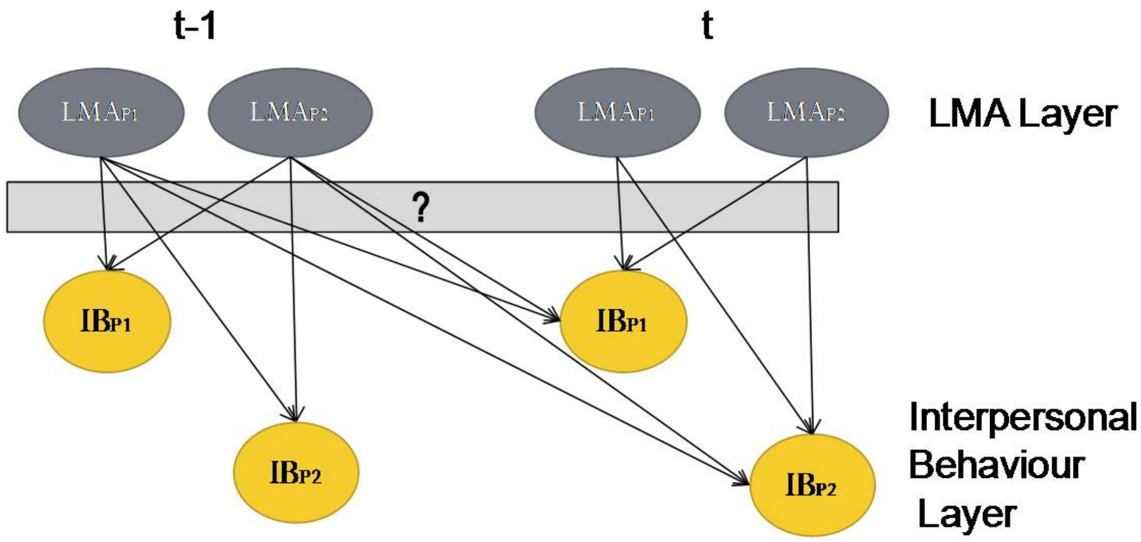


Figure 5.4: Presenting all possible connection between the two layers in different times. P1 and P2 denote the first and second persons, and LMA_{p1} and IB_{p1} denote all defined variables inside of LMA and IB for the first person.

$$Emphasis \in \{consistent, inconsistent\} \tag{5.5}$$

The last defined IB is Emphasis variable, which explains a person's focus in a situation, another person or object. If the person has a wandering mind, his/her behaviour will be variable or inconsistent. Set.5.5 defines the two possible states of Emphasis variable, which are consistent or inconsistent.

5.3.1.3 General schema of the variable spaces

The key point of this work is to explore through all the obtained features in LMA space to estimate the interested classes in IB space. We present how those features can affect the estimation of each IB, and based on that analysis, the sufficient model of each IB is proposed. Fig. 5.4 shows an overview of the different variable spaces with all possible connections between those layers. In the next section, the sufficient connection will be found for each IB.

IBs	IB definition [Pen08]	LMA [Hut74]
Indicator	More energy body motions than others	Effort
Empathy	Copying each other activities and nodding	Space
Interest	Presenting energetic motion	Effort
Emphasis	Movements become jerky or not	Effort, Space

Table 5.2: A brief description of the relation between the IB definitions and LMA components for each of the IBs

5.3.2 Global human interpersonal behaviour model

As mentioned before, this work parametrizes IBs with Laban components, by explicitly defining two different abstraction layers. The reason why this work does not infer IB from input signal features directly, is because information will be lost. There are several works that developed models to classify Laban parameters from input signal features [SD10, KAQ+11, Ret08, BPW93]. Thus, the present model uses Laban movement analysis as observations. We will describe both IB and LMA components, demonstrating the latter has enough information to characterize human behaviour.

As mentioned, four IBs are defined: Indicator, Interest, Empathy, Emphasis. In the following sections, the dependencies between LMA and Pentland's definitions are explored (as can be seen in Table.5.2 briefly) and based on them, a Bayesian model for each of IBs is proposed. As can be seen in the following subsections, Empathy and Emphasis are modeled by dynamic Bayesian approach and explain the reasons for the use of the previous knowledge.

5.3.2.1 Selection of effective LMA components in learning process

Eq.5.6 presents a general Bayesian model, and based on that, we explain the general learning process briefly. For all the variables, we only formulate the learning distributions, and the process is analogous for all.

$$P(A|B) = P(A).P(B|A)/P(B) \tag{5.6}$$

Variable A is formulated as seen in this equation. Observing the second term of the equation, we have the prior distribution $P(A)$, the likelihood $P(B|A)$ and the normalization factor $P(B)$. The likelihood is a conditional probability corresponding to previous knowledge which needs to be learned.

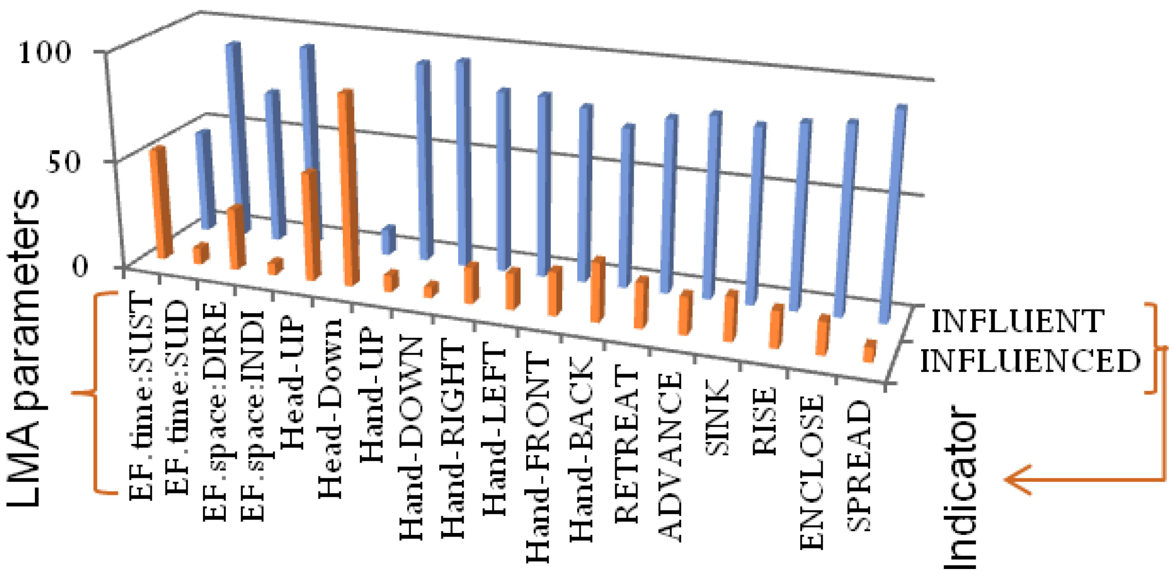


Figure 5.5: Histogram of the different LMA features for Indicator variable

Histogram-like approach is a common supervised learning method. To illustrate this method, let's analyze the learning histogram for the Indicator variable (See Fig. 5.5).

From the collected labeled signals we built an histogram counting all obtained LMA states given the knowledge of the Indicator variable state. This allows us to generate the necessary distributions, and this means that, the method allows us to visualize those LMA parameters that better discriminate the IB's states. By comparing the different states of the same LMA variable (e.g. Effort.Time for states Sudden and Sustained), it is possible to empirically select the ones that exhibit the highest dynamics between the IB variable states. In this investigation, depending on the IBs, the previous knowledge of the mentioned person and others are also used, as can be seen in Fig. 5.4.

5.3.2.2 Indicator model

In different scenarios, such as people conversation or negotiation, it is interesting to realize who has influence over other participants in many aspects of social context. While interacting with others, a person can be either influent or influenced. Indicator is the variable which we define with these two possible states.

In [Pen08] most of the influence signals analysis are based on human speech, described in some examples like a student-teacher argument or salary negotiations. However, Krauss mentioned [KH99] that hand gestures which people produce, play an

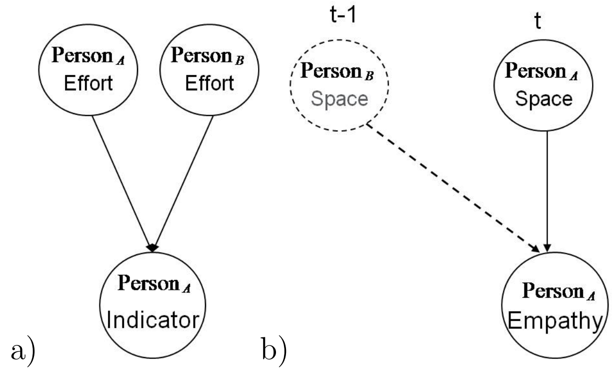


Figure 5.6: a) The dependencies between LMA parameters and Indicator and variable ($Person_A$=first person , $Person_B$=second person). b) Dependency diagram among LMA parameters and Empathy characteristic for person A. The same can be applied for person B.

important role in any communication. We interpret these facts, that when a person is trying to gain influence over others, he/she usually produces more energy through his/her body part motion, such as hand motion in a conversation scenario, to be more representative.

LMA framework encompasses a component, Effort, which analysis is concerned with the changing patterns which occur in the ebb and flow of energy within the body [Hut74]. Thus, to measure the influent and influenced states of a person, Effort component should be sufficient. In this concept, when a person has a more representative body part motion, the probability of being influent over others is higher. Therefore, the probability of a person being an Indicator, will be the probability of being influent or not, given the obtained Effort characteristics, of him/herself and others. The related histogram which is obtained in the learning process (Fig. 5.5) also proves the mentioned analysis.

Based on the dependency diagram in Fig. 5.6-a, the relation between LMA and Indicator is formulated as follows:

$$P\left(Ind_i | \prod_{h=1:n,i=1:m} Ef_i^h\right) = \frac{P(Ind_i) \prod_{h=1:n,i=1:m} P\left(Ef_i^h | Ind_i\right)}{\prod_{h=1:n,i=1:m} P\left(Ef_i^h\right)} \tag{5.7}$$

where Ind_i and Ef_i^h denote respectively, Indicator variable for i^{th}person, and Effort component variable for h^{th} body part of i^{th} person. n and m denote the number of body parts and persons.

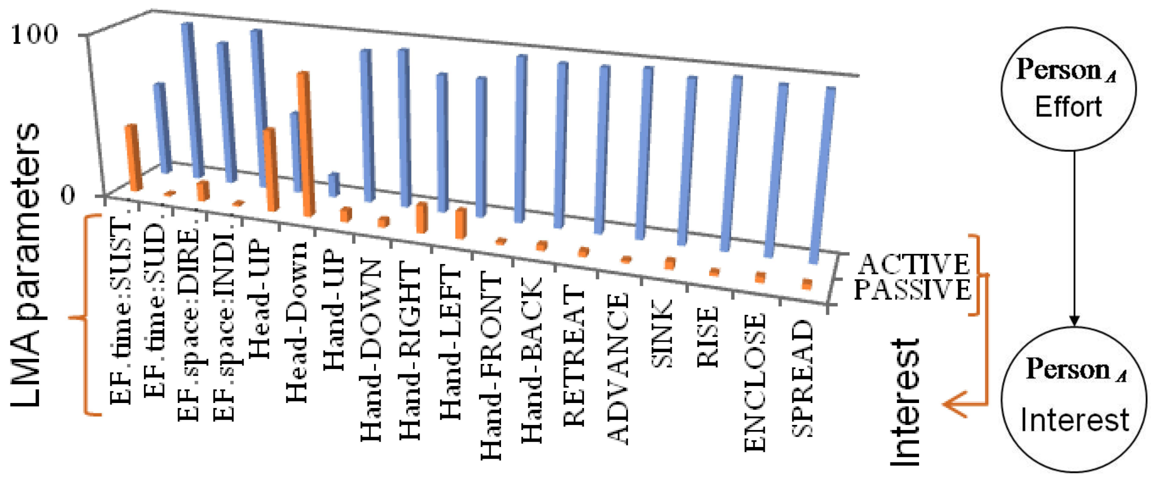


Figure 5.7: Left) Histogram of the different LMA features in different states of Interest variable. Right) Dependency diagram among LMA and Interest characteristic.

5.3.2.3 Interest model

In a social interaction, the activity level of the human body is a visible unconscious signal to present human interest and excitement level. In a communication between people, when a person is interested or excited, he/she presents more energetic motion. A visible example to show the issue presented in [Pen08] is the connection between the activity level and excitement for children in special events like a birthday party. Excited kids usually talk faster and louder, fidget more and run around, and similar of those effects also happen in adults. Thus this IB also relates to how much energy is consumed by the person which can be interpreted by Effort component of LMA.

As seen in Fig. 5.7-left, most of the features are quite dynamic, but those features which include less parameters and also also cover the mentioned analysis are selected.

The difference between Interest and Indicator variables is that for Interest just one person's data is sufficient but for Indicator, we need to have more than one person involved (in a social context).

Based on the dependency diagram shows in Fig.5.7-right, the relation between LMA and Interest behaviour is formulated as follows:

$$P\left(Int_i | \prod_{h=1:n} Ef_i^h\right) = \frac{P(Int_i) \prod_{h=1:n} P\left(Ef_i^h | Int_i\right)}{\prod_{h=1:n} P\left(Ef_i^h\right)} \tag{5.8}$$

where Int_i and Ef_i^h denote Interest variable for i^{th}person and Effort component variable for h^{th} body part of i^{th} person respectively.

5.3.2.4 Empathy model

When people who are deeply engaged in a conversation are on the same wavelength, it is called Empathy [Pen08]. Empathy can be felt by some interactive motion signals. One of the common signals is mimicry. The engaged people copy each other's activities, such as smiling, body gesture, head nodding and etc. during a conversation. Those mimicry motions, usually occur when motion is similar in space, but not necessarily relating to its dynamics. Thus, the LMA spatial-based features are more meaningful and reliable inputs to discretize this IB. Space component of LMA describes body motion trajectory, specific direction, level, distance, or degree of motion [Hut74].

Algorithm 1: Algorithm to decide the similarity value for each LMA parameter, to be used for mimicry histogram generation (see Fig. 5.6-b).

for each $f_A \in LMA\,(A_t)$
f_B: corresponding feature to f_A
if value(f_A^t) = value$\left(f_B^{t-1}\right)$
  f_A^t.mimicry $\leftarrow' same'$
else
  f_A^t.mimicry $\leftarrow' different'$
end.

To prove the idea, Fig. 5.8-left) and Fig. 5.8-right) show generated histograms that will be used for mimicry and uncoordinated states, respectively. In each of these two histograms, the first dimension shows the LMA parameters, and the second dimension indicates the decided value for mimicry or uncoordinated states. For each person (for example A) these values are obtained by considering the difference between the corresponding LMA features among the current person in time t (f_A^t) and the other person in previous time (f_B^{t-1}). Such a process is shown in Algorithm.1 for the mimicry state, and is similar for uncoordinated state.

Comparing the two histograms, it can be seen that Space component has highly distinct behaviors than the remaining. Thus in Eq.5.15, just space component features of the person and previous data of other person are used.

Fig. 5.6-b presents the relation between LMA and Empathy, which is formulated as follows:

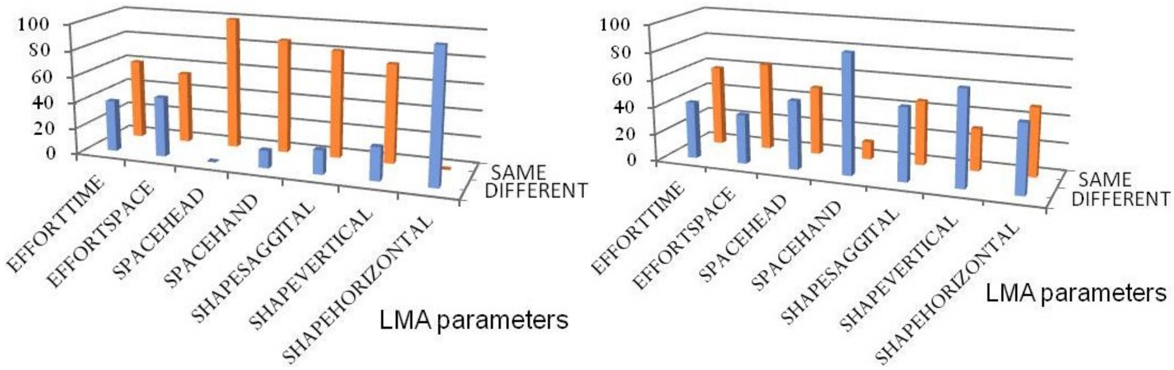


Figure 5.8: Histograms of LMA variable states similarity of two persons between time t and $t-1$ for the both Empathy variable states; Mimicry histogram is represented in the left and Uncoordinated in the right image. The blue and orange bars show statistically how much each of LMA parameters data belong to person A at time t and person B at time t-1, are different and the same, respectively.

$$P\left(Emp_i(t)\,|\prod_{h=1:n}\left(Sp_i^h(t)\prod_{j=1:m,j\neq i}Sp_j^h(t-1)\right)\right)=$$

$$\frac{P(Emp_i(t))\prod_{h=1:n,j=1:m,j\neq i}P\Big(Sp_i^h(t)\,Sp_j^h(t-1)|Emp_i(t)\Big)}{\prod_{h=1:n}P\left(Sp_i^h(t)\prod_{j=1:m,j\neq i}Sp_j^h(t-1)\right)} \tag{5.9}$$

where $Emp_i(t)$ and $Sp_i^h(t)$ denote Empathy variable for i^{th}person and Space component variable for h^{th} body part of i^{th} person at time t, respectively. n and m denote the number of body parts and persons. In the experiment, we just applied a couple of people and three body parts data, however it can be extended to more people and body parts data.

5.3.2.5 Emphasis model

When a person is thinking about different things simultaneously, his/her speech and movements become jerky and inconsistency paced ([VV96] in [Pen08]). It means that depending on context, people's emphasis can be consistent or variable, and the relative consistency or variability of human activity conveys different messages for people. Those messages can play an important role in social aspect. To estimate this IB we should look for variation of both Space and Effort components features of the person along time. When they remain constant, it means the person is focused or his/her

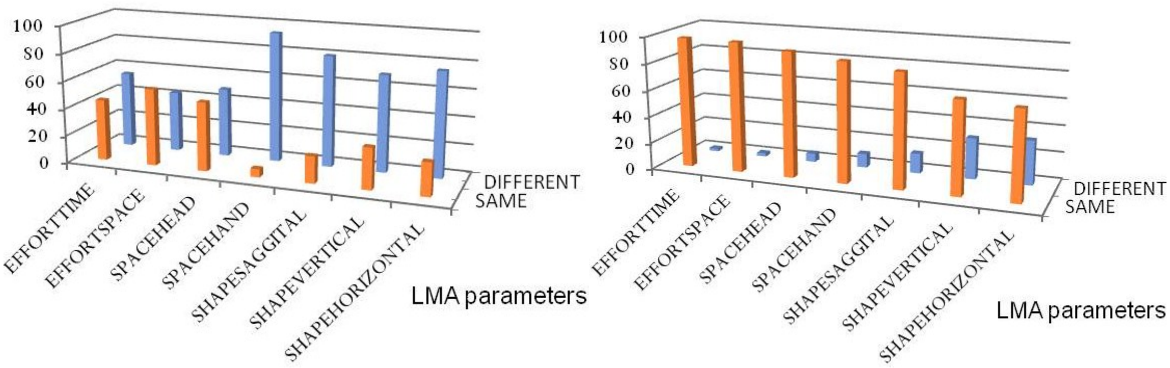


Figure 5.9: Histogram of LMA variable states similarity of one person between time t and $t-1$ for the both Emphasis variable states; Consistent histogram is represented in the left and Inconsistent in the right image. The blue and orange bars show statistically how much each of LMA parameters data belong to person A at time t and t-1, are different and the same, respectively.

behavior is consistent, giving emphasis to that person's actions.

Similar to the previous section, Histograms of the Emphasis model are generated and presented in Fig. 5.9, but instead of using previous data of other person in Empathy model, previous data of the same person is used in the current model. The left image corresponds to the Consistent state, and presents whether LMA parameters for the person at time t, correspond to the LMA parameters of the same person at time $t-1$ or not. The right image presents the same histogram but for Inconsistent state. Comparing the two histograms, it can be seen that Space and Effort components have high distinct behaviors. The following equation expresses the Bayesian model (see Fig. 5.10-a) for Emphasis variable:

$$P\left(Emf_{i(t)} | \prod_{h=1:n}\left(Sp^h_{i(t)}\, Ef^h_{i(t)}\, Sp^h_{i(t-1)}\, Ef^h_{i(t-1)}\right)\right) =$$

$$\frac{P\left(Emf_{i(t)}\right)\prod_{h=1:n}\left(P\left(Sp^h_{i(t)}\, Sp^h_{i(t-1)}\, Ef^h_{i(t)}\, Ef^h_{i(t-1)} | Emf_{i(t)}\right)\right)}{\prod_{h=1:n}\left(P\left(Ef^h_{i(t)}\right)P\left(Ef^h_{i(t-1)}\right)P\left(Sp^h_{i(t)}\right)P\left(Sp^h_{i(t-1)}\right)\right)} \tag{5.10}$$

where $Emf_{i(t)}$ and $Ef^h_{i(t)}$ and $Sp^h_{i(t)}$ denote Emphasis variable for i^{th}person and Effort and Space component variables for h^{th} body part of i^{th} person at time t.

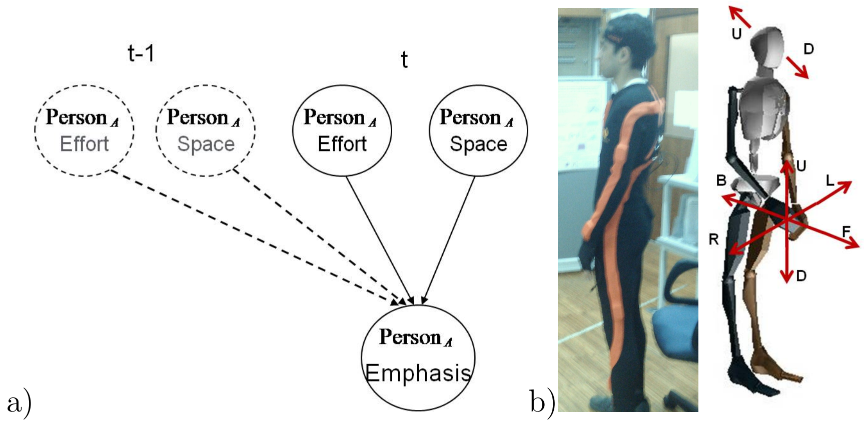


Figure 5.10: a) A dependencies diagram among LMA parameters and Emphasis characteristic. b) A person with a special suit (motion tracker) for 3D data capturing and Space parameters presented on a dummy.

5.3.3 Experiments

A set of experiments have been carried out to demonstrate the effectiveness of the proposed models. The experiments are performed in the smart-room of the MRL of ISR*. Our setup is comprised of a 3D human motion tracking sensor, called MovenSuit† (see Fig.5.10-b), and a network of cameras installed on the ceiling of the room. A set of conversation scenarios is defined in which the contexts are the IB activities performed by two persons. Note that the proposed model has the capability of being used for a multi persons case, however, here due to some limitation in the data acquisition setup we have just used two persons in order to prove the concept. In each scenario the body movements for each person are recorded using the MovenSuit device. After recording the scenarios in each sequence and for each of the acting person, an expert, called annotator, manually annotated the LMA parameters and IB states by observing the recorded videos from the ceiling cameras. The used LMA parameters and IB states are based on our definitions provided on sections 2.1 and 2.2. Each sequence has a length of about 400 seconds.

The annotated data by the expert is used for the learning stage. For classifications, the LMA parameters are automatically extracted from the 3D tracker based on our previous works [KAQ+11]. The frequency for both annotation and classification stages is 1Hz. Among the annotated data, half of them are used for learning and the other half is used for classification.

*http://paloma.isr.uc.pt/mrl/
†http://www.xsens.com/en/general/mvn

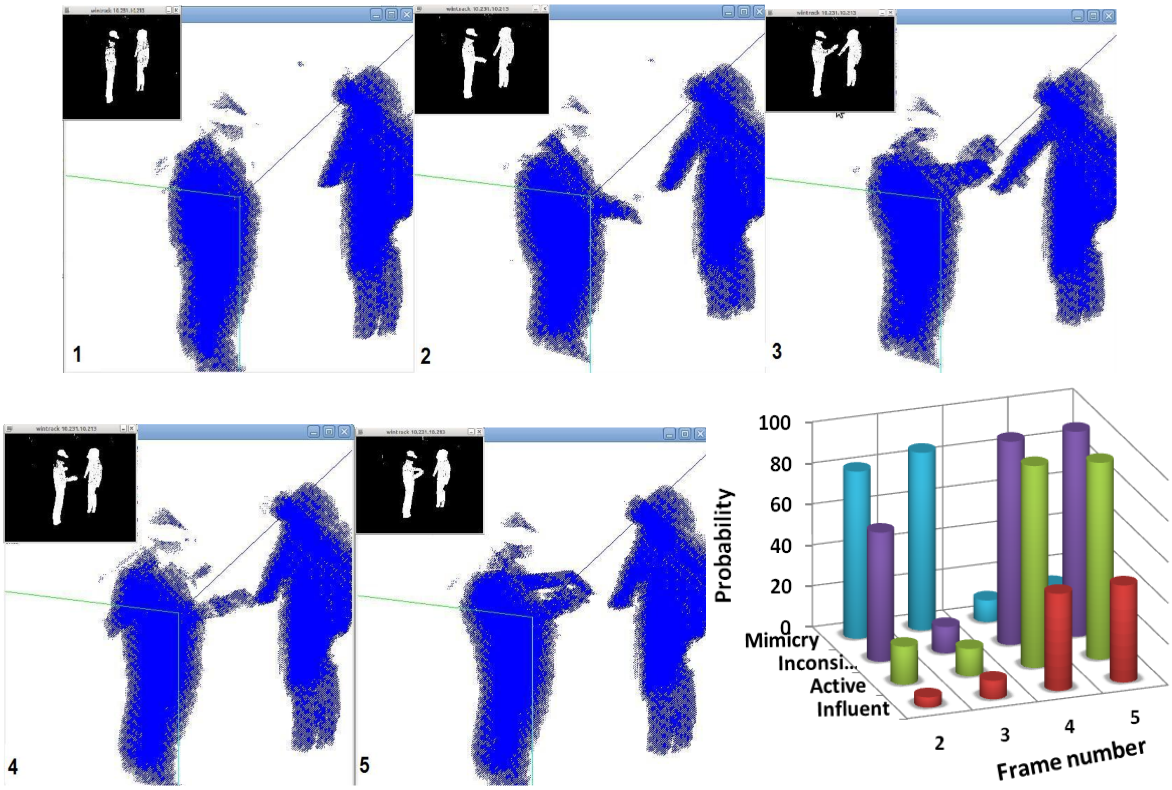


Figure 5.11: An exemplary short sequence including five frames (5 sec. length, labeled from 1 to 5). The extracted LMA features related to these sequence are fed to each IB model. The histogram in this figure represents the output for each IB for the seconds 2 to 5.

For the purpose of classification, the automatically obtained LMA parameters from each frame are fed to the proposed IB models. Fig. 5.11 shows an exemplary short sequence including five frames (5 sec. length, labeled from 1 to 5). The extracted LMA features related to this sequence are fed to each IB model. The histogram in this figure represents the output for each IB for the seconds 2 to 5.

Fig. 5.12 presents the classification results for some long sequence data. In Fig. 5.12-a) the result for the Indicator model is plotted. As can be seen, the classification result converges after passing a maximum of three frames. The convergence for the Interest model, shown in Fig. 5.12-b), is faster. The reason is because as can be seen in Fig. 5.5-right, this model just depends on the data of the same person independent of previous data.

Fig.5.12-c) presents the Empathy model results. The graph is divided in two parts. The first part corresponds to the first 33 frames and presents the results based on using head-space feature (nodding), and the rest is based on comparison algorithm proposed

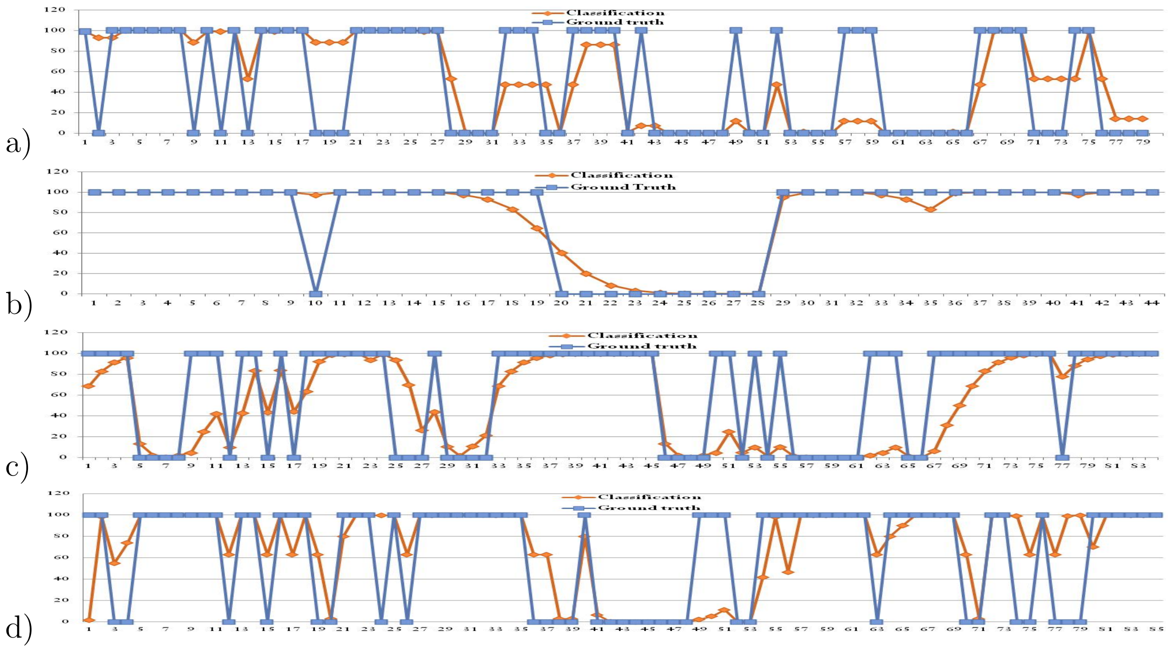


Figure 5.12: Classification results over the time axis for the four IBs states Influent, Active, Mimicry, Inconsistent which are respectively shown in (a), (b), (c) and (d).

	Indicator		Interest		Empathy		Emphasis	
	Inft	Infd	Act.	Pas.	Mim.	Unc.	Con.	Inc.
PCR	72%	70%	93%	90%	80%	74%	88%	53%

Table 5.3: Positive Classification Rate (PCR) results for IBs models

in Alg.1. As can be seen, the first part shows faster convergence but we have slower convergence for the second part. The reason is that the nodding model requires no data from previous time whereas the comparison algorithm needs data from previous moments, which makes the convergence slower.

Fig. 5.12-d) presents the result for Emphasis model. As seen, in most parts the classification result converges to the ground truth. Only a few frames diverge from the ground truth signal (78^{th},79^{th} frame), because the states were changed fast. Table.5.3 summarizes the positive classification results for all IB variables based on the obtained LMA states at every second. In this table, the classifications percentage for the Emphasis state is lower than the other IBs. It is due to the difficulty of interpretation for this behaviour state even for an expert. As can be seen in the presented results, when the observation data changes, it takes a while to converge. We expect to have better results with faster convergence if a higher frequency could be used (currently 1Hz) [KAQ+10].

IBs	IB definition [Pen08]	States [RSAD12b]	LMA [Hut74]
Ind.	More body motions than others	influenced/influent	Effort
Emp.	mimicking and nodding	uncoordinated/mimicry	Space
Int.	Energetic motion	passive/active	Effort
Emf.	Jerky movements	consistent/inconsistent	Effort, Space

Table 5.4: A brief description of IBs with their states, and their relevant LMA components [RSAD12b]

5.4 LMA-based IBs analysis in frequency domain

In this step, we propose using human body motions in frequency domain to estimate the IBs and a SR. In the previous section the IBs were analysed [RSAD12b], but this work explores LMA features on the frequency domain to estimate both IBs and a SR.

5.4.1 Interpersonal primitive and social behaviour

Various social roles in different scenarios were investigated by Alex Pentland, to find relevant connections between features, which could be analysed by computer scientist [Pen08]. To fill the existent big gap between features (voice, body motion) and SRs, a set of IBs was proposed;

$$IB \in \{Indicator, Interest, Empathy, Emphasis\} \tag{5.11}$$

then by combination of the IBs, a number of SRs; Searching, Teaming, Listening and Leading, were described. In this study, we intend to estimate Leading social role by analysing the IBs, which will be estimated by using body parts motion information in frequency domain.

A Bayesian graph in three layers, which was obtained based on our recent work [RSAD12b], is presented in Figure 5.13. LMA parameters of people, which are in a face to face interaction scenario, are located in the first layer. Then IBs will be estimated given those LMA features, and finally in the last layer the SR will be analysed given the IBs. Table 5.4 shows the brief description of the IBs, which were achieved in [RSAD12b].

Based on the Pentland's definition about the SRs, the Leading role described as combination of attention, interest and great focus in thought and purpose [Pen08].

IBs	State
Indicator	high level of influence
Empathy	No relevant
Interest	high activity level
Emphasis	consistency emphasis and rhythm

Table 5.5: IBs states for Leading role [Pen08]

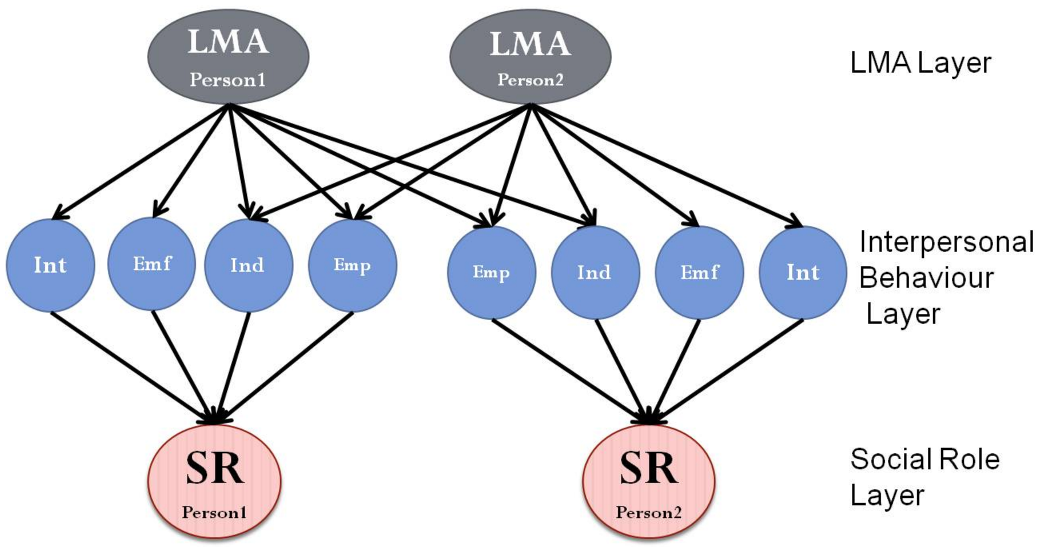


Figure 5.13: proposed BN model of IBs and Leading's social role understanding.

Thus, the IBs states which present the Leading role, can be described as in Table 5.5. As seen, only Empathy did not consider as an effective IB of the mentioned role.

By analysing the Table 5.4 and Table 5.5, Effort component is the most important feature for the Leading role analysis. Thus we consider to use only Effort component as input features to analyse the IBs. Since the Effort-Time component was analysed in frequency domain [KAQ+10], we are able to use the advantages of frequency domain features (as an invariant feature) to reach more high level information.

5.4.2 Social behaviour modelling

Social role is a high level human behaviour information, which consists of different kinds of features related to human being in any society. Based on the obtained results in our recent work [RSAD12b] (modeling the IBs), and Pentland's definition about Leading role which was described before, we attempted to analyse body parts motions in frequency domain to estimate the mentioned SR, with respect to the LMA-Effort property. Thus, to provide the corresponding models, we divided the analysis into a

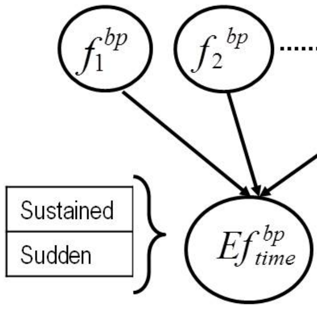


Figure 5.14: Bayesian model for Effort.time sub-component proposed in [KAQ+10]. Ef_{time}^{bp} and f_i^{bp} denote Effort.time variable for bp's body part, and i^{th} frequency-based feature obtained from accelaration signal of bp's body part.

couple of steps; LMA-based human movement and IBs and social role analysis.

5.4.2.1 LMA based human movement analysis

As described before, LMA has several components which were invented to interpret human motions in different domains. Space and Shape were analysed in spatial domain [Ret08, KAQ+11] and Effort in frequency domain [KAQ+11]. Depending on the application, researchers select a number of more representative LMA parameters. This study relies on Effort-Time sub-component which is related to the frequency domain analysis. To estimate Effort-Time, PS signals of different body parts are obtained from the corresponding acceleration signals [KAQ+10]. Figure 5.14 shows the model of the LMA-Effort-Time which contains a couple of states (Sustained/Sudden).

By having the obtained PS signals, the first four coefficients of the signals are collected as input features for the model to estimate the Effort-Time sub-component [KAQ+11]. Eq.5.12 presents the Bayesian rule of the mentioned model (Chapter 2):

$$P\left(Ef_{time}^{bp}|f_1^{bp}f_2^{bp}f_3^{bp}f_4^{bp}\right) = \frac{P(f_1^{bp}|Ef_{time}^{bp})P(f_2^{bp}|Ef_{time}^{bp})P(f_3^{bp}|Ef_{time}^{bp})P(f_4^{bp}|Ef_{time}^{bp})P(Ef_{time}^{bp})}{P(f_1^{bp})P(f_2^{bp})P(f_3^{bp})P(f_4^{bp})} \tag{5.12}$$

where Ef_{time}^{bp} and f_i^{bp} denote to Effort-Time component variable for bp body part motion, and i^{th} coefficient of the PS signal which was obtained from bp body part movement, respectively. The Bayesian equation consists of several parameters as follows;

- $P\left(Ef_{time}^{bp}|f_1^{bp}f_2^{bp}f_3^{bp}f_4^{bp}\right)$, the posterior probability, the probability of Ef_{time}^{bp}

Term(s):	Description
$P\left(Ef_{time}^{bp}\lvert f_1^{bp} f_2^{bp} f_3^{bp} f_4^{bp}\right)$:	posterior probability
$P(f_1^{bp}\lvert Ef_{time}^{bp})$, $P(f_2^{bp}\lvert Ef_{time}^{bp})$, $P(f_3^{bp}\lvert Ef_{time}^{bp})$ and $P(f_4^{bp}\lvert Ef_{time}^{bp})$:	likelihood parameters
$P(Ef_{time}^{bp})$:	prior probability
$P(f_1^{bp})$, $P(f_2^{bp})$, $P(f_3^{bp})$ and $P(f_4^{bp})$:	The normalisation terms

Table 5.6: The movement Bayesian rule variables descriptions

Program / Description / Specification

Pertinent variables :

$Pb \in \{Head, Left\,hand, Right\,hand\}$
$f_{i=1:4}^{p\in\{pb\}} \in \{No, low, Medium, High\}$
$Ef_{time}^{p\in\{pb\}} \in \{Sustained, Sudden\}$

Decomposition :

$$\left(P\left(Ef_{time}^{p}\prod^{i=1:4}\left(f_i^p\right)\right) = \prod^{i=1:4}\left(P\left(f_i^p\right)\right)P\left(\prod^{i=1:4}\left(f_i^p\right)\lvert Ef_{time}^{p}\right)\right)^{p\in\{pb\}}$$

Parametric forms :
$P\left(f_i^p\right)_{i=1:4}^{p\in\{pb\}}$: $uniform$
$P\left(f_{i=1:4}^p\lvert Ef_{time}^p\right)^{p\in\{pb\}}$: $histogram(Matrix)$

Iden.

Ques. : $P\left(Ef_{time}^p\middle\vert \prod^{i=1:4}\left(f_i^p\right)\right)^{p\in\{pb\}}$

Figure 5.15: Bayesian Program for the human movement model.

given f_1^{bp} ,f_2^{bp},f_3^{bp} and f_4^{bp} in the Bayesian rule,

- The likelihood parameters, $P(f_1^{bp}\lvert Ef_{time}^{bp})$, $P(f_2^{bp}\lvert Ef_{time}^{bp})$, $P(f_3^{bp}\lvert Ef_{time}^{bp})$ and $P(f_4^{bp}\lvert Ef_{time}^{bp})$, which are the probability of evidences given the Ef_{time}^{bp} state,

- $P(Ef_{time}^{bp})$, the prior probability, is the probability of Ef_{time}^{bp} before having the evidences,

- The normalisation terms, which are $P(f_1^{bp})$, $P(f_2^{bp})$, $P(f_3^{bp})$ and $P(f_4^{bp})$.

Table 5.6 presents the list of variables with their relevant descriptions. Figure 5.15 presents the BP's correspondent of the proposed human movement model.

Figure 5.16-(a) shows a sample of LMA-Effort-Time signals which are obtained from the PS signals of LLFs as was proposed in Chapter 2.

5.4.2.2 Human interpersonal behaviour and social role analysis

In our previous work [RSAD12b], the LMA components were analysed in each instance (one-second) and were used for IBs estimation. The previous approach was more reliant on the instant evidences, and it just used previous knowledge through DBN rules. However, the process needs longer data (thirty-seconds [Pen08]) to be able to analyse the IBs. Since the time period of performing the IBs is variant and longer than one second, instead of using each instant estimated LMA parameters to analyse IBs [RSAD12b], a sequence of the LMA-Effort-Time data is collected from a sequence of instances (Figure 5.16-(a)).

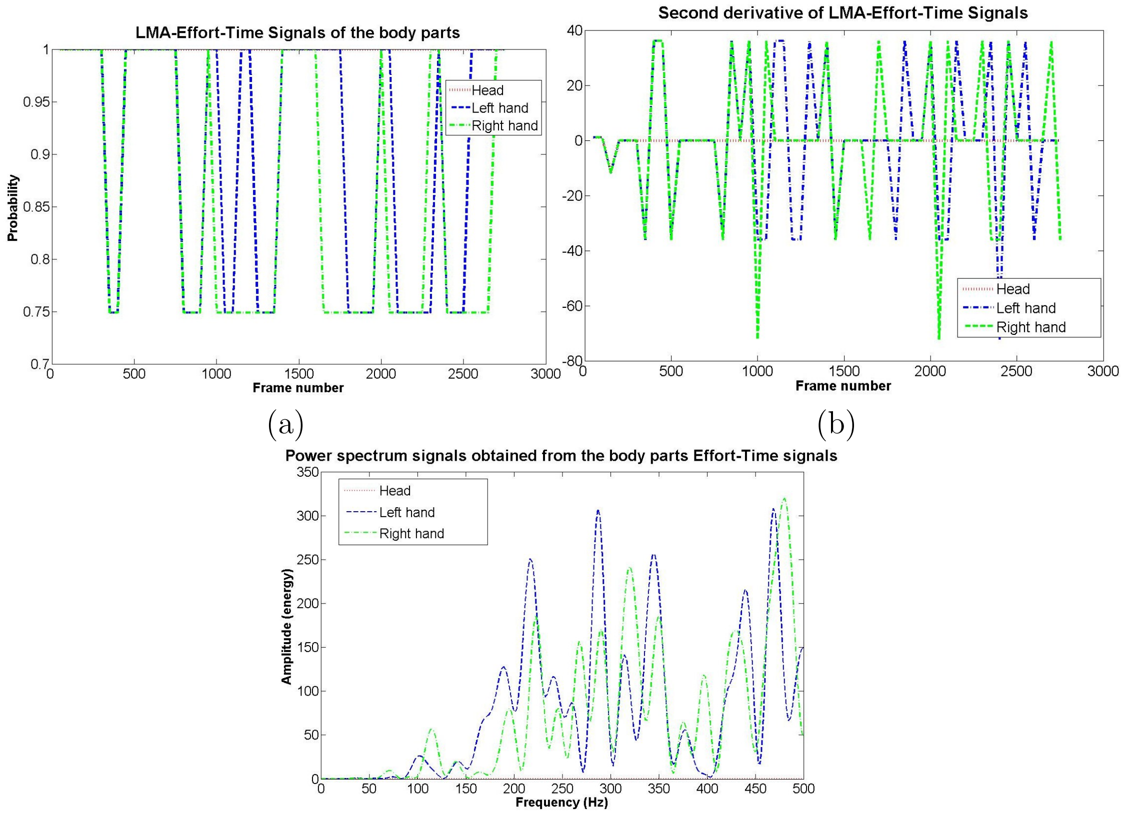
To explore the changes in LMA-Effort-Time signals, the corresponding second derivative of the signals ($\frac{d^2(Ef_{time}^{bp})}{dt^2}$), is calculated (Figure 5.16-(b)). Then PS technique is used to extract frequency based features from the obtained second derivative of the LMA-Effort-Time signals (Figure 5.16-(c)). Then, coefficients of each PS signals are collected. For estimating each IBs, based on the previous work [RSAD12b], the representative body parts are selected, and the needed coefficients from relevant PS signals are collected for feeding our Bayesian model.

- Indicator

It is common that when a person tries to influence others, (s)he usually consumes more energy and move more his/her body parts, e.g. hand in a conversation scenario, to be more representative. PS technique is a proper approach to analyse the consumed energy from performed body parts motion by people. It means that who influences more others, usually has more energy (amplitude) in the PS signals. Thus, Bayesian model of this IB is defined as:

$$P\left(Ind_i|\prod_{h=1:n,i=1:m} {}^aEf_i^h\right) = \frac{P(Ind_i)\prod_{h=1:n,i=1:m} P\left({}^aEf_i^h|Ind_i\right)}{\prod_{h=1:n,i=1:m} P\left({}^aEf_i^h\right)} \tag{5.13}$$

where i, h, n and m denote person index, body parts index, number of body parts and number of involved people, respectively. $P(Ind_i)$,$P\left({}^aEf_i^h\right)$ denote probability of IB of Indicator for person i, and probability of Effort-Time of body part h for person i based on amplitudes (a) of PS signals coefficients, respectively. The Bayesian equation consists of several parameters as can be seen in Table 5.7, which presents the list of variables with their relevant descriptions.



(a) (b)

(c)

Figure 5.16: (a) and (b) show a sample of LMA-Effort-Time signals and their second derivative signals correspondent, respectively. (c) presents PS signals obtained from (b). In (a) presents probability of the sustained state of a person during a conversation scenario. Hands shows some changes, but there is no considerable movement in the body part of head. Then, we produce the second derivative of the (a) which can be seen in (b). (c) shows the energy of signals presented in (b) for each frequency domains.

Figure 5.17 presents the BP's correspondent of the proposed Indicator model.

- Interest

In any social interaction, the activity level of human body is a visible unconscious signal to present human interest and excitement level [Pen08]. Eq.5.14 presents the relations between LMA parameters and Interest IB. The needed features are similar to Indicator, but this one doesn't have connection to other people's behaviour. Thus;

$$P\left(Int_i | \prod_{h=1:n} {}^{a}Ef_i^h\right) = \frac{P\left(Int_i\right) \prod_{h=1:n} P\left({}^{a}Ef_i^h | Int_i\right)}{\prod_{h=1:n} P\left({}^{a}Ef_i^h\right)} \tag{5.14}$$

Term(s):	Description
$P\left(Ind_i \mid \prod_{h=1:n,i=1:m} {}^{a}Ef_i^h\right)$:	posterior probability
$P\left({}^{a}Ef_{i=1:m}^{head} \mid Ind_i\right)$, $P\left({}^{a}Ef_{i=1:m}^{Right-hand} \mid Ind_i\right)$, and $P\left({}^{a}Ef_{i=1:m}^{Left-hand} \mid Ind_i\right)$:	likelihood parameters
$P\left(Ind_i\right)$:	prior probability
$P\left({}^{a}Ef_{i=1:m}^{head}\right)$, $P\left({}^{a}Ef_{i=1:m}^{Right-hand}\right)$, and $P\left({}^{a}Ef_{i=1:m}^{Left-hand}\right)$:	The normalisation terms

Table 5.7: The Indicator Bayesian rule variables descriptions

$$
\text{Program}\begin{cases}
\text{Description}\begin{cases}
\text{Specification}\begin{cases}
\textbf{Pertinent variables}: \\
Pb \in \{Head, Left\,hand, Right\,hand\} \\
Ef_{i=1:2}^{p\in\{pb\}}(a) \in \{No, low, Medium, High\} \\
Ind_{i=1:2} \in \{Influenced, Influent\} \\
\textbf{Decomposition}: \\
\left(P\left(Ind_k \prod_{p\in\{pb\}}^{i=1:2}\left(Ef_i^p(a)\right)\right) = \prod_{p\in\{pb\}}^{i=1:2} P\left(Ef_i^p(a)\right) \prod_{p\in\{pb\}}^{i=1:2} P\left(\left(Ef_i^p(a)\right) \mid Ind_k\right)\right)^{k=1:2} \\
\textbf{Parametric forms}: \\
P\left(Ef_i^p(a)\right)_{i=1:2}^{p\in\{pb\}} : uniform \\
\left(P\left(\left(Ef_i^p(a)\right) \mid Ind_k\right)_{p\in\{pb\}}^{i=1:2,}\right)^{k=1:2} : histogram(Matrix)
\end{cases} \\
\textbf{Iden}.
\end{cases} \\
\textbf{Ques}.: \quad P\left(Ind_k \,\middle|\, \prod_{p\in\{pb\}}^{i=1:2}\left(Ef_i^p(a)\right)\right)^{k=1:2}
\end{cases}
$$

Figure 5.17: Structure of Bayesian program for proposed Indicator model.

Term(s):	Description
$P\left(Int_i \mid \prod_{h=1:n} {}^{a}Ef_i^{h}\right)$:	posterior probability
$P\left({}^{a}Ef_i^{head} \mid Int_i\right)$, $P\left({}^{a}Ef_i^{Right-hand} \mid Int_i\right)$, and $P\left({}^{a}Ef_i^{Left-hand} \mid Int_i\right)$:	likelihood parameters
$P(Int_i)$:	prior probability
$P\left({}^{a}Ef_i^{head}\right)$, $P\left({}^{a}Ef_i^{Right-hand}\right)$, and $P\left({}^{a}Ef_i^{Left-hand}\right)$:	The normalisation terms

Table 5.8: The Interest Bayesian rule variables descriptions

$$
\text{Program}\begin{cases}
\text{Description}\begin{cases}
\text{Specification}\begin{cases}
\textbf{Pertinent variables}: \\
Pb \in \{Head, Left\,hand, Right\,hand\} \\
Ef_{i=1:2}^{p\in\{pb\}}(a) \in \{No, low, Medium, High\} \\
Int_{i=1:2} \in \{Passive, Active\} \\
\textbf{Decomposition}: \\
\left(P\left(Int_i \prod_{p\in\{pb\}} \left(Ef_i^{p}(a)\right)\right) = \prod_{p\in\{pb\}} P\left(Ef_i^{p}(a)\right) \prod_{p\in\{pb\}} P\left(\left(Ef_i^{p}(a)\right) \mid Int_i\right)\right)^{i=1:2} \\
\textbf{Parametric forms}: \\
P\left(Ef_i^{p}(a)\right)_{i=1:2}^{p\in\{pb\}} : uniform \\
P\left(\left(Ef_i^{p}(a)\right) \mid Int_i\right)_{p\in\{pb\}}^{i=1:2} : histogram(Matrix)
\end{cases} \\
\textbf{Iden}.
\end{cases} \\
\textbf{Ques}.: \; P\left(Int_i \middle| \prod_{p\in\{pb\}} \left(Ef_i^{p}(a)\right)\right)^{i=1:2}
\end{cases}
$$

Figure 5.18: Structure of Bayesian program for proposed Interest model.

where Int_i and Ef_i^h denote Interest variable for i^{th}person and Effort component variable based on amplitudes (a) of PS signals coefficients for h^{th} body part of i^{th} person respectively. The Bayesian equation consists of several parameters as can be seen in Table 5.8, which presents the list of terms with their relevant descriptions.

Figure 5.18 presents the BP's correspondent of the proposed Interest model.

- Empathy

This IB is highly connected to the spatial based features [RSAD12b]. However, head nodding signals, which is one of the important mimicry characteristics, can be analysed in frequency domain. Thus, besides the spatial based features (Shape component), PS signal of head Effort-Time data is used to feed the model, and Eq.5.15 presents the relations between LMA parameters and Empathy's IB:

Term(s):	Description
$P\left(Emp_i \mid {}^aEf_i^{head}\, Sp_i(t)\, Sp_j(t-1)\right)$:	posterior probability
$P\left({}^aEf_i^{head} \mid Emp\right)$ and $P\left(Sp_i(t)\, Sp_j(t-1) \mid Emp_i\right)$:	likelihood parameters
$P(Emp_i)$:	prior probability
$P\left({}^aEf_i^{head}\right)$, $P(Sp_i(t))$, and $P\left(Sp_j(t-1)\right)$:	The normalisation terms

Table 5.9: The Empathy Bayesian rule variables descriptions

$$
\text{Program}\begin{cases}
\text{Description}\begin{cases}
\text{Specification}\begin{cases}
\textbf{Pertinent variables}: \\
Ef_{i=1:2}^{head}(a) \in \{No, low, Medium, High\} \\
Sp_{T\in\{t,t-1\}}^{i=1:2} \in \{Rising, Still\, sinking\} \\
Emp_{i=1:2} \in \{Mimicry, Uncoordinated\} \\
\textbf{Decomposition}: \\
\left(P\left(Emp_i\, Ef_i^{head}(a)\, Sp_t^i\, Sp_{t-1}^j\right) = P\left(Ef_i^{head}(a)\right) P\left(Sp_t^i\right) P\left(Sp_{t-1}^j\right)\right)^{i=1:2, j\neq i} \\
\qquad P\left(Ef_i^{head}(a) \mid Emp_i\right) P\left(Sp_t^i\, Sp_{t-1}^j \mid Emp_i\right) \\
\textbf{Parametric forms}: \\
P\left(Ef_i^{head}(a)\right)_{i=1:2} : uniform \\
P\left(Sp_T^i\right)_{T\in\{t,t-1\}}^{i=1:2} : uniform \\
P\left(Ef_i^{head}(a) \mid Emp_i\right)^{i=1:2} : histogram(Matrix) \\
P\left(Sp_t^i\, Sp_{t-1}^j \mid Emp_i\right)^{i=1:2, j\neq i} : histogram(Matrix)
\end{cases} \\
\textbf{Iden}.
\end{cases} \\
\textbf{Ques}.: \quad P\left(Emp_i \,\middle|\, Ef_i^{head}(a)\, Sp_t^i\, Sp_{t-1}^j\right)^{i=1:2, j\neq i}
\end{cases}
$$

Figure 5.19: Structure of Bayesian program for proposed Empathy model.

$$
P\left(Emp_i \mid {}^aEf_i^{head}\, Sp_i(t)\, Sp_j(t-1)\right) = \frac{P(Emp_i)\, P\left({}^aEf_i^{head} \mid Emp_i\right) P\left(Sp_i(t)\, Sp_j(t-1) \mid Emp_i\right)}{P\left({}^aEf_i^{head}\right) P(Sp_i(t))\, P(Sp_j(t-1))} \tag{5.15}
$$

where Emp_i,${}^aEf_i^{head}$ and $Sp_i(t)$ denote Empathy variable for i^{th}person, Effort-Time variable based on amplitudes (a) of PS signals coefficients for head's body part of i^{th} person, and spatial-based features, respectively, were $j \neq i$. For the spacial-based features, we used a property of Shape component in vertical axes, which has triple states; Sinking, Still, Rising. Since, the spatial-based feature extraction isn't this article concern, for more information about it, we refer our recent work [RSAD12b]. The Bayesian equation consists of several parameters, as can be seen in Table 5.9, which presents the list of variables with their relevant descriptions.

Figure 5.19 presents the BP's correspondent of the proposed Empathy model.

Term(s):	Description
$P\left(Emf_i \mid \prod_{h=1:n} {}^{d}Ef_i^h\right)$:	posterior probability
$P\left({}^{d}Ef_i^{head} \mid Emf_i\right)$, $P\left({}^{d}Ef_i^{Right-hand} \mid Emf_i\right)$, and $P\left({}^{d}Ef_i^{Left-hand} \mid Emf_i\right)$:	likelihood parameters
$P(Emf_i)$:	prior probability
$P\left({}^{d}Ef_i^{head}\right)$, $P\left({}^{d}Ef_i^{Right-hand}\right)$, and $P\left({}^{d}Ef_i^{Left-hand}\right)$:	The normalisation terms

Table 5.10: The Emphasis Bayesian rule variables descriptions

- Emphasis

When a person is thinking about different things simultaneously, his/her body parts movements become jerky and inconsistency paced ([VV96] in [Pen08]). It means that depending on the context, people's emphasis can be consistent or variable (inconsistence). To analyse this IB, usually, who is more consistent has energy in a higher frequency domain than who is less. Thus, the frequency domain of first PS signal's peak, is used to feed the model. Eq.5.16 presents the relations between LMA parameters and Emphasis IB in frequency domain.

$$P\left(Emf_i \mid \prod_{h=1:n}\left({}^{d}Ef_i^h\right)\right) = \frac{P(Emf_i)\prod_{h=1:n}\left(P\left({}^{d}Ef_i^h \mid Emf_i\right)\right)}{\prod_{h=1:n} P\left({}^{d}Ef_i^h\right)} \tag{5.16}$$

where Emf_i and ${}^{d}Ef_i^h(t)$ denote Emphasis variable for i^{th}person and Effort component variables based on frequency domain (d) of PS signal's coefficient for h^{th} body part of i^{th} person, respectively. The Bayesian equation consists of several parameters, as can be seen in Table 5.10, which presents the list of variables with their relevant descriptions. Figure 5.20 presents the BP's correspondent of the proposed Emphasis model.

- Social Role of Leading

Social roles present the different types of relationships between people in a community. In [Pen08], various examples proposed to illustrate how the social roles can be estimated by combination of the IBs. Figure 5.13 presents the global BN of the whole proposed system. This is performed by finding the Bayesian rule of the SR, that maximizes Eq.5.17.

$$
\text{Program}\begin{cases}
\text{Description}\begin{cases}
\text{Specification}\begin{cases}
\textbf{Pertinent variables}: \\
Pb \in \{Head, Left\,hand, Right\,hand\} \\
Ef_{i=1:2}^{p\in\{pb\}}(a) \in \{No, low, Medium, High\} \\
Emf_{i=1:2} \in \{Consistance, Variable\} \\
\textbf{Decomposition}: \\
\left(P\left(Emf_i \prod_{p\in\{pb\}} \left(Ef_i^p(d)\right)\right) = \prod_{p\in\{pb\}} P\left(Ef_i^p(d)\right) \prod_{p\in\{pb\}} P\left(\left(Ef_i^p(d)\right)|Emf_i\right)\right)^{i=1:2} \\
\textbf{Parametric forms}: \\
P\left(Ef_i^p(d)\right)_{i=1:2}^{p\in\{pb\}} : uniform \\
P\left(\left(Ef_i^p(d)\right)|Emf_i\right)_{p\in\{pb\}}^{i=1:2} : histogram(Matrix)
\end{cases} \\
\textbf{Iden}.
\end{cases} \\
\textbf{Ques}.: \quad P\left(Emf_i \middle| \prod_{p\in\{pb\}} \left(Ef_i^p(d)\right)\right)^{i=1:2}
\end{cases}
$$

Figure 5.20: Structure of Bayesian program for proposed Emphasis model.

$$
P\left(SR|\prod_{l=1:k} IB_l\right) \tag{5.17}
$$

where l, k, IB_l and SR denote the IB index (for the four IBs), number of IBs, l^{th} IB, and the interesting social role (Leading) variable. Figure 5.21 presents the corresponding BP for the proposed Leading social role model.

Figure 5.22 presents a sample of IBs states during a Leading social role analysis. As can be seen in the figure, the considered leader is highly interested and performs many body motions to be influent for others. (S)He moves often and consistently, with no concerns for the other person movements (low empathy). Thus, in the diagram, low probability of Indicator means that the person is in influenced, otherwise in influent state, and low probability of Empathy means the person is in uncoordinated, otherwise in mimicry state. Low probability of Interest means the person is in passive, otherwise in active state, and low probability of Emphasis means the person is in consistent, otherwise in inconsistent or variable state.

In this study, we tested the model on peer interaction scenarios. For modelling the variables by having more persons, estimation of LMA components for all people can be applied individually, in parallel approach. For Interest and Emphasis variables, since they are not related to others, can be estimated individually, however, Empathy and Indicator variables need other's knowledge, which increase the complexity of the modeling process. Thus, to simplify the process, influence model can be used [DLC+07].

$$
\text{Program}\begin{cases}
\text{Description}\begin{cases}
\text{Specification}\begin{cases}
\textbf{Pertinent variables}: \\
IB \in \{Indicator, Interest, Empathy, Emphasis\} \\
Indicator \in \{Influenced, Influent\} \\
Interest \in \{Passive, Active\} \\
Empathy \in \{Mimicry, Uncoordinated\} \\
Emphasis \in \{Consistance, Variable\} \\
SR \in \{Leading, No-Leading\} \\
\textbf{Decomposition}: \\
\left(P\left(SR_i \prod_{B\in\{IB\}} (B_i)\right) = P(SR_i) \prod_{B\in\{IB\}} P(B_i|SR_i)\right)^{i=1:2} \\
\textbf{Parametric forms}: \\
\left(B_{i=1,2}^{B\in\{IB\}}\right) : Estimated\ from\ previous\ step \\
P(B_i|SR_i)_{B\in\{IB\}}^{i=1:2} : histogram(Matrix)
\end{cases} \\
\textbf{Iden}.
\end{cases} \\
\textbf{Ques}.: \quad P\left(SR_i \middle| \prod_{B\in\{IB\}} (B_i)\right)^{i=1:2}
\end{cases}
$$

Figure 5.21: Structure of Bayesian program for proposed Leading-role model.

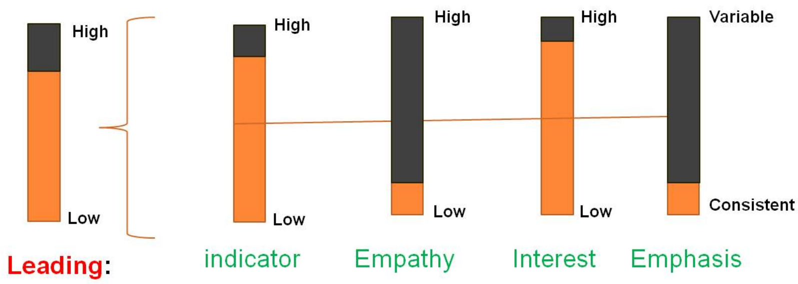


Figure 5.22: A probabilistic-based social role analysis.

Figure 5.23: The motion tracker suit used by an actor and its correspondent skeleton.

5.4.3 Experiments

For collecting data, a motion tracker suit, which gives body parts 3D position, is used (Figure 5.23). Different people put on the suit and after doing the calibration process (for the suit's sensors), people performed arbitrary conversation. Outputs of the suit's sensors as a XML file for each record series, were stored. The mentioned experiment process was performed by ten different people. Each person in different acquisition times participated in an arbitrary conversation with another person. Afterwards, the states of the IBs and the SR of each person, were labeled for the learning and classification process.

To extract features, sliding-window approach is used [KAQ+11]. The window size was defined as one second, which shifts by a half of the window size [KAQ+11]. The frame rate is 120 Hz. More than one hundred sequences (each sequence contains less than 30 seconds of the body motions data) are collected, which are performed in different IBs and SR (as leader or not-leader) states.

For the classification step, each sequence is analysed to extract LMA signals. Then frequency based features from the second derivative of LMA signals are extracted as was explained in previous section. Those frequency-based features are used to estimate the IBs and the SR states probability. Half of the data was used for the learning, and the others for the test process. Figure 5.24, Figure 5.26 and Figure 5.28 present three scenarios, which consist of different levels of signal processing from the LLFs to the SR estimation;

- Scenario A: a couple of people, a supervisor and a student are in a formal conversation, and the supervisor is guiding the student about his work. Figure 5.24-(a) presents the skeletons of the people during the conversation. The right person is the supervisor.

- Scenario B: two colleagues are talking about a project. One of them, a PhD student, is describing the project, which is related to his PhD study, to a new researcher of the laboratory. Figure 5.26-(a) presents a sequence of the people's movements and the corresponding skeletons during the conversation. The person on the right is the PhD student.

- Scenario C: two colleagues are arguing. The person on the right is attacking the person on the left about what (s)he did, and the person on the left is defending him/herself. Figure 5.28-(a) presents a sequence of the people's movements and corresponding skeletons during the conversation.

In each of those figures, we present the corresponding signal processing in different steps;

- First step: shows the scenario of conversation by a sequence of images of body movements performed by the people involved, labeled as (a).

- Second step: a sequence of PS signals obtained from the body parts acceleration signals during the conversation, which can be seen on the second row of the figures, for each person, labeled (b) and (c). The mentioned diagrams present the changes of PS signals of each body part during conversations.

- Third step: presents a sequence Effort-Time results provided from the sequence PS signals, obtained in the previous step (estimated by the Eq. 5.12). The sequence Effort-Time results is presented in the third row of the figures, for each of the involved persons, labeled (d) and (e).

- Fourth step: the PS signal obtained from the corresponding second derivative of Effort-Time signal for each body part, is presented in the fourth row of the figures, labeled as (f) and (g).

- Fifth step: the obtained probabilities of IBs and the SR are presented in the last row, as labeled (h) and (i), each of them belongs to one of the involved people.

Results discussion:

Figure 5.24-(a) presents a sequence of the scenario A. Figure 5.24-(b) and Figure 5.24-(c) present the right and left person PS signals in every 50 frames (50/120 second time shift), respectively. As we expected, in the Figure 5.24-(b), we can see more energy in the supervisor's body parts PS signals, than in the student ones which can be seen in Figure 5.24-(c) (about 10 times).

The sequence of Effort-Time data from the person on the right, shows high variation in the corresponding signals of the supervisor's hands (Figure5.24-(d)), than the student's ones, which is presented in Figure 5.24-(e). It shows that the supervisor wants to influence the student, and the student presents a few motions to show that he is interested in the conversation.

Then, Figure 5.24-(f) and Figure 5.24-(g) present that in the estimated PS signals from the supervisor's corresponding Effort-Time signals, has several high amplitudes (peaks) in various frequency domains for his hands, but for the student it is not the same.

Figure 5.24-(h) and Figure 5.24-(i) present the estimated four IBs and SR probabilities. As can be seen in the Figure 5.24-(f) and Figure 5.24-(g), the coefficients of the right person's PS signals has a higher amplitude than the left person. Thus, the right person has more probabilities in Influent and Active states than the other. Since there are no head nodding signals, we don't see a considerable difference between the couple's Mimicry states.

The first coefficient of the person on the right appears in a lower frequency domain than the other one, thus, we see less probability distribution in consistence state of Emphasis variable for the right person than the left one. However, several coefficients in higher frequency domains for the right person can be seen, which illustrate that the Variable state's probability should be lower than we obtained. In the future, that information can be considered to improve the model. It can be inferred that the person who is in the Leading role state, consumes more energy in his body motions, however, the probability of Variable state of the person, is much higher than the other, which makes sense (the one has more energy in his/her body movement, usually has more chance to lose his/her consistency).

Figure 5.25 shows second by second the conversation scenario A. The number of each frame is shown on the corresponding skeleton images. Between each two se-

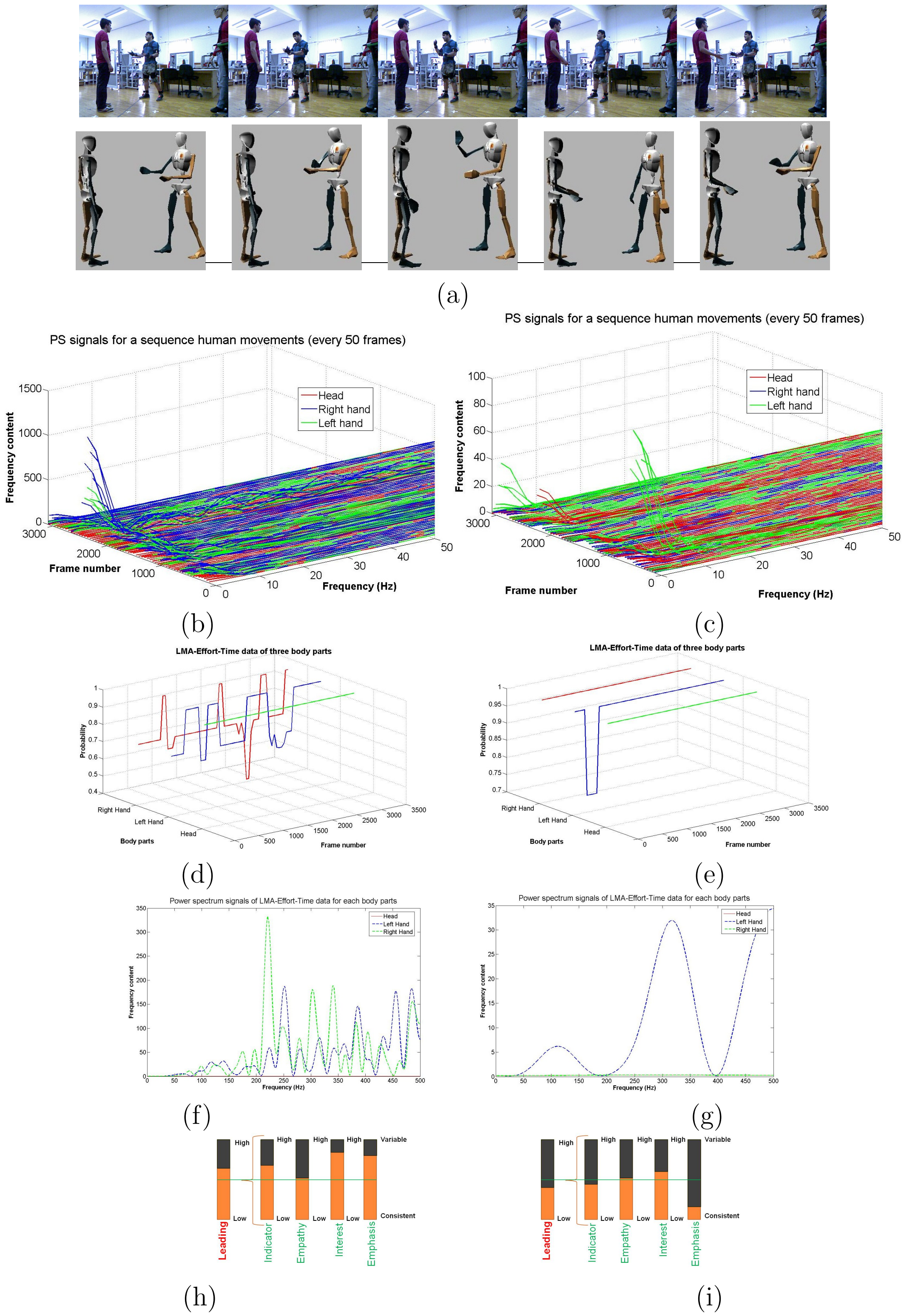


Figure 5.24: A sample of data processing for a sequence of a couple of persons movements (through the proposed scenario A), is presented (skeletons and images of the people can be seen in (a)) . left column belongs to the right person (who is the leader) and the right column belongs to the other. (b) and (c) show PS signals of the human body parts acceleration for every 1/2.4 second. (d) and (e) present the obtained LMA-Effort-Time probability signals of each body parts of the people (1:sustained state/ 0:sudden state). (f) and (g) show the PS signals which are obtained from (d) and (e). (h) and (i) present the outputs of model which are the IBs and the SR probability for the left and right persons, respectively.

quence frames, two diagrams show the PS signals obtained from LMA data of the people, and one comparative diagram for their IBs and SR probabilities with order of Indicator(influent), Interest(active), Empathy(mimicry), Emphasis(consistent) and SR(Leading). It should be mentioned that in each step the results are estimated by using all obtained data until the current frame. It can be seen that during the conversation, the person on the right is more active to influence the other, and one on the left is following the right person by showing some body motion as a good listener. Thus, the most probability for leading state always belongs to the person on the right, indicated with blue colors.

In Figure 5.26 another example of a conversation scenario (scenario B) can be seen, with a sequence of images of the people's movements and their corresponding skeletons. In this trial, the left person shows low interest in the conversation, and performed smooth body movements during the conversation, as can be realized on the obtained corresponding LMA signals. The Effort.Time signals for the right person shows that the person is very active to influence the other person. Thus, it caused the high probability Indicator's IB and the Leading SR, which shows the meaningful relation between Indicator's IB and the Leading role. From the obtained results, we can understand that through those presented conversations there are no considerable signals related to the Empathy IB (head nodding and mimicry movements).

Figure 5.27 shows second by second the conversation scenario B. The number of each frame is shown on the corresponding skeleton images. Between each two sequence frames, two diagrams show the PS signals obtained from LMA data of the people, and one comparative diagram for their IBs and SR probabilities with order of Indicator(influent), Interest(active), Empathy(mimicry), Emphasis(consistent) and SR(Leading). It should be mentioned that in each step the results are estimated by using all obtained data until the current frame. It can be seen that during the conversation, one of the persons is not very interested in the conversation subject. Thus, the most probability for leading state always belongs to the person on the right, indicated by blue colors.

In Figure 5.28 other types of conversation scenario (scenario C) can be seen, with a sequence of images and corresponding skeletons of the people. In this conversation, both persons are highly engaged in the conversation, and trying to convince the other that they are right. The left person is in a more defensive position and moves his/her body parts more smoothly than the other. PS signals and LMA data of both persons

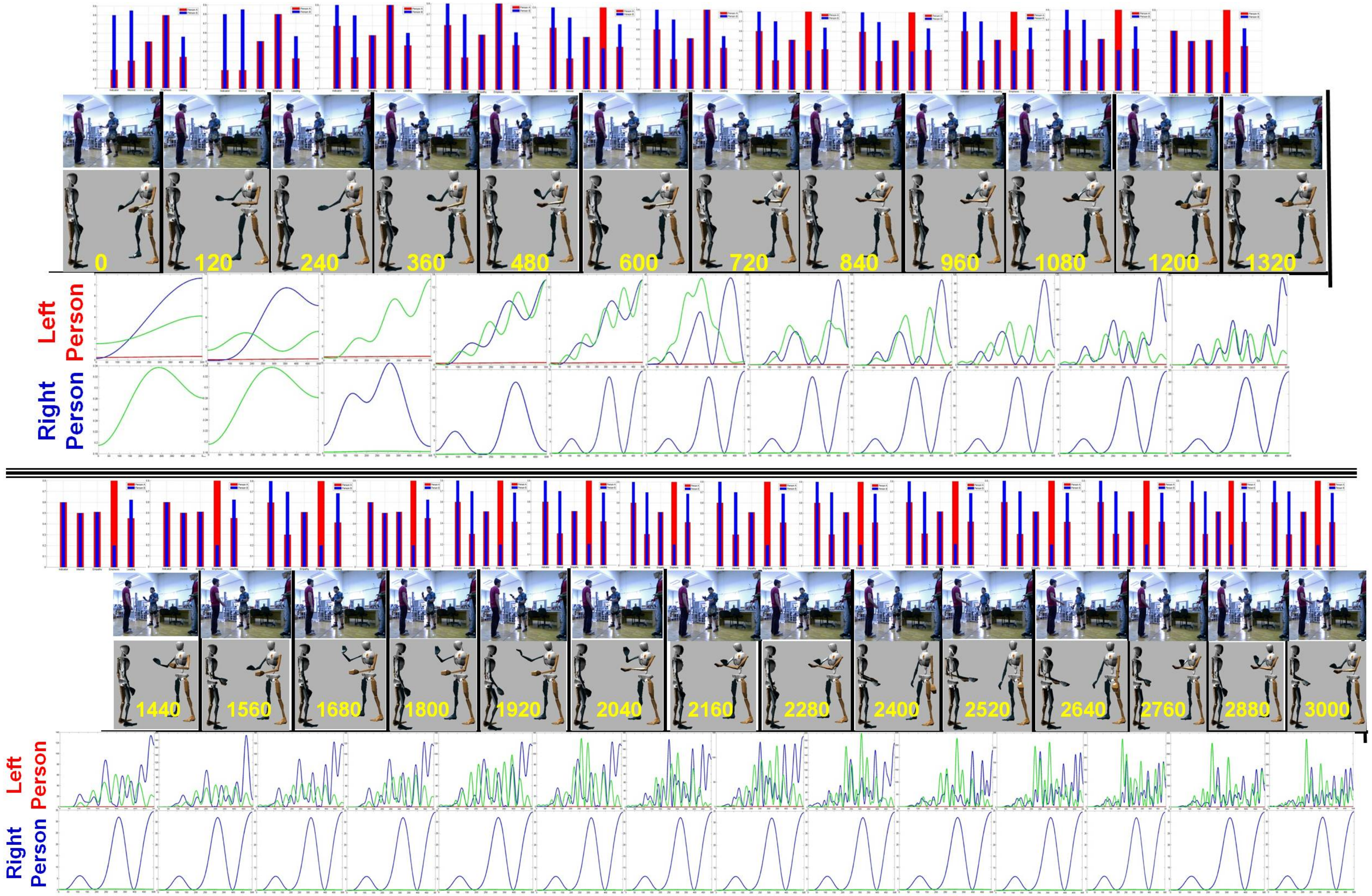


Figure 5.25: A sequence (second by second) process of the conversation scenario A, with both persons PS signal analysis and their corresponding IBs and SR probabilities through a comparative diagram in the each second.

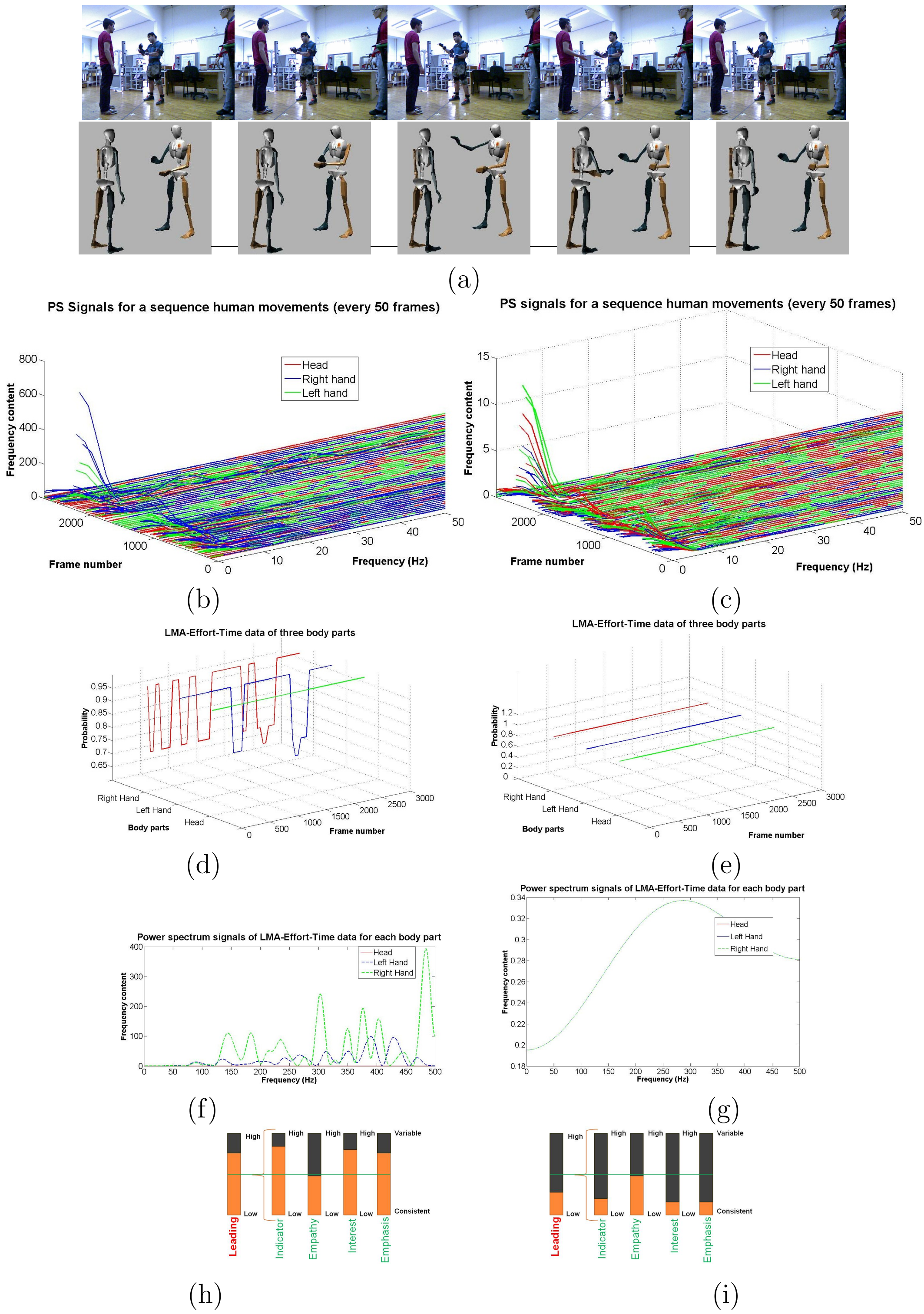


Figure 5.26: A sample of data processing for a sequence of a couple of people movements (through the proposed scenario B) (part (a)) is presented. left column belongs to the right person (as a leader) and the right column belongs to the another. (b) and (c) show PS signals of human body parts acceleration for every 1/2.4 second. (d) and (e) present the obtained LMA-Effort-Time probability signals of each body parts of the people (1:sustained state/0:sudden state). (f) and (g) show the PS signals which are obtained from (d) and (e). (h) and (i) present the outputs of model which are the IBs and the SR probability for the right and left persons, respectively.

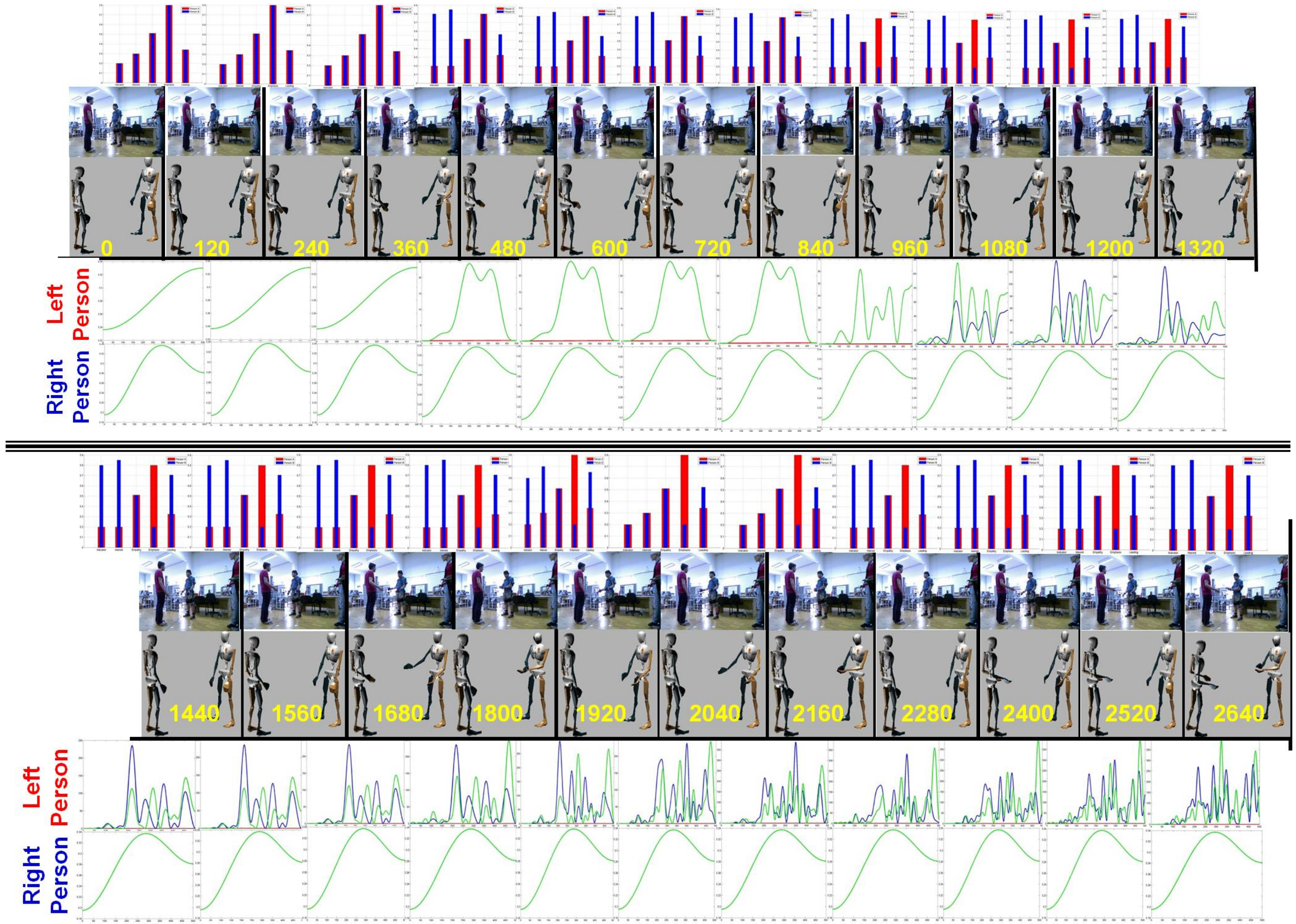


Figure 5.27: A sequence (second by second) process of the conversation scenario B, with both persons PS signal analysis and their corresponding IBs and SR probabilities through a comparative diagram in the each second.

IBs & SR	Indicator		Interest		Empathy		Emphasis		Leading Role	
State	Inft.	Infd.	Act.	Pas.	Mim.	Unc.	Con.	Inc.	Leading	No-Leading
PCR	78%	95%	98%	95%	71%	74%	97%	98%	0.93%	0.88%

Table 5.11: Positive Classification Rate (PCR) results for the IBs and the SR models

show that they are performing body motions to influence the other, but the right person, presented in the left column, performed more energetic movements, which can be seen in Figure 5.28-(c) and Figure 5.28-(e). The PS signals of LMA.Effort.Time signals for the right person, show higher coefficients in lower frequency domains (even for his head). It means that he is more active to influence another person. Thus, we see high probability Indicator's IB and the Leading SR for the right person, however, the left person signals shows that (s)he is highly engaged in the conversation, but (s)he hasn't succeeded in influencing the other.

Figure 5.29 shows second by second the conversation scenario C. The number of each frame is shown on the corresponding skeleton images. Between each two sequence frames, two diagrams show the PS signals obtained from LMA data of the people, and one comparative diagram for their IBs and SR probabilities with order of Indicator(influent), Interest(active), Empathy(mimicry), Emphasis(consistent) and SR(Leading). It should be mentioned that in each step the results are estimated by using all obtained data until the current frame. It can be seen that during the conversation, the higher probability of each IBs and leading role between people are changed. For instance; during 840 to 960 frames the leading state is changed from the right person to the left one, indicated by blue and red colors respectively. However, in overall result, the right person is presented with a more leading role state than the other.

Table 5.11 presents the overall results for all sequence performed scenarios. The results prove that by using frequency-based features in LMA space, the accuracy of IBs recognitions are more precise, than the previous approach [RSAD12b] (from about 77% to about 88%). We should mention that in the previous approach the analysis results were based on one second window slide data, while, the obtained results in this work are analysed based on the whole data of each sequence (less than 30 seconds). Thus using longer data, improves the accuracy of the results, as it was mentioned in [Pen08], which concluded that the adequate time for such an analysis is thirty seconds. It should be mentioned that this study concerns only body motion based IB analysis, and our annotation and the learning process were just based on the body motion features. In the future, the adequate duration segmentation for the proposed model

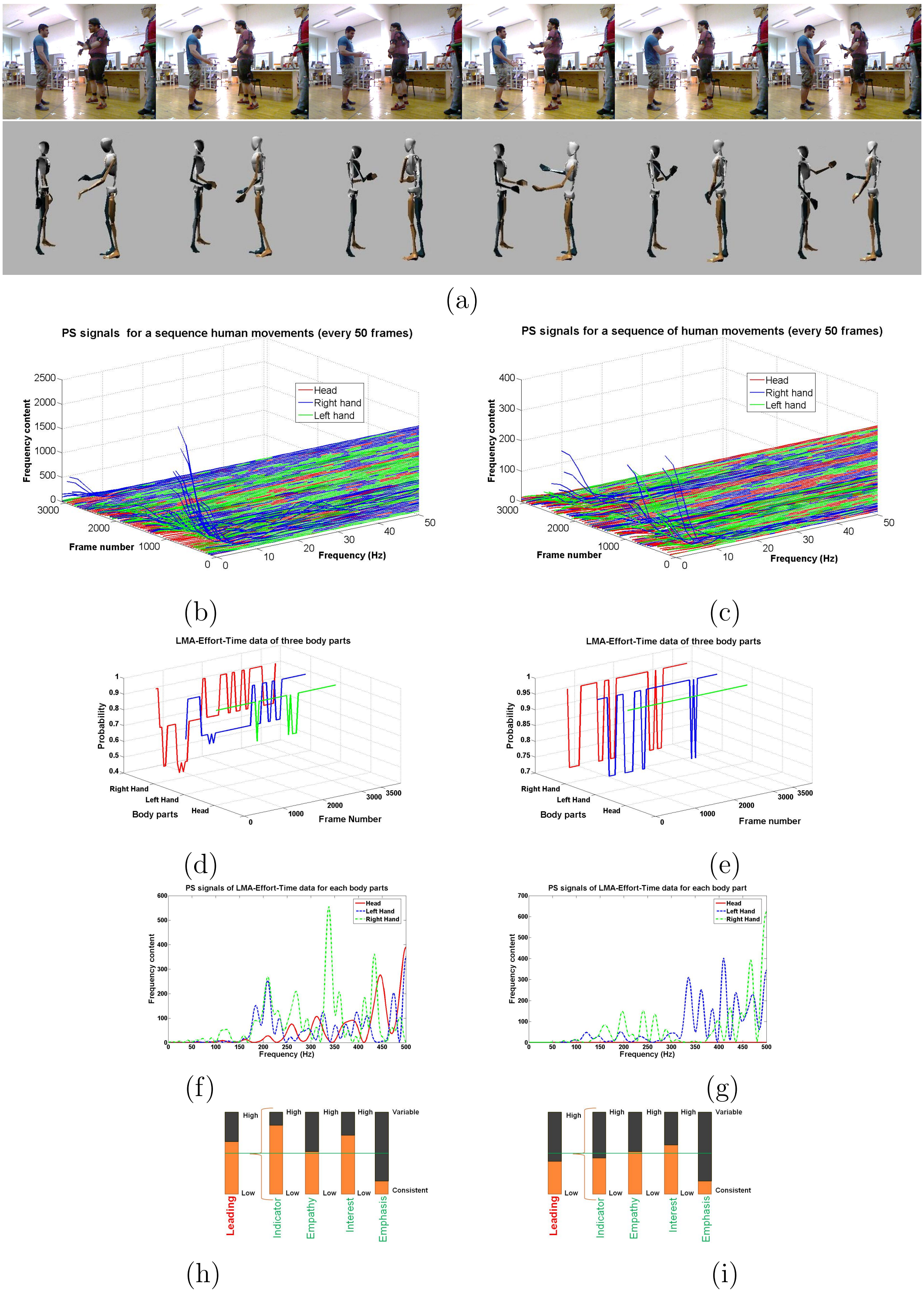


Figure 5.28: A sample of data processing for a sequence of a couple of people's movements and their corresponding skeletons (through the proposed scenario C) (part (a)) are presented. The left column belongs to the right person (as a leader) and the right column belongs to the another. (b) and (c) show PS signals of human body parts acceleration for every 1/2.4 second. (d) and (e) present the obtained LMA-Effort-Time probability signals of each body parts of the people (1:sustained state/0:sudden state). (f) and (g) show the PS signals which are obtained from (d) and (e). (h) and (i) present the outputs of model which are the IBs and the SR probability for the right and left persons, respectively.

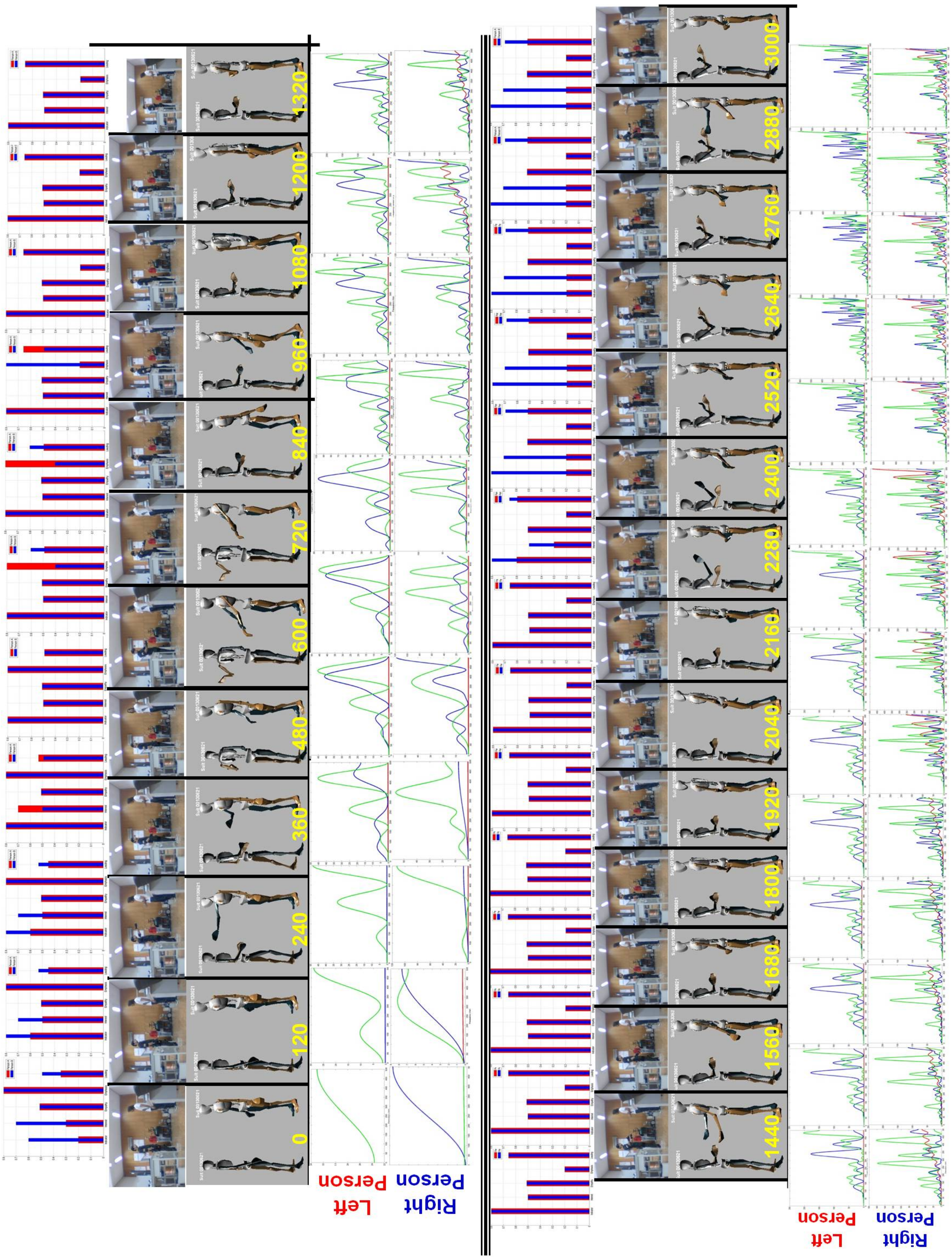


Figure 5.29: A sequence (second by second) process of the conversation scenario C, with both persons PS signal analysis and their corresponding IBs and SR probabilities through a comparative diagram in the each second.

can be investigated.

5.5 Discussion

The body part motion-based IBs are analysed in a couple of different approaches. In the first proposed approach, we used a common approach for learning and recognition process. We proposed a sliding window approach to analyse LLFs to estimate LMAs and then IBs on each sliding window. By shifting the slide window we apply the same process on the new slide, to have new LMA and IBs results by using the obtained previous knowledge. Thus, the confusion table was obtained based on all obtained results for those sliding windows process. However, social scientists claim that some seconds are needed to understand IBs properly. To follow the mentioned role, another approach was proposed. We still use the sliding window approach to obtain LMA parameters, but we don't go directly to analyse IBs. A sequence of sliding windows was selected to obtain their LMA data. Then, based on obtained LMA data which we call LMA signals, we explore frequency properties to estimate IBs. It means that in this approach we used several sliding windows information, which contain some seconds of body motion information, to estimate IBs. Then, the defined SR was estimated based on those IBs. As can be seen in the obtained results accuracy, a magnificent improvement can be seen. In the first step, we have every second IBs states for each person, but in the second approach we have more reliable results.

5.6 Conclusion

In this study, a new approach to parameterize human Interpersonal Behavior (IB) using body motion description (LMA) evidence was proposed. To find the dependencies between body part motions-based features and the IBs, we inspired ourselves in the definitions of social signal by Alex Pentland and the human motion descriptor of Rudolph Laban. We used Bayesian Network (BN) to define our models and a histogram approach to perform supervised learning. The results are encouraging, and motivate us to further explore this work. Then, based on the Pentland's definition and our previous work, we realized that most of the IBs are highly connected to Effort components. Then, we attempted to explore through obtained Effort signals in frequency domain to find more reliable features to estimate the IBs and the Leading role. The existent dependencies are implemented by using BN approach. The results present the capability of the frequency based features to analyse those IBS and the social role. To improve the model, and using both approaches advantages, it is possible to merge those approaches by having another sliding window analysis on the LMA level (such as LLF level). We intend to extend this work to analyse different SRs based on the other tracker data such as kinect-based or camera network based.

Chapter 6

Overall conclusions and future work

In this thesis study, body motion-based human behaviour was investigated in a bottom-up strategy, as can be seen in Fig.1.1. A motion tracker suit was used to collect different class of human activities data. The data was the 3D position, velocity, acceleration, etc. of different body parts which were stored as XML files. However, we obtained multi-modal based dataset from PROMETHEUS European Project, and we applied it in some of our related experiment tests [KAM$^+$11].

Human movement analysis is a very complex problem because of the existing several dependencies and high flexibility between different human body parts and scene context. Thus, to simplify the complexity of the system and fill the existing big gap between LLFs and human activities, a human movement analysis system was used. Laban Movement Analysis (LMA) is the system which provides several components, which contain the needed features, to interpret any human movements in a scene, by choreography scientists. Meanwhile we could use the LMA semantics to interpret human movements in different level of analysis.

For implementing the idea, Bayesian approaches have been used. A Bayesian approach presents many advantages on using prior knowledge and modeling the dynamic dependencies between parameters of object states. In the related fields, this approach is popular and researchers have been keen on applying it. There are several reasons to use the Bayesian approach in the mentioned application: Bayes theorem is valid in all common interpretations of probability, can represent and solve decision problems under uncertainty, it is a common approach to predict, an explicit approach to follow states, does not need large amounts of data to be trained, and it is able to fuse different types of data in one model.

The first level of analysis was estimating LMA components by obtained LLFs. The contribution of this study in this step, was exploring through frequency domain features to estimate Effort property of LMA. Effort component was provided to explain how human consume energy during his/her body part motions. Thus, FFT and Power Spectrum (PS) techniques were used on the acceleration signals of each body parts, to extract the needed invariant coefficients and features for estimating Effort.Time subcomponent [KAQ+10]. Those acceleration signals obtained in body-centred system, and the reference coordinate was located in the body centre (independent of the scene). Meanwhile, to analyse individual actions, we used another LMA component in spatial domain "Shape" to complement Effort component which was obtained in frequency domain [KAQ+11].

In the next step, we attempt to model the Relationship component of LMA, which contains all possible relations between each body part motion with other body parts, outside stationary objects and another person, to estimate different human context-based activities [RD13a], such as human-object and human-human interaction. To improve the model, context-based knowledge was also used to analyse human activities [RD13b].

Since the human movement is variant in temporal and spatial domains, several false detections usually appear during some of the classification process. The false detections usually happen in the boundary of two different movements during a sequence of human movements. Thus, an adaptive sliding window approach, which used the previous estimation entropy as a feedback to the next analysis, was proposed to deal with such a common problem in human movement understanding.

The proposed framework allowed us to explore in more complex human behaviour in social context. Inspired by of Alex Pentland investigation [Pen08], four basic Inter-personal Behaviours (IBs) were analysed based on the LMA components [RSAD12a]. To improve the obtained IBs recognition accuracy, frequency-based analysis in LMA level was proposed to estimate not only the IBs, but also a social role "Leading", which could be obtained by those obtained IBs.

We agree with Aggarwal and Ryoo's opinion, as they mentioned in [AR11]; " In the near future, hierarchical approaches together with strong action level detectors will be explored for reliable recognition of complex activities". As they have covered in [AR11]," hierarchical approaches have their advantages in recognition of high-level activities performed by multiple persons, and they must be explored further in the future to support demands from surveillance systems and other applications".

We believe that the obtained framework has the capability to be extended to different complex applications and scenarios. We intend to explore more in social context scenarios and robotic-based applications, which are going to be the near future interesting challenges in the related communities. In this work, we did not study 3D body part pose estimation problems from image-based data, because it was out of scope of this study. Meanwhile, we realized that 3D body part pose estimation from image-based data is still an active challenge in the relevant communities. However, by using some devices such as Kinect, we are quietly confident that the problem will be solved soon. Based on the reason, we are preparing a global dataset of different human activities using both motion trackers and kinect. Then, we will continue this research based on the new dataset.

Bibliography

[ACLS94] J.K. Aggarwal, Q. Cai, W. Liao, and B. Sabata. Articulated and elastic non-rigid motion: a review. In In Proc. IEEE Workshop on Motion of Non-Rigid and Articulated Objects, pages 2–14, 1994. Proceedings of the IEEE Workshop on Motion of Non-Rigid and Articulated Objects.

[AFKD10] H. Aliakbarpour, J. F. Ferreira, K. Khoshhal, and J. Dias. A novel framework for data registration and data fusion in presence of multi-modal sensors. In Emerging Trends in Technological Innovation, volume 314/2010 of IFIP Advances in Information and Communication Technology, pages 308–315. Springer Boston, 2010 edition, February 2010.

[AKA+02] Jan Allbeck, Karin Kipper, Charles Adams, William Schuler, Elena Zoubanova, Norman I. Badler, Martha Palmer, and Aravind K. Joshi. Acumen: amplifying control and understanding of multiple entities. In International Conference on Autonomous Agents, pages 191–198, Bologna, Italy, 2002. Association for Computing Machinery.

[AM09] T. Abbas and B.A. MacDonald. Robust trajectory segmentation for programming by demonstration. In Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on, pages 1204 –1209, 27 2009-oct. 2 2009.

[AR11] J.K. Aggarwal and M.S. Ryoo. Human activity analysis: A review. ACM Comput. Surv., 43(3):16:1–16:43, April 2011.

[AWSR05] Dejan Arsic, Frank Wallhoff, Bjorn Schuller, and Gerhard Rigoll. Video based online behaviordetection usingprobabilistic multi-stream fusion. IEEE International conference on image processing, 2:606–609, 2005.

[Bea06] Faisal I. Bashir and et al. View-invariant motion trajectory-based activity classification and recognition, 2006.

[BI98] A.F. Bobick and Y.A. Ivanov. Action recognition using probabilistic parsing. In Computer Vision and Pattern Recognition, 1998. Proceedings. 1998 IEEE Computer Society Conference on, pages 196 –202, jun 1998.

[BL80] I. Bartenieff and D. Lewis. Body Movement: Coping with the Environment. Gordon and Breach Science, New York, 1980. ISBN: 978-0677055008.

[BN10] P. Banerjee and R. Nevatia. Dynamics based trajectory segmentation for uav videos. In Advanced Video and Signal Based Surveillance (AVSS), 2010 Seventh IEEE International Conference on, pages 345 –352, 29 2010-sept. 1 2010.

[Bob97] A.F. Bobick. Movement, activity and action: the role of knowledge in the perception of motion. Philosophical Transactions of the Royal Society B: Biological Sciences, 352(1358):1257–1265, 1997.

[Bor10] James Borg. Body language : 7 easy lessons to master the silent language. FT Press, 2010. ISBN: 978-0137002603.

[BPW93] Norman I. Badler, Cary B. Phillips, and Bonnie Lynn Webber. Simulating humans: computer graphics animation and control. Oxford University Press, Inc., New York, NY, USA, 1993. ISBN: 978-0195073591.

[Bre97] C. Bregler. Learning and recognizing human dynamics in video sequences. San Juan, Puerto Rico, 1997. Conference on Computer Vision and Pattern Recognition.

[BSP+04] Jernej Barbič, Alla Safonova, Jia-Yu Pan, Christos Faloutsos, Jessica K. Hodgins, and Nancy S. Pollard. Segmenting motion capture data into distinct behaviors. In Proceedings of Graphics Interface 2004, GI '04, pages 185–194, School of Computer Science, University of Waterloo, Waterloo, Ontario, Canada, 2004. Canadian Human-Computer Communications Society.

[CCK04] Fangxiang Cheng, William Christmas, and Josef Kittler. Periodic human motion description for sports video databases. In Proceedings of the 17th International Conference on Pattern Recognition (ICPR04), 2004.

[CCZB00] Diane Chi, Monica Costa, Liwei Zhao, and Norman Badler. The emote model for effort and shape. In SIGGRAPH '00: Proceedings of the 27th annual conference on Computer graphics and interactive techniques, pages 173 – 182, New York, NY, USA, 2000. Proceedings of the 27th annual conference on Computer graphics and interactive techniques, ACM Press/Addison-Wesley Publishing Co.

[CGV+11] Min Chen, Sergio Gonzalez, Athanasios Vasilakos, Huasong Cao, and VictorC.M. Leung. Body area networks: A survey. Mobile Networks and Applications, 16(2):171–193, 2011.

[CL03] Isaac COHEN and Hongxia LI. Inference of human postures by classification of 3d human body shape. In IEEE International Workshop on Analysis and Modeling of Faces and Gestures, ICCV 2003, pages 74–81, 2003.

[CRG+05] A. John Campbell, M. Clare Robertson, Steven J. La Grow, Ngaire M. Kerse, Gordon F. Sanderson, Robert J. Jacobs, Dianne M. Sharp, and Leigh A. Hale. Randomised controlled trial of prevention of falls in people aged more than 75 with severe visual impairment: the vip trial. BMJ, 331:817, 2005.

[Cro00] Michael Cross. Chapter 6 : Power Spectrum, Chaos course on the web, 2000.

[CS95] C. Cedras and M. Shah. Motion-based recognition: a survey. Image Vision Comput, 13:129–155, 1995.

[Daw02] Mark Ruane Dawson. Gait recognition. Technical report, Department of Computing Imperial College of Science, Technology & Medicine London, June 2002.

[DBM03] Julien Diard, Pierre Bessiere, and Emmanuel Mazer. A survey of probabilistic models, using the bayesian programming methodology as a unifying framework. In In The Second International Conference on Computational Intelligence, Robotics and Autonomous Systems (CIRAS 2003), 2003.

[Dev06] Ruth Devlaeminck. Human motion tracking with multiple cameras using a probabilistic framework for posture estimation. Master's thesis, Purdue University, August 2006.

[DFL+12] Vincent Delaitre, David F. Fouhey, Ivan Laptev, Josef Sivic, Abhinav Gupta, and Alexei A. Efros. Scene semantics from long-term observation of people. In Proceedings of the 12th European conference on Computer Vision - Volume Part VI, ECCV'12, pages 284–298, Berlin, Heidelberg, 2012. Springer-Verlag.

[DLC+07] Wen Dong, Bruno Lepri, Alessandro Cappelletti, Alex Sandy Pentland, Fabio Pianesi, and Massimo Zancanaro. Using the influence model to recognize functional roles in meetings. In Proceedings of the 9th international conference on Multimodal interfaces, ICMI '07, pages 271–278, New York, NY, USA, 2007. ACM.

[dLS02] Rocio Diaz de Leon and Luis Enrique Sucar. Continuous activity recognition with missing data. In Proceedings of the 16th International conference on pattern recognition, volume 1, pages 439–442, 2002.

[DP93] T. Darrell and A. Pentland. Space-time gestures. In Computer Vision and Pattern Recognition, 1993. Proceedings CVPR '93., 1993 IEEE Computer Society Conference on, pages 335 –340, jun 1993.

[Eng06] ISA N. Engleberg. Working in Groups: Communication Principles and Strategies. Houghton Mifflin Company, 2006. ISBN: 978-0618677429.

[ENVRZ09] Magy Seif El-Nasr, Athanasios V. Vasilakos, C. Rao, and Joseph A. Zupko. Dynamic intelligent lighting for directing visual attention in interactive 3-d scenes. IEEE Trans. Comput. Intellig. and AI in Games, 1(2):145–153, 2009.

[FC05] Zhou Feng and Tat-Jen Cham. Video-based human action classi.cation with ambiguous correspondences. In Computer Vision and Pattern Recognition - Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on, page 82, june 2005.

[Fis36] R. A. Fisher. The use of multiple measurements in taxonomic problems. Annals of Eugenics, 7(7):179–188, 1936.

[For10] Donelson R. Forsyth. Group Dynamics. Wadsworth/Cengage, Belmont, CA, 2010. ISBN: 978-0495599524.

[FW06] A. Foroud and I.Q. Whishaw. Changes in the kinematic structure and non-kinematic features of movements during skilled reaching after stroke: A laban movement analysis in two case studies. Journal of Neuroscience Methods, 158:137–149, 2006.

[Gav99] D. M. Gavrila. The visual analysis of human movement: A survey. Computer Vision and Image Understanding, 73:82–98, 1999.

[GKD09] Abhinav Gupta, Aniruddha Kembhavi, and Larry S. Davis. Observing human-object interactions: Using spatial and functional compatibility for recognition. IEEE Trans. Pattern Anal. Mach. Intell., 31(10):1775–1789, October 2009.

[HMNP11] M.B. Holte, T.B. Moeslund, N. Nikolaidis, and I. Pitas. 3d human action recognition for multi-view camera systems. In 3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), 2011 International Conference on, pages 342 –349, may 2011.

[HNB04] Somboon Hongeng, Ram Nevatia, and Francois Bremond. Video-based event recognition: Activity representation and probabilistic recognition methods. computer vision and image understanding, 96:129–162, 2004.

[Hun03] Hayley Hung. Towards modelling the semantics of natural human body movement. Technical report, PhD-MPhil Transfer report, 2003.

[Hut74] Ann Hutchinson. Labanotation. Oxford Univ. Press, 1974. ISBN: 0193113090.

[JH02] Neil Johnson and David Hogg. Representation and synthesis of behaviour using gaussian mixtures. Image and Vision Computing, 20:889–894, 2002.

[JHYGP09] D.B. Jayagopi, H. Hung, Chuohao Yeo, and D. Gatica-Perez. Modeling dominance in group conversations using nonverbal activity cues. Audio, Speech, and Language Processing, IEEE Transactions on, 17(3):501–513, 2009.

[Kah03] Kanav Kahol. Gesture segmentation in complex motion sequences. Master's thesis, Arizona State University, 2003.

[KAM+11] Kamrad Khoshhal, Hadi Aliakbarpour, Kamel Mekhnacha, Julien Ros, Joao Quintas, and Jorge Dias. Lma-based human behaviour analysis using hmm. In Luis M. Camarinha-Matos, editor, DoCEIS, volume 349 of IFIP, pages 189–196. Springer, 2011.

[KAQ+10] Kamrad Khoshhal, Hadi Aliakbarpour, Joao Quintas, Paulo Drews, and Jorge Dias. Probabilistic lma-based classification of human behaviour understanding using power spectrum technique. In 13th International Conference on Information Fusion2010, EICC Edinburgh, UK, July 2010.

[KAQ+11] Kamrad Khoshhal, Hadi Aliakbarpour, Joao Quintas, Martin Hofmann, and Jorge Dias. Probabilistic lma-based human motion analysis by conjugating frequency and spatial based features. In 12th international Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS'11), Delft, The Netherlands, April 2011. TU Delft; EWI; MM; PRB.

[KBOZ04] Timor Kadir, Richard Bowden, Eng-Jon Ong, and Andrew Zisserman. Minimal training, large lexicon, unconstrained sign language recognition. In British Machine Vision Conference 2004. British Machine Vision Conference 2004, 2004. Winner of the Industrial Paper Prize.

[Ken90] Adam Kendon. Conducting Interaction: Patterns of behavior in focused encounters. Cambridge University Press, 1990. ISBN: 978-0521389389.

[KH99] Robert M. Krauss and Uri Hadar. Gesture, speech, and sign, chapter The role of speech-related arm/hand gestures in word retrieval, pages 93–116. Oxford University Press, 1999.

[KI95] Sing Bing Kang and K. Ikeuchi. Toward automatic robot instruction from perception-temporal segmentation of tasks from human hand motion. Robotics and Automation, IEEE Transactions on, 11(5):670 –681, oct 1995.

[KN05] Shehzad Khalid and Andrew Naftel. Classifying spatiotemporal object trajectories using unsupervised learning of basis function coefficients. In 3rd ACM international workshop on video surveillance and sensor networks, pages 45–52, 2005.

[KSH05] Yan Ke, R. Sukthankar, and M. Hebert. Efficient visual event detection using volumetric features. In Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on, volume 1, pages 166 – 173 Vol. 1, oct. 2005.

[KSH07] Yan Ke, R. Sukthankar, and M. Hebert. Event detection in crowded videos. In Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on, pages 1 –8, oct. 2007.

[Lau94] A. Laurentini. The visual hull concept for silhouette-based image understanding. IEEE Trans. Pattern Anal. Mach. Intell., 16(2):150–162, February 1994.

[LBDM04] Olivier Lebeltel, Pierre Bessiere, Julien Diard, and Emmanuel Mazer. Bayesian robot programming. Auton. Robots, 16(1):49–79, January 2004.

[LDG+04] M. Leo, T. Dorazio, I. Gnoni, P. Spagnolo, and A. Distante. Complex human activity recognition for monitoring wide outdoor environments. In Proceedings of the 17th International Conference on Pattern Recognition (ICPR04), volume 4, pages 913–916, 2004.

[LDS04] Marco Leo, Tiziana Dorazio, and Paolo Spagnolo. Human activity recognition for automatic visual surveillance of wide areas. In Proceedings of the 2nd International workshop on video surveillance and sensor networks, pages 124–130, 2004.

[LLF00] C. Lu, H. Liu, and N.J. Ferrier. Multidimensional motion segmentation and identification. In Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on, volume 2, pages 629 –636 vol.2, 2000.

[LLRE07] Lian Loke, Astrid T. Larssen, Toni Robertson, and Jenny Edwards. Understanding movement for interaction design: frameworks and approaches. Personal Ubiquitous Comput., 11:691–701, December 2007.

[LN06] Fengjun Lv and Ramakant Nevatia. Recognition and segmentation of 3-d human action using hmm and multi-class adaboost. In Proceedings of the 9th European conference on Computer Vision - Volume Part IV, ECCV'06, pages 359–372, Berlin, Heidelberg, 2006. Springer-Verlag.

[LOB11] Javier Ortiz Laguna, Angel García Olaya, and Daniel Borrajo. A dynamic sliding window approach for activity recognition. In Proceedings of the 19th international conference on User modeling, adaption, and personalization, UMAP'11, pages 219–230, Berlin, Heidelberg, 2011. Springer-Verlag.

[LS09] Jorge Dias Luis Santos, Jose Prado. Human robot interaction studies on laban human movement analysis and dynamic background segmentation. In IROS 2009, The 2009 IEEE/RSJ International Conference on Intelligent RObots and Systems, 2009.

[MA07] Sushmita Mitra and Tinku Acharya. Gesture recognition: A survey. IEEE TRANSACTIONS ON SYSTEMS, MAN AND CYBERNETICS PART C: APPLICATIONS AND REVIEWS, 37, MAY 2007.

[MB05] Kervin Murphy and Berkeley. Bayesnet toolbox. In http://bnt.insa-rouen.fr/, 2005.

[MbK06] Thomas B. Moeslund, Adrian Hilton b, and Volker KruÂ¨ger. A survey of advances in vision-based human motion capture and analysis. Computer Vision and Image Understanding, pages 90–126, 2006.

[MCB+01] Gerard Medioni, Isaac Cohen, Francois Bremond, Somboon Hongeng, and Ramakant Nevatia. Event detection and analysis from video streams.

IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, 23(8):873–889, AUGUST 2001.

[Mcl04] Geoffrey J. Mclachlan. Discriminant Analysis and Statistical Pattern Recognition (Wiley Series in Probability and Statistics). Wiley-Interscience, August 2004. ISBN-13: 978-0471691150.

[MG01] Thomas B. Moeslund and Erik Granum. A survey of computer vision-based human motion capture. Computer Vision and Image Understanding, 81:231–268, 2001.

[MGIJ03] Bartlett M.S., Littlewort G., Fasel I., and Movellan J.R. Real time face detection and expression recognition: Development and application to human-computer interaction. CVPR Workshop on Computer Vision and Pattern Recognition for Human-Computer Interaction., 2003.

[ML98] P. Morguet and M. Lang. Spotting dynamic hand gestures in video image sequences using hidden markov models. In Image Processing, 1998. ICIP 98. Proceedings. 1998 International Conference on, pages 193 –197 vol.3, oct 1998.

[MPF03] Alessandro Mecocci, Massimo Pannozzo, and Antonio Fumarola. Automatic detection of anomalous behavioural events for advanced real-time video surveillance. In ClMSA 2003 - lntemalional Symposium on Computational Intelligence for Measurement Systems and Applicalions, pages 29–31, Lugano, Switrerland, July 2003.

[MS10] Andrea Mannini and Angelo Maria Sabatini. Machine learning methods for classifying human physical activity from on-body accelerometers. Sensors, 10:1154–1175, 2010.

[Nay06] M. Naylor. Annotated digital video for intelligent surveillance and optimised retrieval. Technical report, ADVISOR project website (http://www-sop.inria.fr/orion/ADIVSOR), 2006.

[NFM05] J.C. Nascimento, M. A. T. Figueiredo, and J. S. Marques. Segmentation and classification of human activities. In HAREM 2005 - International Workshop on Human Activity Recognition and Modelling, Oxford, UK, September 2005.

[NMS02] T. Nakata, T. Mori, and T. Sato. Analysis of impression of robot bodily expression. Journal of Robotics and Mechatronics, 14:27–36, 2002.

[OKGfA04] Abhijit S. Ogale, Alap Karapurkar, Gutemberg Guerra-filho, and Yiannis Aloimonos. View-invariant identification of pose sequences for action recognition. In In VACE, 2004.

[ORP00] Nuria M. Oliver, Barbara Rosario, and Alex P. Pentland. A bayesian computer vision system for modeling human interactions. IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, 22(8), AUGUST 2000.

[OSK02] K. Oka, Y. Sato, and H. Koike. Real-time fingertip tracking and gesture recognition. Computer Graphics and Applications, IEEE, 22(6):64 – 71, nov/dec 2002.

[Pen00] Alex Pentland. Looking at people: Sensing for ubiquitous and wearable computing. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22:107–119, 2000.

[Pen08] Alex (Sandy) Pentland. Honest Signals: How They Shape Our World. The MIT Press, 2008. ISBN: 978-0262515122.

[PL95] Alex Pentland and Andrew Liu. Modeling and prediction of human behavior. IEEE Intelligent vehicles 95, pages 350–355, 1995.

[Pop07] Ronald Poppe. Vision-based human motion analysis: An overview. Computer Vision and Image Understanding, 108:4–18, 2007.

[Pop10] Ronald Poppe. A survey on vision-based human action recognition. Image Vision Comput., 28:976–990, June 2010.

[PPNH06] Maja Pantic, Alex Pentland, Anton Nijholt, and Thomas Huang. Human computing and machine understanding of human behavior: a survey. In ICMI '06: Proceedings of the 8th international conference on Multimodal interfaces, pages 239–248, New York, NY, USA, 2006. ACM.

[PPNH07] Maja Pantic, Alex Pentland, Anton Nijholt, and Thomas Huang. Human Computing and Machine Understanding of Human Behavior: A Survey, volume 4451/2007 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2007.

[PR00] V. Pavlovic and J.M. Rehg. Impact of dynamic model learning on classification of human motion. In Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on, volume 1, pages 788 –795 vol.1, 2000.

[PRO] PROFANE. Prevention of falls network european.

[PVW07] P. Peursum, S. Venkatesh, and G. West. Tracking-as-recognition for articulated full-body human motion analysis. In Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on, pages 1 –8, june 2007.

[RA00] Yong Rui and P. Anandan. Segmenting visual actions based on spatio-temporal motion patterns. In CVPR, pages 1111–1118. IEEE Computer Society, 2000.

[RA06] M. S. Ryoo and J. K. Aggarwal. Recognition of composite human activities through context-free grammar based representation. In CVPR '06: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pages 1709–1718, Washington, DC, USA, 2006. IEEE Computer Society.

[RA08] M. S. Ryoo and J. K. Aggarwal. Recognition of high-level group activities based on activities of individual members. In WMVC '08: Proceedings of the 2008 IEEE Workshop on Motion and video Computing, pages 1–8, Washington, DC, USA, 2008. IEEE Computer Society.

[RA09] M. S. Ryoo and J. K. Aggarwal. Semantic representation and recognition of continued and recursive human activities. Int. J. Comput. Vision, 82:1–24, April 2009.

[Rao48] C. Radhakrishna Rao. The utilization of multiple measurements in problems of biological classification. Journal of the Royal Statistical Society - Series B, 10(2):159–203, 1948.

[RD07] J. Rett and J. Dias. Human robot interaction based on bayesian analysis of human movements. Proceedings of EPIA 07, Lecture Notes in AI, Springer Verlag, Berlin., 2007.

[RD13a] Kamrad Khoshhal Roudposhti and Jorge Dias. Probabilistic human interaction understanding: Exploring relationship between human body motion and the environmental context. Pattern Recognition Letters, 34(7):820 – 830, 2013. Scene Understanding and Behaviour Analysis.

[RD13b] Kamrad Khoshhal Rousposhti and Jorge Dias. Improved semantic-based human interaction understanding using context-based knowledge. In IEEE International Conference on Systems, Man, and Cybernetics (IEEE SMC 2013), Manchester, October 2013. IEEE.

[RDA08] Joerg Rett, Jorge Dias, and Juan-Manuel Ahuactzin. Brain, Vision and AI, chapter Laban Movement Analysis using a Bayesian model and perspective projections, pages 183–192. InTech, 2008. ISBN: 978-953-7619-04-6.

[Ret08] Joerg Rett. ROBOT-HUMAN Interface using LABAN Movement Analysis Inside a Bayesian framework. PhD thesis, University of Coimbra, 2008.

[RJ01] P. Remagnino and G.A. Jones. Classifying surveillance events from attributes and behaviour. In Proceedings of the biritish machine vision conference, pages 685–694, Manchester, 2001.

[RSAD12a] Kamrad Khoshhal Roudposhti, Luis Santos, Hadi Aliakbarpour, and Jorge Dias. Parameterizing interpersonal behaviour with laban movement analysis - a bayesian approach. In Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on, pages 7 –13, june 2012.

[RSAD12b] Kamrad Khoshhal Roudposhti, Luis Santos, Hadi Aliakbarpour, and Jorge Dias. Parameterizing interpersonal behaviour with laban movement analysis - a bayesian approach. In CVPRW'12, Providence, Rhode Island, USA, June 2012. IEEE.

[RTVM10] S. Rao, R. Tron, R. Vidal, and Yi Ma. Motion segmentation in the presence of outlying, incomplete, or corrupted trajectories. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 32(10):1832 –1845, oct. 2010.

[Rub06] Laurence Z. Rubestein. Falls in older people: epidemiology, risk factors and strategies for prevention. Age and Ageing, 35-S2:ii37–ii41, 2006.

[RVRE08] Hossein Ragheb, Sergio Velastin, Paolo Remagnino, and Tim Ellis. Human action recognition using robust power spectrum features. In 15th IEEE International Conference on Image Processing, ICIP 2008., pages 753–756, 2008.

[RYS02] C. Rao, A. Yilmaz, and M. Shah. View-invariant representation and recognition of actions. Internation Journal of Computer Vision, pages 203–226, 2002.

[SD10] Luis Santos and Jorge Dias. Human motion classifier based on laban movement analysis. In 4th International Conference on Cognitive Systems, COGSYS 2010, ETH Zurich, Switzerland, 2010.

[SNTP11] Xiaofan Sun, Anton Nijholt, Khiet P. Truong, and Maja Pantic. Automatic visual mimicry expression analysis in interpersonal interaction. In IEEE Computer Vision and Pattern Recognition Workshops, CVPRW 2011, pages 40–46, USA, August 2011. IEEE Computer Society.

[STMea09] Dilip Swaminathan, Harvey Thornburg, Jessica Mumford, and et al. A dynamic bayesian approach to computational laban shape quality analysis. Advances in Human-Computer Interaction, 2009:17, 2009.

[SWCS08] Qinfeng Shi, Li Wang, Li Cheng, and A. Smola. Discriminative human action segmentation and recognition using semi-markov model. In Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, pages 1 –8, june 2008.

[SZJ+09] Guangyi Shi, Yuexian Zou, Yufeng Jin, Xiaole Cui, and Wen J. Li. Towards hmm based human motion recognition using mems inertial sensors. In Proceedings of the 2008 IEEE, International Conference on Robotics and Biomimetics, 2009.

[THB00] Dirk Ormoneit Trevor, Trevor Hastie, and Michael Black. Functional analysis of human motion data, 2000.

[TVC08] P. Turaga, A. Veeraraghavan, and R. Chellappa. Statistical analysis on stiefel and grassmann manifolds with applications in computer vision. In Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, pages 1 –8, june 2008.

[VKD08] S.N. Vitaladevuni, V. Kellokumpu, and L.S. Davis. Action recognition using ballistic dynamics. In Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, pages 1 –8, june 2008.

[VPB09] Alessandro Vinciarelli, Maja Pantic, and Hervé Bourlard. Social signal processing: Survey of an emerging domain. Image Vision Comput., 27(12):1743–1759, 2009.

[VSP09] A. Vinciarelli, H. Salamin, and M. Pantic. Social signal processing: Understanding social interactions through nonverbal behavior analysis. In Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on, pages 42 –49, june 2009.

[VV96] Willem B. Verwey and Hans A. Veltman. Detecting short periods of elevated workload. a comparison of nine workload assessment techniques. Journal of Experimental Psychology. Applied, 2(3):270–285, 1996.

[WB99] A.D Wilson. and A.F Bobick. Parametric hidden markov models for gesture recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 21(9):884–900, 1999.

[WCL+08] Chia-Chi Wang, Chih-Yen Chiang, Po-Yen Lin, Yi-Chieh Chou, I-Ting Kuo, Chih-Ning Huang, and Chia-Tai Chan. Development of a fall detecting system for the elderly residents. In Bioinformatics and Biomedical Engineering, 2008. ICBBE 2008. The 2nd International Conference on, pages 1359–1362, 2008.

[WHT03] Liang Wang, Weiming Hu, and Tieniu Tan. Recent developments in human motion analysis. Pattern Recognition Society, pages 585–601, 2003.

[WOQX05] Xinyu WU, Yongsheng OU, Huihuan QIAN, and Yangsheng XU. A detection system for human abnormal behavior. In Intelligent Robots and Systems, 2005. (IROS 2005). 2005 IEEE/RSJ International Conference, pages 1204–1208, 2005.

[WRB06] Daniel Weinland, Remi Ronfard, and Edmond Boyer. Automatic discovery of action taxonomies from multiple views. In Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on, volume 2, pages 1639 – 1645, 2006.

[WRB11] Daniel Weinland, Remi Ronfard, and Edmond Boyer. A survey of vision-based methods for action representation, segmentation and recognition. Comput. Vis. Image Underst., 115(2):224–241, February 2011.

[Yac99] Yaser Yacoob. Parameterized modeling and recognition of activities. computer vision and image understanding, 73(2):232–247, February 1999.

[YS05] Alper Yilmaz and Mubarak Shah. Recognizing human actions in videos acquired by uncalibrated moving cameras. pages 150 – 157. Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1, 2005, IEEE Computer Society Washington, DC, USA, 2005.

[YS07] A. Yilmaz and M. Shah. A differential geometric approach to representing the human actions. Computer Vision and Image Understanding, 2007.

[YZC03] Zhanfeng Yue, Liang Zhao, and Rama Chellappa. View synthesis of articulating humans using visual hull. In Proceedings of the IEEE International conference on multimedia and Expo, volume 1, pages 489–492, 2003.

[ZB05] L. Zhao and N.I. Badler. Acquiring and validating motion qualities from live limb gestures. Graphical Models, pages 1–16, 2005.

[Zha01] Liwei Zhao. Synthesis and Acquisition of Laban Movement Analysis Qualitative Parameters for Communicative Gestures. PhD thesis, University of Pennsylvania, 2001.

[ZLP06] Massimo Zancanaro, Bruno Lepri, and Fabio Pianesi. Automatic detection of group functional roles in face to face interactions. In Proceedings of the 8th international conference on Multimodal interfaces, ICMI '06, pages 28–34, New York, NY, USA, 2006. ACM.

[ZMI01] L. Zelnik-Manor and M. Irani. Event-based analysis of video. In Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on, volume 2, pages II–123 – II–130 vol.2, 2001.

[ZSV04] Hua Zhong, Jianbo Shi, and M. Visontai. Detecting unusual activity in video. In Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on, volume 2, pages II–819 – II–826 Vol.2, june-2 july 2004.