Characterization and Modelling of Complex Motion Patterns

Fabio Martínez Carrillo

Universidad Nacional de Colombia
Facultad de Ingeniería, Departamento de Ingeniería de Sistemas e Industrial
Bogotá D.C., Colombia
2014

Characterization and Modelling of Complex Motion Patterns

Fabio Martínez Carrillo

In fulfillment of the requirements for the degree of:
Doctor en Ingeniería - Ingeniería de Sistemas y Computación

Advisor:
Eduardo Romero Castro, Ph.D.

Co-Advisor:
Antoine Manzanera, Ph.D.

Research Field:
Applied Computing - Image and Video Processing
Research Group:
Computer Imaging and Medical Applications Laboratory - CIM@LAB

Universidad Nacional de Colombia
Facultad de Ingeniería, Departamento de Ingeniería de Sistemas e Industrial
Bogotá D.C., Colombia
2014

A mis padres, mi hermana y mi esposa
. . . quienes son las bases fundamentales en mi vida

Insanity is doing the same thing over and over again and expecting different results.
Locura es hacer la misma cosa una y otra vez esperando obtener diferentes resultados.
— **Albert Einstein**

Acknowledgements

First of all, I want to express my most sincere and immense gratitude to my advisor, professor Eduardo Romero, who helped me to transform step by step and with a lot of patience my academic and professional dreams in reality. The most important lesson was that everyone has at least one opportunity and always the most important resource is the people. Thanks a lot for the guidance, the discussion and the permanent contribution on the research work. *To write many emails can be the start point to a exciting adventure.*

I also specially thank to the professor Antoine Manzanera for the important contributions in my work, for his invaluable support, academic guidance and to the opportunity of carried out an inthership at ENSTA-Paristech. Also to the professor Oscar Acosta for the opportunity to spend some time in LTSI-Rennes laboratory, and to invaluable assistant in the work produced jointly. I want thank to whole people that share with me during my stage on France, their disinterested collaboration allowed me an unforgettable experience.

Thanks to all people in CIM@lab research group that have contributed at different scales in my work and personal enrichment. In particular to my co-authors and friends Charlens Álavarez, Josué Ruano , David Trujilo, Luisa Fernanda Cardenas, Christian Cifuentes, and Angelica Athourtua. To my great friends Oscar Anaya, David Romo, Juan Carlos Caicedo, Carlos Vargas, Andrea Rueda and also to my old teammates Angel Cruz for the broad support received during this process. I thank to the *Universidad Nacional de Colombia* that have sponsored my Ph.D. formation with the intership: *becas para estudiantes sobresalientes de posgrado.*

I am deeply grateful with the support provided by my family. A lot of thanks to my parents and sister, who are example of persistence, constancy and fully dedication to the others. Endless thanks to my loved wife Cristina, with whom I have shared the most wonderful experiences and who has lived with me the challenges that are involved to achieve shared aims. Finally, I want to extend my gratitude to my wife's family for the support during this process.

Abstract

Movement analysis is the principle of any interaction with the world and the survival of living beings completely depends on the efficiency of such analysis. Visual systems have remarkably developed efficient mechanisms that analyze motion at different levels, allowing to recognize objects in dynamical and cluttered environments. In artificial vision, there exist a wide spectrum of applications for which the study of complex movements is crucial to recover salient information. Yet each domain may be different in terms of scenarios, complexity and relationships, a common denominator is that all of them require a dynamic understanding that captures the relevant information. Overall, current strategies are highly dependent on the appearance characterization and usually they are restricted to controlled scenarios. This thesis proposes a computational framework that is inspired in known motion perception mechanisms and structured as a set of modules. Each module is in due turn composed of a set of computational strategies that provide qualitative and quantitative descriptions of the dynamic associated to a particular movement. Diverse applications were herein considered and an extensive validation was performed for each of them. Each of the proposed strategies has shown to be reliable at capturing the dynamic patterns of different tasks, identifying, recognizing, tracking and even segmenting objects in sequences of video.

Keywords: motion analysis, dense Optical Flow, background substraction, action recognition, Tracking, gait analysis, polyp detection, hummingbird flight patterns, cardiac MRI analysis, video-surveillance.

Resumen

El análisis del movimiento es el principio de cualquier interacción con el mundo y la supervivencia de los seres vivos depende completamente de la eficiencia de este tipo de análisis. Los sistemas visuales notablemente han desarrollado mecanismos eficientes que analizan el movimiento en diferentes niveles, lo cual permite reconocer objetos en entornos dinámicos y saturados. En visión artificial existe un amplio espectro de aplicaciones para las cuales el estudio de los movimientos complejos es crucial para recuperar información saliente. A pesar de que cada dominio puede ser diferente en terminos de los escenarios, la complejidad y las relaciones de los objetos en movimiento, un común denominador es que todos ellos requieren una comprensión dinámica para capturar información relevante. En general, las estrategias actuales son altamente dependientes de la caracterización de la apariencia y por lo general están restringidos a escenarios controlados. Esta tesis propone un marco computacional que se inspira en los mecanismos de percepción de movimiento conocidas y esta estructurado como un conjunto de módulos. Cada módulo esta a su vez compuesto por un conjunto de estrategias computacionales que proporcionan descripciones cualitativas y cuantitativas de la dinámica asociada a un movimiento particular. Diversas aplicaciones fueron consideradas en este trabajo y una extensa validación se llevó a cabo para cada uno de ellas. Cada una de las estrategias propuestas ha demostrado ser fiable en la captura de los patrones dinámicos de diferentes tareas identificando, reconociendo, seguiendo e incluso segmentando objetos en secuencias de video.

Palabras clave: análisis de movimiento, flujo optico denso, extracción de fondo, reconocimiento de acciones, seguimiento, análisis de marcha, detección de polipos, patrones de vuelo del colibrí, análisis de sequencias cardiacas MRI, video-vigilancia.

Contents

List of Figures

List of Tables

1 Introduction

Visual systems interact with their environment by developing models that transform an over-fragmented external world into semantic internal clusters of coherent information. The world is basically parsed by analysing dynamic information, usually using a set of captors or receptors that can be though as a bank of filters that segment the original information in terms of very basic primitives [112, 253]. The world is then perceived by coherently grouping such primitives under a metric that is learned when these systems are forced to analyse motion. A dynamic framework results then crucial for the visual system to be part of the world. Upon such analysis, living beings build the rest of the complexity of the visual processing, achieving visual integration and of course recognition [8]. Interestingly enough, this information is also used to predict events or future actions that follow a certain coherency, learned from those complex motion analyses.

This central role of the motion analysis, at representing the world, has shaped the different functions of visual systems during the evolution process. Different types of biological specialization have improved the robustness to noise and the time processing, basically searching to optimally respond to any external stimulus. Overall, the human visual system provides a large set of adaptive mechanisms that decompose specific information, which is projected onto different representations at low level and then integrated at the upper structures of brain analysis, as illustrated in Figure **1-1**. Firstly, the temporal change of luminance is captured and transformed by the retina, deeply encoded in the nuclei of the brain-stem and mapped towards the primary visual cortex (V1) (blue region in Figure **1-1**). In this primary cortex, a spatio-temporal correlation of the encoded information is performed, perceiving at this level global shapes with a very short history, that is to say, nearly instantaneous motion. The inner mechanism passes by specialized neurons that act as spatio-temporal filters of low level features, composing a first low frequency approximation of the global scene organization [148, 210]. The primary visual cortex is also responsible for transmission of the encoded information onto the two higher-level pathways: the dorsal and ventral stream paths, entailed with larger and more complex receptive fields that facilitate a further specialized processing. The former, the dorsal stream path (associated to the green region in Figure **1-1**) comprises several cortical regions, such as the V6 and the Middle

Temporal (MT) areas. This path is responsible for capturing the oriented coherent motion associated to high temporal frequencies [129, 179]. The regions associated with the dorsal path accounts for the 3D perception of the scene and early identification of objects in motion [179]. The latter, the ventral stream path (the purple region in Figure **1-1**) includes the V2, the V4 and the inferior temporal lobe. This path is charged of associating the captured temporal visual features with semantic concepts learned as a prior model, classically known as the visual work memory [42]. Visual analysis in the dorsal stream follows an egocentric motion perspective, i.e., the viewer is established as reference of scene. In contrast, regions related with the ventral pathway follow an allocentric frame, for which the object in motion is the reference.

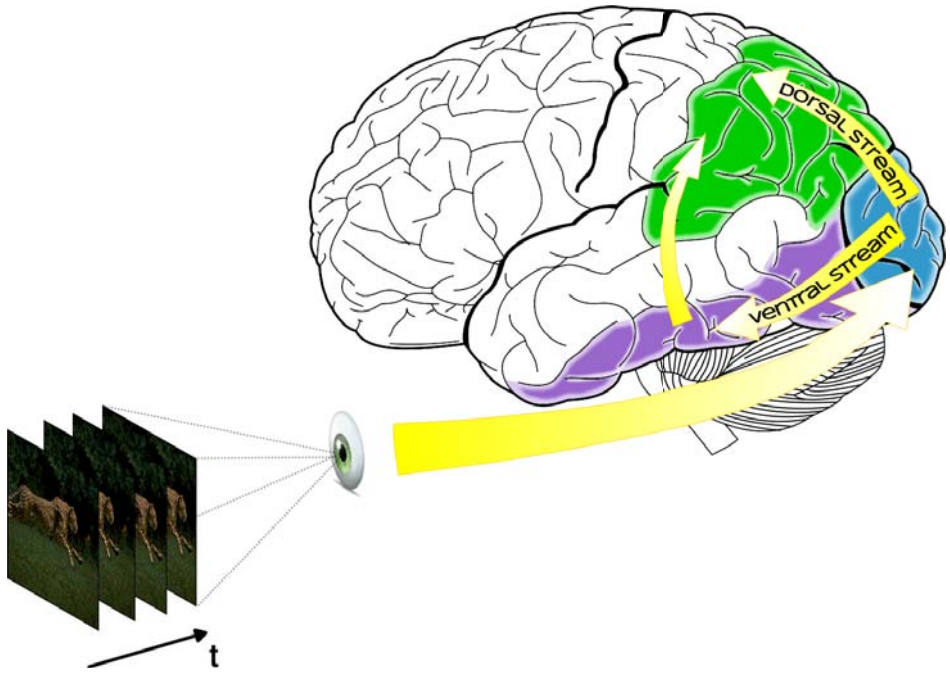


Figure **1-1**: Motion perception in human visual systems. Three main region are responsible for perceiving motion information. The primary visual cortex is represented in blue, a structure charged of temporally correlating motion information. In the second level, the dorsal stream is responsible for characterizing dynamic patterns, as performed by an optical flow representation. Finally, in the third level, the ventral stream is responsible to match semantic concepts with a set of motion observations. Each module can independently perceive motion or interact with the other regions.

Finally, The information processed in these regions is sent to the other regions and then fed back, at different temporal cycles, optimizing the general information process

and obtaining at the end a very accurate map of the visual world.

From an engineer perspective, a proper understanding of such mechanisms opens up new perspectives, and not only because it facilitates the introduction of new techniques, but above all since it allows to organize smart pipelines of new methods to attack any problem. This statement is at the very base of this thesis and has guided us during this research. Different applications, in several domains, highly depend on motion analysis, for instance the robotic area at understanding both the human locomotion and particular dynamics of some animals. In the rehabilitation domain, diagnosis, treatment and follow-up depend on a strong objective support at studying the dynamics of the human body, or a particular movement in a specific disease such as certain abnormal gait patterns in brain palsy or musculo-skeletal disorders [224]. Many surveillance applications require of spatio-temporal characterization of human movements to facilitate classification and recognition of abnormal activities. In some cases, motion analyses of organs such as the heart might potentially change our comprehension of the disease and therefore the patient management.

1.1 Research Problem

So far the limited understanding of complex motion patterns has prevented a natural representation of the dynamics of different objects in different contexts. Emerging areas have compelled researchers to objectively analyze complex recorded movements. Nonetheless, these techniques, highly dependent on the appearance, are typically restricted to controlled scenarios. Likewise, these complex movements are usually measured with specific devices that alter natural gestures and introduce noise during the the acquisition. Usually, motion analysis uses very simple prior hypotheses with very few generality and little extension of such methods to other contexts, thereby considerably restricting a proper understanding and characterization of the different conditions in many dynamic systems.

1.1.1 Research Question

In spite of the variability of complex motion patterns, the visual system is capable of capturing very different processes while understanding the underlying mechanism, using essentially the same mechanism. This consideration has led us to the main research question in this work: *how to learn a* ***computational framework*** *from visual systems to perform complex motion tasks in different artificial vision problems.*

1.2 A computational framework for motion analysis

Motion is perceived under an analysis framework that integrates information captured by a set of specialized components, filtering out different types of noise and structuring visual concepts. These systems are very adaptable, they balance holistic or part-based representations when needed if the interaction circumstance demands understanding of complex scenarios. The remarkable point is that such simple mechanisms are apt to cover a practically infinite number of visual perceptions.

In this thesis we have proposed a computational framework that emulates the primary motion analysis of the human visual system, constructed upon a set of computational tools and mathematical formulations that provide qualitative and quantitative descriptions of the dynamics. This framework locally characterizes moving objects and permits a global understanding of their fundamental relationships. As the visual system, three components are at the base of three complementary levels of processing. A first module emulates the primary visual cortex function at computing the temporal change coherence among consecutive frames, recovering a coarse shape representation. A temporal anaysis of variance follows those temporal changes with a certain level of statistical dependence, since no changes are observed at any scene during a short time interval.

A second module of the proposed framework aims to simulate the motion characterization of the secondary pathways, the dorsal stream, as the obtained optical flow pattern that would correspond, in such case, to the orientation maps of the motion flow defined at this visual processing level. Visual systems compute a vector field as the distribution of apparent velocities, allowing an analysis of the relative motion of objects w.r.t the observer. These bottom-up patterns result decisive for visual learning, perception of structure, self-localization and even for stereo vision. In this work, a set of statistical descriptors of a dense optical flow map, were successfully evaluated in classification and recognition tasks under uncontrolled conditions.

Finally, a third module includes the integration of visual motion patterns with learned semantic concepts, accomplishing the track of complex objects under very noisy conditions. Overall, visual systems associate learned motion patterns or even raw observations with prior belief, i.e., learned knowledge of the world. This association can be seen as a set of likelihood mappings which allow a probabilistic description of learned concepts with spatio-temporal observations. This prior belief is recreated in visual systems as knowledge models that are able to efficiently interact with the environment using a restricted set of salient points. The proposed module, a probabilistic standpoint of the motion problem, has been devoted to the development of appropriate priors that adapt different types of problems, coupled with a tracking strategy that imitates the visual system motion perception. This module has been extensively proven in different

types of gait problems.

1.3 Complex motion patterns in diverse scenarios

There exists a wide spectrum of problems for which the quantitative and qualitative motion analyses are crucial and many intractable problems result better approached from this perspective, even if they are different in terms of their sources of noise, scenarios or levels of non-linearity. This thesis has been devoted to globally understand the essence of the motion analysis and to use such proposal as a different angle in classical research problems. Hereafter we introduce the different applications for which such motion analysis was applied, even though some of them look completely different. On the contrary, as we will discuss later, these problems share fundamental challenges such as the identification of abnormal patterns or discovering of hidden relationships among a set of objects of interest.

1.3.1 Clinical Gait Analysis

Distortion of gait patterns are one of the main clinical manifestations of many diseases, among others diabetes, parkinson, brain palsy or accident sequelae. The analysis of human gait attempts to objectively assess pathologies by following up the hidden gait dynamic variables. The set of techniques dedicated to perform this analysis is what is currently known as the gait laboratory, a tool devised to quantify a disease and to compare the gait with normal patterns [224]. Conventional gait analysis is invasive and alters the natural movement gestures, necessitating strong variations to achieve diagnosis, i.e., this approach is hardly useful in early stages. In this thesis, the gait analysis has been addressed from a markerless perspective to analyze normal and pathological motion patterns. From this standpoint, the gait analysis is extended to classifying, tracking and recognizing tasks. Since the gait is the result of complex relationships among different subsystems, an original model herein formulated and developed served to emulate complex motion patterns.

1.3.2 Human Action Recognition for Surveillance Applications

Human action recognition is a challenging task in many surveillance applications [4, 229]. The challenge is to automatically detect normal and abnormal actions, occurring in ordinary visual scenes and to classify actions. The challenge is related to the huge variability, in terms of relationships between the scene components, people appearance and poorly controlled conditions. In this work, the human action recognition is

aimed to analyze normal and abnormal human motion patterns in real scenarios. This video-processing analysis attempts to automatically classify and recognize of human activities.

1.3.3 MRI-based Motion Analysis

one of the principal causes of death and disability [213]. Magnetic Resonance Cine (MRC) is a non invasive video technique that provides cardiac anatomic information, allowing the analysis of the heart to globally estimate the myocardial function [135]. Overall, these analyses are carried out by expert physicians, whereby results are highly subjective. In this thesis the heart motion characterization is addressed towards the automation of this analysis, especially useful in early disease detection, case in which the boundary between normal and abnormal is quite diffuse, even for expert physicians.

1.3.4 Polyp shape and size estimation from spatio-temporal patterns

Digestive diseases constitute worldwide one of the most common public health problems. Polyp size quantification results crucial to determine the surgical procedure, either by immediately extracting the polyp or just by increasing the patient follow up. Currently, the video endoscopy estimation is the gold standard procedure to determine the polyp size. However, estimation with this method is highly subjective, among others because the large variability due to the capturing device and the specialist experience. Automatic approaches may help to overcome such problem by obtaining more repeatable estimations, reducing the measure subjectivity and allowing better predictions of the pathology evolution.

1.3.5 Discovering Hummingbird motion patterns in video-sequences

The emulation of flying animal movements would allow micro air vehicles (MAVs) to perform complex maneuvers and to expend less energy [198]. Among these MAVs, hovering machines have the ability to perform accurate maneuvers within a limited space [198]. The design of hovering MAVs can be approached by observing animals such as Hummingbirds [130]. Their maneuverability is based on their ability to suddenly turn at any direction. However, crucial hummingbird flight issues are yet not fully understood. The only available information about these turns comes from other species like gliding or flapping birds, and hovering insects [130,132], but these animals are very

different since the hummingbird is entailed with simple wings that operate over a larger range of speeds [11]. Typical devices that capture information are very limited and generally based on few reference points that must be superimposed with the animal structure. These devices usually introduce artefacts that alter natural gestures of the hummingbird movement since markers need to be placed on the bird, markers that can be easily occluded like the shoulder joint.

1.4 Contributions and Academic Products

The main contribution of this work is a computational framework, as a set of tools, that analyzes and characterizes complex motion patterns. Three different scalable and complementary motion analysis phases were herein proposed. Hereafter we present the main applications covered by the proposed methods along with the achieved academic products.

1.4.1 Locally estimating temporal shape patterns

The motion analysis in this phase allows to recover those moving objects in a sequence of video. From this analysis it was possible for instance to recover human silhouettes during walking and to perform automatic markerless classification and recognition of pathological patterns. This analysis module was also used in segmentation tasks, in which a set of target structures were recovered by characterizing them with motion strategies. Two different applications were tested in classification tasks: Right ventricle on MRI sequences and Polyps on endoscopy. This work was published in:

- **Fabio Martínez**, Josué Ruano, Martín Gómez, Eduardo Romero. Estimating the size of polyps during actual endoscopy procedures using a spatio-temporal characterization. Submitted to Journal of the Computerized Medical Imaging and Graphics. 2014.
- Josué Ruano, **Fabio Martínez**, Eduardo Romero.*A 3D endoscopy reconstruction as a saliency map for analysis of polyp shapes.* Submitted in SIPAIM 2014.
- Josué Ruano, **Fabio Martínez**, Eduardo Romero.*Shape estimation of gastrointestinal polyps using motion information.* International Seminar on Medical Information Processing and Analysis, SIPAIM -2013. Mexico
- **Fabio Martínez**, Juan Carlos León, Eduardo Romero. *Pathology Classification of Gait Human Gestures.* International Conference on Computer Vision Theory and Applications (Visapp). Algarve - Portugal (2011).

- Juan Carlos León,**Fabio Martínez**, Eduardo Romero. *A robust background subtraction algorithm using a* Σ- Δ *estimation.* International Conference on Computer Vision Theory and Applications (Visapp). Italy. (2012)

- Juan Carlos León, **Fabio Martínez**, Eduardo Romero. Classification of Pathological Gait Markerless Pattern. SIPAIM: VI Seminario Internacional de Procesamiento y Análisis de Imágenes Médicas. Bogotá. Colombia (2010).

- Juan Carlos León, **Fabio Martínez**, Eduardo Romero. A Comparison of Σ- Δ Background Substraction Algorithms for Gait Analysis. SIPAIM: VII Seminario Internacional de Procesamiento y Análisis de Imágenes Médicas. Bucaramanga. Colombia (2011).

- David Trujillo, **Fabio Martínez**, Eduardo Romero.Characterizing the eye trajectory during the gait towards Parkinson stage identification. Submitted in SIPAIM 2014.

- Angélica Atehortúa, **Fabio Martínez**, Eduardo Romero. *A Novel Right Ventricle Segmentation Approach from Local Spatio-Temporal MRI Information.*. Iberoamerican Congress on Pattern, CIARP -2013. Cuba -2013.

- Angélica Atehortúa, **Fabio Martínez**, Eduardo Romero. *A novel right ventricle segmentation strategy using local spatio-temporal MRI information with a prior regularizer term.*International Seminar on Medical Information Processing and Analysis, SIPAIM -2013. México.

- Angélica Atehortúa, **Fabio Martínez**, Eduardo Romero. *Automatic right ventricle (RV) segmentation by propagating a basal spatio-temporal characterization.*International Seminar on Medical Information Processing and Analysis, submitted SIPAIM -2014

1.4.2 Spatio-temporal descriptors from optical flow patterns

A second proposed module includes a set of spatio-temporal descriptors that capture motion patterns from the computed optical flow. These descriptors are mainly used to capture the dynamic of targets captured in noisy conditions. These descriptors were used in surveillance applications for recognition and classification of normal and abnormal human activities. Also, the characterization of optical flow patterns allowed to characterize normal and pathological cardiac trajectories. In addition, the computation of biomechanical hummingbird flight patterns was calculated during hovering

trajectories using a fully markerless strategy. The academic products of this module are:

- **Fabio Martínez**, Antoine Manzanera, Eduardo Romero. An overcomplete spatio-temporal multi-scale motion descriptor for human action recognition in video-surveillance from an optical flow characterization. Submitted to IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI). 2014.

- **Fabio Martínez**, Antoine Manzanera, Eduardo Romero. *Representing activities with layers of velocity statistics for multiple human action recognition in surveillance applications.* SPIE Electronic Imaging. San Francisco. USA. February 2014.

- **Fabio Martínez**, Antoine Manzanera, Eduardo Romero. *A motion descriptor based on statistics of optical flow orientations for action classification in video-surveillance.* Int. Conf. on Multimedia and Signal Processing (CMSṔ12). Shanghai, China. 2012

- **Fabio Martínez**, Antoine Manzanera, Eduardo Romero. Automatic Analysis and Characterization of the hummingbird flight using dense optical flow features. Submitted to Journal of Bioinspiration & Biomimetics. 2014.

- **Fabio Martínez**, Antoine Manzanera, Eduardo Romero. *Analysing the hovering flight of the hummingbird using statistics of the optical flow field.* ICPR Workshop on Visual observation and analysis of animal and insect behavior. Tsukuba, Japan (VAIB 2012)

- **Fabio Martínez**, Antoine Manzanera, Cristina Santa Marta, Eduardo Romero. Characterization of Motion Cardiac Patterns in Magnetic Resonance Cine. International conference on Image Information Processing. INDIA (2011)

1.4.3 Dynamic Analysis by integrating prior knowledge

Complex motions such as the human gait have been widely studied, aiming to understand the underlying gait mechanisms and the relations that produce optimal energy consumption. Several studies have proposed dynamical models that, under some restrictive conditions, can recreate these phenomena. The proposed framework may naturally includes prior gait models, which potentially may be fused with a set of observations to analyze pathological gait patterns. In addition, some prior models have been extended to include additional relationships of complex motion and emulate more realistic trajectories. The academic products obtained in this module are:

- **Fabio Martínez**, Francisco Gómez, Eduardo Romero. *A kinematic method for computing the motion of the body centre-of-mass (CoM) during walking: a Bayesian approach.* Journal of Computer Methods in Biomechanics and Biomedical Engineering. 14(6):561-72, 2011

- **Fabio Martínez**, Christian Cifuentes, Eduardo Romero. *Simulation of normal and pathological gaits using a fusion knowledge strategy.* J Neuroeng Rehabil. 10(1):73, 2013.

- Christian Cifuentes, **Fabio Martínez**, Eduardo Romero. Physics-based model to simulate the Parkinsonian gait. SIPAIM: VII Seminario Internacional de Procesamiento y Análisis de Imágenes Médicas. Bucaramanga. Colombia (2011)

- Christian Cifuentes, **Fabio Martínez**, Eduardo Romero. A 3D physics-based model to simulate normal and pathological gait patterns. International Conference on Computer Graphics Theory and Applications (GRAPP). Rome - Italy. (2012)

- Fernanda Sarmiento, **Fabio Martínez**, Eduardo Romero. Automatic Characterization of the Parkinson Disease by Classifying the ipsilateral coordination and spatiotemporal gait patterns. submittted to SIPAIM 2014

- Luisa Fernanda Cardenas, **Fabio Martínez**, Eduardo Romero. Simulation of parkinsonian gait by fusing a physical model and trunk learned patterns. submittted to SIPAIM 2014

On the other hand, a bottleneck of the modern prostate radiotherapy treatments is the accurate delineation of the prostate, a procedure usually performed by delineating this organ. The problem of segmenting the prostate and the neighboring organs was herein approached from a dynamical standpoint. Yet this problem is static, the notion of movement allows to evolve towards the true shape. In this case, the learned temporal evolution amounts to a shape deformation analysis, for which a set of prostate shapes is arranged according to the probability of occurrence. This vector of learned shapes encodes the variance of the perceived objects. The objects, in this case the prostates, are mapped to a feature space that facilitates to dynamically find the most similar shape. Academic products obtained from this perspective were:

- **Fabio Martínez**, Eduardo Romero, Gael Dreán, Antoine Simon, Pascal Haigron, Renaud de Crevoisier, Oscar Acosta. *Segmentation of pelvic structures for planning CT using a geometrical shape model tuned by a multi-scale edge detector.* Accepted for publication in Journal of Physics in Medicine and Biology.(2014)

- Charlens Álvarez, **Fabio Martínez** and Eduardo Romero. *A novel atlas-based approach for MRI prostate segmentation using multiscale points of interest.* International Seminar on Medical Information Processing and Analysis, SIPAIM -2013. Mexico.

- **Fabio Martínez**, Eduardo Romero, Gael DrÃ«án, Antoine Simon, Pascal Haigron, Renaud de Crevoisier, Oscar Acosta.*Segmentation of pelvic structures from planning CT based on a statistical shape model with a multiscale edge detector and geometrical likelihood measures.* Image-Guidance and Multimodal Dose Planning in Radiation Therapy. InternationalWorkshop in MICCAI 2012.

- Carlos Mosquera, **Fabio Martínez**, Oscar Acosta, Renaud de Crevoisier, Eduardo Romero. *Prostate Localization Using a Probabilistic Model.* SIPAIM: VII Seminario Internacional de Procesamiento y Análisis de Imágenes Médicas. Bucaramanga. Colombia (2011).

- **Fabio Martínez**, Oscar Acosta, Renaud de Crevoisier, Eduardo Romero. *Local jet features and statistical models in a hybrid bayesian framework for prostate estimation in CBCT images.* SPIE Medical Images. San Diego, California. United States, USA (2012).

1.5 Thesis Outline

The remaining chapters of the thesis are organized as follows:

- **Chapter 2:** ***Locally estimating temporal shape patterns***

  This chapter presents the first level of the proposed computational framework. This module introduces a set of strategies that capture dynamic patterns by computing a temporal correlation between consecutive frames. Such computed motion history allows to follow shapes in motion and analyze their temporal changes. Diverse applications were approached, namely, the automatic classification of gait pathologies, the measure of polyps in endoscopy and the analysis of the heart ventricle.

- **Chapter 3:** ***Spatio-temporal descriptors from optical flow patterns***

  This chapter presents a second level of motion analysis from an optical flow field representation. The developed descriptors capture the velocity orientation and characterize actions of objects in motion. Contributions were carried out in different applications such as recognition of human activities in surveillance

applications, computation of the hummingbird dynamic patterns during flight and also global heart trajectories analysis to classify normal and pathological movements.

- **Chapter 4:** ***Dynamic Analysis by integrating prior knowledge***

  This chapter presents the third module of the proposed framework, in which a set of prior semantic concepts were associated with dynamic patterns. In particular, for the analysis of pathological gait, it was proposed a sophisticated physical prior that emulates normal and pathological patterns, while an original strategy for tracking the center of mass was also proposed.

  Likewise, these dynamical analysis strategies were applied for delineating the prostate and neighboring organs in the context of a radiotheraphy treatment. For this purpose, a set of learned organ shapes were arranged according to the probability occurrence and then were iteratively mapped to a common space, together with the captured observations, and then to establised the most likely shapes.

- **Chapter 5:** ***Conclusions and Perspectives.*** This final chapter presents the main conclusions of the proposed work, highlighting the main contributions and their impact. In addition, some main issues were discussed as well as future research directions and perspectives.

2 Locally estimating temporal shape patterns

At a first level of motion perception, the primary visual cortex temporally correlates visual information, resulting in a perception of moving shapes. Such analysis ends up by capturing dynamic object patterns. This chapter presents a set of applications, for which the analysis consists basically in correlating consecutive frames to recover the main regions during motion. The analysis has allowed to characterize pathological gaits (see in section 2.1, and 2.2), to identify and measure polyps in routine endoscopy procedures (see in section 2.3) and to track the right ventricle on cine-MRI sequences (see in section 2.4 and 2.5).

2.1 Classification of Pathological Gait Markerless Patterns

Gait patterns may be distorted in a large set of pathologies. In the clinical practice, the gait is studied using a set of measurements which allow identification of pathological disorders, thereby facilitating diagnosis, treatment and follow up. These measurements are obtained from a set of markers, carefully placed in some specific anatomical locations. This conventional procedure is obviously invasive and alters the natural movement gestures, a great drawback for diagnosis and management of the early disease stages. Instead, markerless approaches attempt to capture the very nature of the movement with practically no intervention on the movement patterns. This article introduces a novel markerless strategy for classification of normal and pathological gaits, using view-based video descriptor of the sagital trajectory, stored in a temporal summarization. The strategy was evaluated in three groups of patients: normal, musculoskeletal disorders and parkinson's disease, obtaining a sensitivity around 80%. *The complete content of this section has been published as a research article in the proceedings of* ***SIPAIM-2010*** *(see [94]).*

2.1.1 Introduction

Distortion of gait patterns are the first clinical manifestation of many diseases, among others diabetes, brain palsy, cerebral vascular accidents, neuromuscular dystrophies or accident sequelae. The analysis of human gait attempts to objectively assess pathologies by following up the hidden gait dynamic variables. The set of techniques dedicated to perform this analysis is what is currently known as the gait laboratory, a modern tool devised to quantify a disease, to compare the gait with normal patterns and to efficiently perform the dynamic alignment of lower member prostheses [163, 224]. Most of this gait analysis is carried out with a set of markers, carefully placed in some specific anatomical locations. This conventional procedure is invasive and alters the natural movement gestures leading to wrong pattern measures.

On the other hand, gait dynamic patterns are by nature highly variable and can be easily contaminated with noise. In early stages, most of these diseases differ by very little from what is considered a normal pattern so that classification is a very challenging problem, even for the expert clinicians. This picture may be worsen if one considers that the basic examination tool, the markers, can move very easily or even be unobservable, contaminating the resulting measurement. These factors together lead to subjective clinical analyses with the consequent limitation in the reproduction of the clinic management of the patient [158, 288].

Ultimately, this problem has undergone a fundamental transformation since the objective is not anymore the movement reconstruction from the anatomical markers, but the accurate tracking of the movement pattern i.e. the markerless strategy. Research areas as computer vision, automatic surveillance, animation and image processing have already developed some markerless strategies for diverse applications, namely, biometric identification, abnormal motion detection, scene reconstruction and activity classification [166, 266, 277]. These methods attempt an interpretation of human movements using nothing but the shape and dynamics of the body. From the medical application standpoint, marker strategies have demonstrated their limitations and new procedures are required. This article presents a precise and efficient markerless framework to identify and classify different kinds of normal and pathological movements. This approach uses as input a sagital view video of a patient walking. Every frame is processed to extract the human silhouette, with which we build a view-based video descriptors that are a temporal summarization of the motion history. Hu moments are then computed for each descriptor, a feature vector is obtained and used to classify patterns as normal or pathological using a classical Support Vector Machine strategy. Evaluation was performed on a database with 48 videos from 12 patients, with 3 types of movements: normal, musculoskeletal disorders and parkinson's disease.

2.1.2 Materials and Methods

The proposed strategy segments the silhouette and uses it to construct a gait descriptor: the motion history images(MHI). This descriptor is finally used as a feature in a classical SVM strategy. The whole method is illustrated in Figure **2-1**.

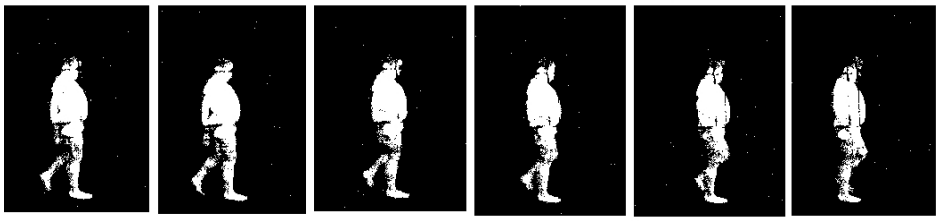
Figure **2-1**: Markerless strategy to Classify Pathological Gait Patterns. This approach consist in a set of step: (a)walking sequence of video, (b) Silhouette extraction process using $\Sigma - \Delta$ algorithm, (c) HMI video descriptor construction, (d) classification method

2.1.2.1 Background Estimation for Silhouette Extraction

Temporal description of the patient gait patterns is central at describing structural changes. Many strategies have been proposed already, they are currently known as background estimation methods [88,141,196]. These methods use a sequence of images It and build a model of the static scene M_t. Output of the model is a Image D_t, where the background is represented by $D_t(x) = 0$ and the foreground is $D_t(x) = 1$.

Among the background estimation methods, the non linear operator $\Sigma - \Delta$ is one of the most robust. This estimator oversampled a signal at higher rates than the specified by the Nyquist theorem, increasing correlation between the adjacent frames, evaluated for each pixel [196]. The $\Sigma - \Delta$ operator behaves as a background tracker $M_t(x)$, dynamically updated by comparing each image $I_t(x)$ with the current background $M_t(x)$, using a simple updating rule: If $I_t(x)$ is greater (lower) than $M_t(x)$, then a positive increase (decrease) $+\Delta$ is performed. The implemented $\Sigma - \Delta$ algorithm is shown in 1, whose results illustrated in Figure 2(a).

(a) Walking sequence



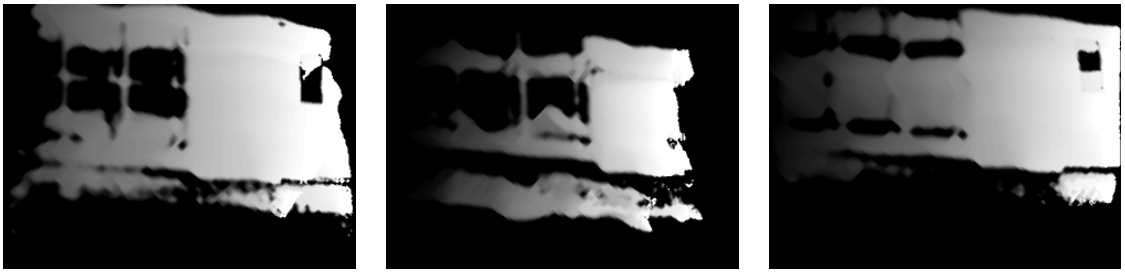
(b) Background estimation

Figure **2-2**: Silhouette extraction estimation. (a) Walking sequence of a patient with a musculoskeletal disorder. (b) Results of the $\Sigma - \Delta$ background estimation

2.1.2.2 Video Descriptor

Once the silhouette is extracted from the video, the next step is to build a video descriptor that represents the dynamic of each walking. In our case, we use the Motion History Image (MHI) represented by $H_\tau(x, y, t)$ [6, 34] which describes the motion by the segmented silhouette changes $D(x, y, t) = 1$. This video descriptor consists in a sequence of consecutive silhouettes, recorded in a single image, i.e., moving pixels are brighter, as follows:

$$H_\tau(x, y, t) = \begin{cases} \tau & \text{if } D(x, y, t) = 1 \\ -max(0, H_\tau(x, y, t-1) - 1) & \\ \text{otherwise.} & \end{cases}$$



(a) (b) (c)

Figure **2-3**: Motion History Image descriptor. (a) Normal Gait. (b)Musculo Eskeletical disease.(c) Parkinson Disease

Algorithm 1 $\Sigma - \Delta$ Algorithm

Initialization: $M_0(x) = I_0(x)$
For each Frame t
$M_t(x) = M_{t-1}(x) + sgn(I_t x - M_{t-1}(x))$

$\Delta_t(x) = \ M_t(x) - I_t x$

Initialization: $V_0(x) = \Delta_t(x)$

For each Frame t
for each pixel x such that $\Delta_t(x) = 0$
$V_t(x) = V_{t-1}(x) + sgn(N \times \Delta_t(x) - V_{t-1}(x))$

if $\Delta_t(x) < V_t(x)$ **then**
  $D_t(x) = 0$
else
  $D_t(x) = 1$
end if

where τ is the spatial window that defines the duration of the sagital patient motion. Figure 2.2 shows typical results for the video descriptor, in different kinds of movements. Figure 3(a) shows the typical smooth pattern of the normal walking, while figure 3(b) shows a irregular pattern, with abrupt changes and long duration, typical of the musculoskeletal disorders [224]. Finally, figure 3(c) shows the video descriptor of a parkinson's disease walking, notice the short movement duration due to the short step length, characteristic of this kind of movement.

2.1.2.3 Gait Data

Validation was carried out with recorded sagital views, registered at the gait lab of the National University, under semi-controlled illumination conditions. The dataset consists of a set of videos captured from 12 patients, each was recorded 4 times while walking, for a total of 48 video sequences. The Dataset was divided as follows:

- 4 patients diagnosed with musculoskeletal disorders.
- 4 patients diagnosed with parkinson's disease (No depressive disorder present).
- 4 patients with normal gait.

2.1.2.4 Attributes for Classification

The Classification phase of the proposed method requires a set of attributes to be extracted from each HMI image. As image moments represent global characteristics of the image objects and provide information about various geometrical features [208], it is expected that the differences between the HMI of normal and pathological gaits can be captured by the mentioned descriptors. Therefore the first seven Raw, Central, Scale Invariant and Hu moments were selected to characterize each of the images, building a feature vector of dimension 7 for each HMI image. An additional Gray scale descriptor was included, built from a rescaled version of the luminance channel of the original color image (it was rescaled by a factor of 10).

2.1.3 Results

Classification was performed using a conventional machine learning method, a Support Vector Machine (SVM) [41]. The SVM was trained with a set of attribute vectors, extracted from already labelled HMI images. In this phase, two types of kernels were employed, polynomial kernels and Radial Basis Function (RBF). Parameters of the cost function, gamma (for RBF kernels) and the exponent (for polynomial kernels) were estimated using the sequential minimal optimization algorithm [270]. The Gait Dataset, with 48 videos, was split into 3 groups according to the type of gait. Each video is represented by 5 HMI and each class is represented by a total of 80 HMI images, from which 69 were selected for training and 11 images for test. The following tables summarize the obtained results.
Tables 1 and 2 show that the rescaled HMI is much better than the description based on moments, but at a higher computational cost. On other hand, scaled invariant and Hu moments, with a sensitivity of 80% can be considered as appropriate for most

Descriptor	Samples	True Positives	True Positives Ratio
Raw Moments	33	21	63%
Central Moments	33	24	72%
Scale Invariant Moments	33	26	79%
Hu Moments	22	11	66%
Gray Descriptor	22	11	66%

Table **2-1**: Classiffication results using Linear Kernels

Descriptor	Samples	True Positives	True Positives Ratio
Raw Moments	33	16	49%
Central Moments	33	16	49%
Scale Invariant Moments	33	11	33%
Hu Moments	33	17	51%
Gray Descriptor	33	32	96%

Table **2-2**: Classification results using RBF Kernels

applications, however results with HMI and a sensitivity of 96% can even be considered for actual clinical applications.

2.1.4 Conclusion

This paper presented a novel markerless method to identify and classify normal and pathological human walking. The whole strategy consists in extracting the silhouette of the patient for each video frame and use this information for building a motion history image descriptor (HMI). Image moments are used to build a feature vector from the HMI descriptors, which is then used to classify patterns as normal or pathological using a classical Support Vector Machine strategy. The results obtained show this method could complement the conventional gait analysis and a first approximation to a markerless analysis.

2.2 Pathology Classification of Gait Human Gestures from noise observations

Gait patterns may be distorted in a large set of pathologies. In the clinical practice, the gait is studied using a set of measurements which allows identification of

pathological disorders, thereby facilitating diagnosis, treatment and follow up. These measurements are obtained from a set of markers, carefully placed in some specific anatomical locations. This conventional procedure is obviously invasive and alters the natural movement gestures, a great drawback for diagnosis and management of the early disease stages, when accuracy is a crucial issue. Instead, markerless approaches attempt to capture the very nature of the movement with practically no intervention on the movement patterns. These techniques remain still limited concerning their clinical applications since they do not segment with sufficient precision the human silhouette. This article introduces a novel markerless strategy for classifying normal and pathological gaits, using a temporal-spatial characterization of the subject from 2 different views. The feature vector is constructed by associating the spatial information obtained with SURF and the temporal information from a $\Sigma - \Delta$. The strategy was evaluated in three groups of patients: normal, musculoskeletal disorders and parkinsons disease, obtaining a precision and a recall of about 60% *The complete content of this section has been published as a research article in the proceedings of* ***VISSAP-2011*** *(see [93])*

2.2.1 Introduction

Distortion of gait patterns are the first clinical manifestation of many diseases, among others diabetes, brain palsy or accident sequelae. The analysis of human gait attempts to objectively assess pathologies by following up the hidden gait dynamic variables. The set of techniques dedicated to perform this analysis is what is currently known as the gait laboratory, a tool devised to quantify a disease and to compare the gait with normal patterns [126, 224]. Most of this gait analysis is carried out with a set of markers, carefully placed upon some specific anatomical locations. This conventional procedure is invasive and alters the natural movement gestures, necessitating strong variations to achieve diagnosis, i.e., this approach is hardly useful in early stages.

On the other hand, gait dynamic patterns are by nature highly variable and can be easily contaminated with noise. In early stages, most of these diseases differ by very little from what is considered a normal pattern so that classification is a very challenging problem, even for the expert clinicians. This picture may be worsen if one considers that the basic examination tool, the markers, can move very easily or can even be unobservable, contaminating the resulting measurement. These factors together lead to subjective clinical analyses with the consequent limitation in the reproduction of the clinic management of the patient [158, 288].

Ultimately, this problem has undergone a fundamental transformation since the objective is not anymore the movement reconstruction from the anatomical markers, but the accurate tracking of the movement pattern i.e. the markerless strategy. Research areas

as computer vision, automatic surveillance, animation and image processing have already developed some markerless strategies for diverse applications, namely, biometric identification, abnormal motion detection, scene reconstruction and activity classification [166, 266]. However, there are several problems related to extracting the object of interest from some scenarios, mainly due to the blurred boundaries between the background and foreground [64, 204], an issue that can result in wrong characterizations. This article presents an efficient markerless methodology to identify and classify different kinds of normal and pathological movements. A non linear Sigma-Delta ($\Sigma - \Delta$) operator is used to obtain a temporal movement description as a set of pixels. Most of them correspond to a particular patient shape while some small scattered groups belong to the background. Afterwards, we compute a bounding box around of largest group and therein we calculate some local features per frame, using the Speeded Up Robust Features (SURF). A weighting function allows associating some of these spatial features with relevant temporal information. This weighted feature vector is used to classify patterns as normal or pathological, applying a classical Support Vector Machine strategy. Evaluation was performed on a database with 96 videos from 32 patients, with three types of movements: normal, musculoskeletal disorders and Parkinson's disease. Sensitivity and specificity are used to assess the utility of this method.

2.2.2 Gait data

Experimentation was carried out with video sequences recorded from 3 views frontal, lateral and 45 degree view, registered at the gait laboratory of the National University of Colombia, under semi-controlled illumination conditions. This dataset consists of a set of videos captured from 20 patients, each one was recorded 4 times while walking, for a total of 240 video sequences. The Dataset was divided as follows:

- 8 patients diagnosed with musculoskeletal disorders for a total of 13500 frames.
- 7 patients diagnosed with parkinsons disease (No depressive disorder present) for a total of 15500 frames.
- 5 patients with normal gait for a total of 14000 frames

2.2.3 The proposed method

Our proposed method begins calculating the temporal information using a $\Sigma - \Delta$ operator. A bounding box is superimposed upon the region with the largest rate of change and the local features are calculated, within this box, using SURF. A weighting function chooses the more relevant SURF features, those with a similar spatial location

to the pixels detected by the $\Sigma - \Delta$ operator, i.e., the features that contain temporal and spatial information. The obtained feature vector is used to classify patterns as normal or pathological, applying a classical SVM, as illustrated in figure **2-4**.

2.2.3.1 $\Sigma - \Delta$ Temporal Estimator

Temporal description of the patient gait patterns is central at describing structural changes. Many strategies have been proposed already, they are currently known as background estimation methods [88,141,196]. These methods use a sequence of images It and build up a model of the static scene M_t. The model output is an image D_t , where the background is represented by $D_t(x) = 0$ and the foreground is $D_t(x) = 1$.
In our dataset the silhouette extraction is a difficult task because of the similarity between the foreground and the background. Hence we use a non linear $\Sigma - \Delta$operator to obtain a motion descriptor which detects the most probable localization of the foreground. This estimator oversamples a signal at higher rates than the especified by the Nyquist teorem, increasing correlation between the adjacent frames at each pixel [196]. The $\Sigma - \Delta$ operator behaves as a background tracker $M_t(x)$, dynamically up dated by comparing each image It (x) with the current background $M_t(x)$, using a simple updating rule: If It (x) is greater (lower) than $M_t(x)$, then a positive increase (decrease) Δ is performed. The implemented $\Sigma - \Delta$ is shown in the Algorithm 1. Upon the region with the largest movement pattern, we compute a center of mass, on top of which we place a bounding box that contains the object of interest. This process is speeded up using an integral image representation of the original images, reducing the computational cost by 94 % [269]

2.2.3.2 Speeded Up Robust Features (SURF)

Once the bounding box is extracted, we calculate some local features of it using the Speeded Up Robust Features (SURF) descriptor [136]. This descriptor highlights the salient points within the bounding box so that each salient point is described by magnitude, orientation and feature vectors. The SURF method provides invariant image description, allowing a robust representation against illumination, scale and rotation changes, a useful aspect in our problem due to the semicontrolled scenario, different views and patients.
The SURF description is obtained by initially computing the Hessian matrix $H(X, \sigma)$, as follows:

$$H(X,\sigma) \begin{bmatrix} L_{x,x}(X,\sigma) & L_{x,y}(X,\sigma) \\ L_{x,y}(X,\sigma) & L_{y,y}(X,\sigma) \end{bmatrix} \tag{2-1}$$

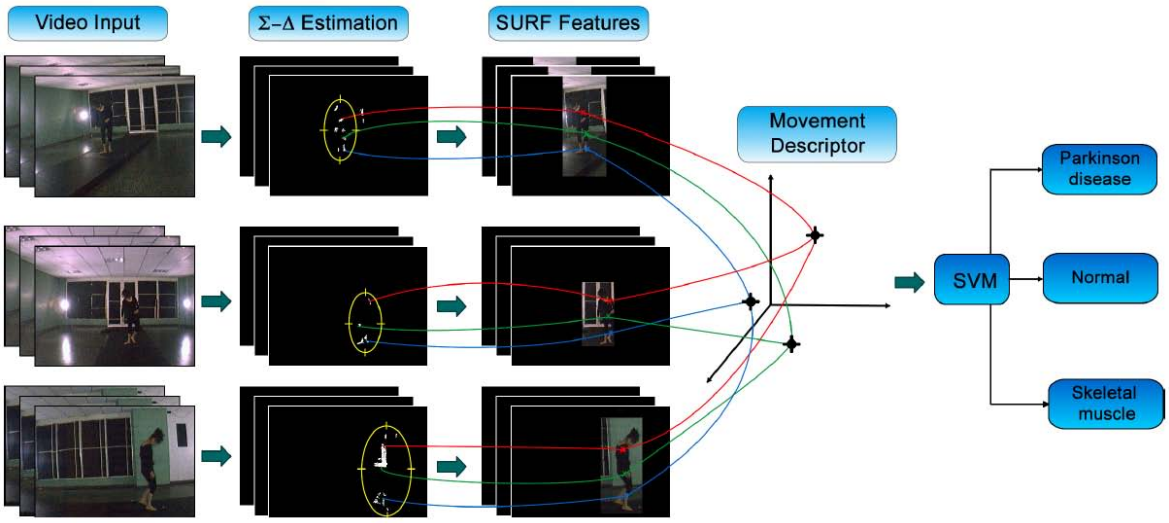


Figure **2-4**: The markerless strategy consists in determining a feature vector to describe normal and pathological movement, using a temporal-spatial gait characterization. Motion is classified using a Support Vector Machine strategy

where X is a specific point, s is the scale and $L_{xx}(X, \sigma)$ is the second Gaussian convolution. This step relies on an integral image to reduce the computational time. Afterwards, SURF constructs a circular region surrounding the points of interest, attempting to assign a unique orientation by estimating the Haar wavelet coefficients in both directions and thereby gaining invariance to image rotations. SURF descriptors are thus constructed by extracting square regions around the points of interest, which are divided in four sub-regions.

2.2.3.3 Feature Extraction

SURF features are used to obtain a summarization of the gait sequence, they operate exclusively on the bounding boxes. Once the set of SURF features is calculated, the values of the SURF descriptor vector are weighted, following the pixel intensity distribution obtained from the $\Sigma - \Delta$ operator. Higher values are assigned to vectors whose locations belong to regions with high movement. The proposed summarization is a collection of weighted vectors, arranged according to their frame number, on the gait sequence.

As the SURF features produce a variable number of points of interest for different sequences, the final descriptor of a gait sequence is obtained at quantizing the complete set of vectors into $5, 10, 20, 40$ and 50 clusters using the Expectation Maximization

algorithm yielding 5 different descriptors for a single sequence.

2.2.4 Experimental Results

Classifcation was performed using a Support Vector Machine (SVM), trained with a set of attribute vectors, extracted from labeled gait sequences. In this phase, two types of kernels were used, polynomial and Radial Basis Function (RBF) kernels. A sensitivity analysis of the parameters, gamma (RBF kernels) and the exponent (polynomial kernels), were estimated using the sequential minimal optimization algorithm [270], the parameter which yielded the larger number of true positives.

Class	**Precision**		**Recall**	
	RBF	Poly	RBF	Poly
M	0.67	0.75	0.33	0.75
N	0.6	0.7	0.95	0.66
P	0.72	0.61	0.41	0.64

Table **2-3**: Table shows the precision, recall and sensitivity for the different evaluated classes, i.e., the musculo-skeletal disorder (M), the normal pattern (N) and the parkinsonian gait (P), using both the RBF and polynomial kernels.

Overall, the SVM strategy shows precision and recall figures above 0.6, except for the musculo-skeletal patterns, for which the RBF is 0.33, a very large difference that can be attributed to the fact that the group of musculo-skeletal is composed of a larger number of patterns and therefore the variance is much larger. The RBF kernel shows a recall of 0.95 for the normal group, indicating that the RBF kernel works better with the data with smaller variance. Of course the fact that the normal group was the larger group (8 cases, compared with 7 and 6) can bias these results, together with the fact that the chosen parameters were set by the fact that they detected the larger number of true positives.

Class	M		N		P	
	RBF	Poly	RBF	Poly	RBF	Poly
M	4	9	5	1	3	2
N	1	2	20	14	0	5
P	1	1	8	5	8	11

Table **2-4**: Confusion Matrix using RBF and polynomial kernels for the three evaluated classes

Likewise the confusion matrix shows that correlation between the normal class is the higher.

2.2.5 Conclusion

This paper has introduced a novel markerless method that allows to characterize normal and pathological human gait patterns. The whole markerless strategy consists in determining a feature vector for describing normal and pathological movement, using a temporal-spatial gait characterization from 3 different views. The feature vector is constructed by associating the spatial information obtained from SURF and the temporal information from a $\Sigma - \Delta$ operator.
Motion is classified using a classical Support Vector Machine strategy. Results demonstrate that this method can complement the conventional gait analysis since it assigns objective pattern measurements. The methodology presented in this work constitutes a first approximation to understanding the complex dynamic of the gait. From this kind of analyzes, we expect it would be possible to set up an assembly of descriptors which allow to accurately describe motions patterns and quantify gait semantics.

2.3 Estimating the size of polyps during actual endoscopy procedures using a spatio-temporal characterization

Polyp size fully determines the surgical endoscopy management since gastroenterologists only send a polyp sample to the pathology examination if the polyp diameter is larger than 10 mm, a measure that is achieved by examining the lesion with a calibrated endoscopy tool. Such measure is very challenging because it must be performed during

a procedure subjected to a complex mix of noise sources, namely anatomical variability, drastic illumination changes and abrupt camera movements. This work introduces a semi-automatic method that estimates a polyp size by propagating an initial manual delineation in a single frame to the whole video sequence using a spatio-temporal characterization of the lesion, during a routine endoscopic examination. The proposed approach achieved a Dice Score of 0.7 in real endoscopy video-sequences, when comparing with an expert. In addition, the method obtained a Root Mean Square Error (RMSE) of 0.87 *mm* in videos artificially captured in a cylindric structure with spheres of known size that simulated the polyps. Finally, in real endoscopy sequences, the diameter estimation was compared with measures obtained by a group of four experts with similar experience, obtaining a RMSE of 4.7 *mm* for a set of polyps measuring from 5 to 20 *mm*. An ANOVA test performed for the five groups of measurements (four experts and the method) showed no significant differences ($p < 0.01$). *The complete content of this section has been submitted for publication to the journal of* ***Computerized Medical Imaging and Graphics***

2.3.1 Introduction

Colorectal cancer is the seventh death cause worldwide [214], a frequently asymptomatic illness characterized by a set of malign polyps along the digestive tract [7,233]. Typically, this disease is discovered during an endoscopy procedure, case in which the polyp size is used as the main endoscopic sign that supports the decision of an immediate resection, i.e., if the polyp is smaller than 10 *mm* [211], it is removed, otherwise a sample is sent to pathology and the procedure is reprogrammed for an extirpation [7,233]. Usually, the polyp size is estimated by measuring the lesion with a linear colonoscopy probe or by using the aperture of the endoscopy forceps as a repair for comparison [114]. This estimation is a very difficult task, highly subjective and dependent on the expert training [159]. In addition, several technical problems may arise during the procedure, such as: 1) optical distortion (Barrel's effect), 2) difficult handling of the endoscope because of the bowel tone and 3) exacerbated physiological conditions like increased motility or secretion [159]. Current advances on video processing open up an actual possibility of identifying polyps during an endoscopy, with some potential advantages, namely: 1) real time estimation, 2) lessinvasiveness, i.e, no additional tool is needed, 3) low cost procedure, and 4) a lesion characterization that might be used as support to the diagnosis. However, an appropriate polyp estimation is yet limited, among others because of the high variability of both the endoscopy procedure and the polyp shape [9, 95], interference light patterns due to the bowel motion and blurred captures because of the varying illumination conditions [159].

Different approaches have been so far proposed for segmenting and estimating the polyp size, under certain geometrical, appearance and size restrictions that often restrain the accuracy of these methods [140]. Aiming to delineate the polyp, Liu et al. [187] simulate a geometrical 3D intestinal tract, computed from a flow deformation map that matches a set of salient points from consecutive frames. Such method results computationally expensive and prone to errors because the salient points are hardly correlated. This strategy only predicts the presence/absence of the polyp, and also it requires an initial manual intervention. In [61], a per-frame polyp shape is estimated by using a set of classical geometrical and color descriptors, a strategy that fails under non controlled illumination conditions. In contrast, Bernal et al. [31] approximate the polyp shape by using per-frame static features that must follow a polyp appearance model. This characterization may fail if the polyp is blurred in the endoscopy video, as usually observed in real scenarios. For estimating the polyp size, Chadebecq et al. [159] introduced a prior RoI bounding the polyp and tracked the lesion using a temporal rigid transformation. Afterward, an infocus blur allowed a RoI size estimation. However, in real conditions the camera movements may be so rapid that the RoI easily losses the polyp.

Recent strategies involve the fusion of video-endoscopy with ultrasound images. Nevertheless, such echo-endoscopy device is mainly indicated in case of extramural polyps, i.e., it makes possible to measure advanced stages of the polyp and its use is highly expert dependent [243]. Other strategies include the virtual endoscopy from CT images [67,81,271], a 3D reconstruction that requires long exposition to ionizing radiation, report low sensitivity rates [67,81,294] and is purely diagnostic and polyps finally must be removed during an endoscopy procedure.

The main contribution of this work is a method that delineates the polyp and estimates its size in a video sequence, using a local spatio-temporal characterization and an automatic defocus strategy. For doing so, a per-pixel motion descriptor is initially built while the camera is moving, assuming only a statistical dependence with the precedent frame. An additional Bayes strategy couples the per-pixel motion descriptor with prior motion information, approximating the shape during occlusion phases. A polyp size is then estimated from the obtained polyp delineation, using a focused estimation of the whole sequence with a calibrated camera model.

2.3.2 Proposed Approach

The present strategy segments, tracks and measures polyps during an endoscopy procedure. An initial manual polyp delineation in the first frame captures the main features to be used. This characterization and the motion history endoscopy coarsely follow the

polyp in the sequence. Afterwards, a classical second order kalman filter, a bayesian tracking strategy, is used to refine the polyp segmentation, obtained from the spatio-temporal characterization. Once a polyp is identified and segmented, the polyp size is computed using an offline depth defocus strategy. The pipeline of the proposed approach is illustrated in Figure **2-5** and further described in the following subsections.

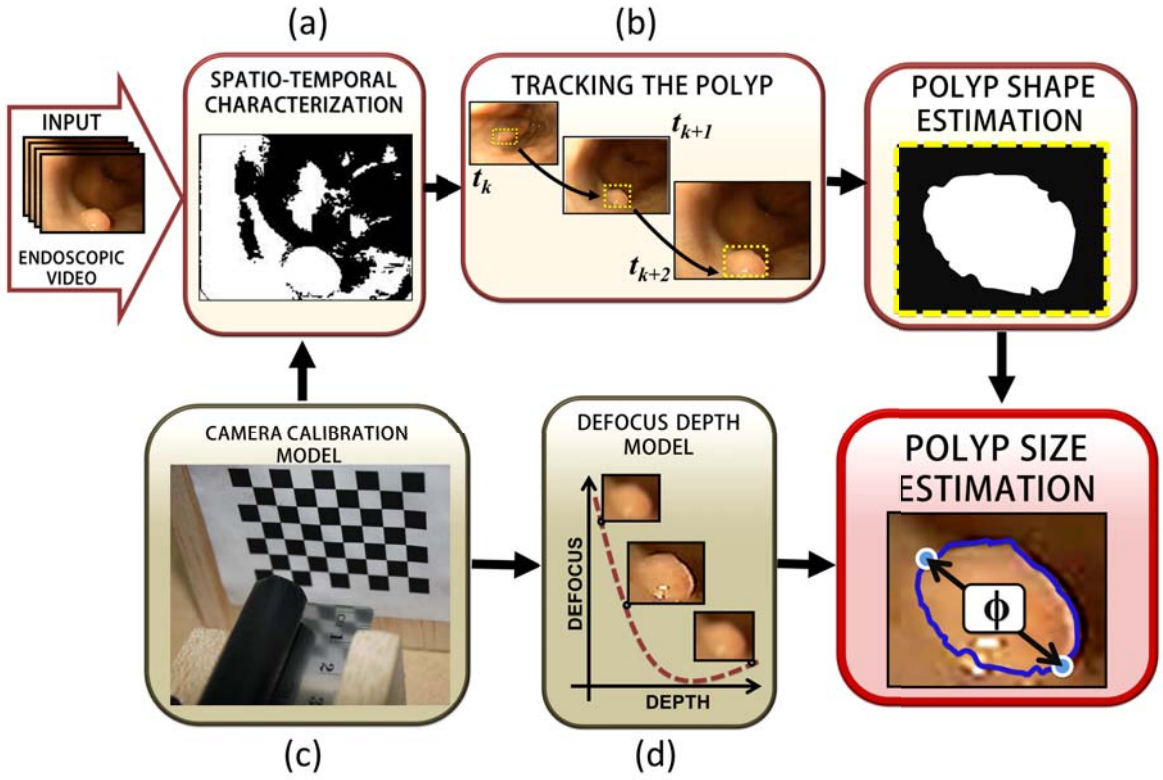


Figure **2-5**: Pipeline of the proposed approach. The method is composed of four main steps. Firstly in (a), a spatio-temporal characterization allows to coarsely follow the polyp. In (b) a polyp tracking strategy was then used to refine the polyp segmentation. Finally a camera calibration model (c) and a depth defocus strategy (d) was used to measure the maximum size diameter of the polyp segmentation obtained.

2.3.2.1 Radial distortion correction

In general, radial endoscopy distortion produces non linear incremental deformations from the optical center to the outer regions, affecting the object relative size and position [184, 206]. The wide-angle lens distortion (barrel's effect) was corrected by estimating the camera parameters using a bank of artificial images (see Figure **2-5**(c)).

Assuming an orthogonal coordinate system, every point in the image space $\mathbf{x}$ is related to the real world $\tilde{\mathbf{x}}$ by a pinhole model defined as $\mathbf{x} = \frac{f}{z}\tilde{\mathbf{x}}$, where f is the focal length and z the distance from the object to the camera lens.
This model estimates the focus camera length, the scale factor, the distortion coefficient and the optical center point. This nonlinear distortion was aproximated by power series and corrected as $r_n = r_d(1 + k * r_d^2)$, being k the radial distortion coefficient and r_d the image with the corrected distortion.

2.3.2.2 Preprocessing and polyp initialization

A polyp is an intestinal protuberance whose appearance may be easily confounded with the surrounding tissues, leading most segmentation procedures to fail. The proposed approach starts by an expert delineation of the polyp contour in the first frame to capture specific polyp features. The polyp contour $\mathbf{X_t}$ is represented as a parametric curve defined as: $\mathbf{X_t} = \left\{ \mathbf{x}_t \ {}_{i=0}^{n}, (\bar{x}, \bar{y}) \right\}$, where $\mathbf{x}_t \ {}_{i=0}^{n}$ is a set of n points contouring the polyp with its centroid defined in $(\bar{x}, \bar{y})$. Such delineation defines a neighbour RoI around the lesion, the RoI_t with size $RoI_t = \mathbf{X_t} + \xi$, being ξ a tolerance value. Afterwards, the histogram of the whole sequence was equalized and a gaussian filter, with $\sigma = 0.7$, was applied to remove the granular noise.

2.3.2.3 Quantifying the polyp motion patterns

During an endoscopy navigation, the expert always tries to track the polyp with translational movements, attempting to generate a depth perception by amplifying the motion of nearby structures[1]. Using a background subtraction strategy, the proposed approach estimates the region with more motion, within which the polyp shape is approximated by those pixels with the largest temporal variance [196]. For so doing, a per pixel history motion $M_t(x, y)$, storing the relative motionless digestive tract, is firstly calculated as $M_t(x, y) = M_{t-1}(x, y) + sign(I_t(x, y) - M_{t-1}(x, y))$, where $I_t(x, y)$ is a particular frame at time t. A likelihood function $\Delta_t(x, y)$ measures the instantaneous pixel motion at time t w.r.t the background history motion, being $\Delta_t(x, y) = \left| M_t(x, y) - I_t(x, y) \right|$. The lesion is then composed of those pixels causally and anti-causally moving during most of the sequence and computed as: $\mathbf{\Delta_t}(\mathbf{x}, \mathbf{y}) = (\alpha_t)\Delta_t^{forward}(x, y) + (1 - \alpha_t)\Delta_t^{backward}(x, y)$, where $\alpha = \frac{t}{N}$ is a temporal weight parameter. Finally, moving pixels that better represent the polyp shape are selected by simple thresholding the estimation $\mathbf{\Delta_t}(\mathbf{x}, \mathbf{y})$ with a learned scalar parameter τ as: $Sb_t(x, y) = \mathbf{\Delta_t}(\mathbf{x}, \mathbf{y}) \geq \tau$, yielding the motion segmentation.

[1] classically known as motion parallax (right-left movements) [240] and kinetic depth perception (rear-front movements) [273]

2.3.2.4 Tracking the RoI

The initial position of the RoI_t that bounds the polyp delineation is then propagated to the rest of the image space and motion history sequences. For doing so, the motion history is correlated for every pair of consecutive frames [181], as:

$$RoI_t(x,y) = \arg\_\max_{RoI_t} \sum_{i=0}^{(n-1)} \sum_{j=0}^{(m-1)} \mathbf{\Delta_t}(\mathbf{i},\mathbf{j}) * RoI_{t-1}(x-i, y-j) \tag{2-2}$$

where $RoI_t(x,y)$ is an estimation of the polyp location corresponds then to that maximally correlated RoI.
Such RoI in the motion history space is mapped to the image space, where a minimal per-pixel euclidean distance w.r.t. the precedent RoI_{t-1}, allows to obtain an additional polyp segmentation $Si_t(x,y)$, the intensity segmentation. Then, an improved segmentation is obtained by fusing the two mentioned segmentations as the intersection of the intensity $Si_t(x,y)$ and motion $Sb_t(x,y)$ (defined in the previous subsection), $Z_t = \ Si_t(x,y) \oplus Sb_t(x,y)$.
Additionally, a classical morphological operator removes the remaining noise, basically groups of isolated polyp regions. Finally, The obtained segmentation is transformed to a polar space, where a simple smoothing preserves the global shape.

2.3.2.5 Tracking the polyp

During an actual endoscopy procedure, a polyp may be missed because of some abrupt camera motions or presence of some digestive fluid that might partially occlude the intestinal tract. With a proper frame-rate capture, it is reasonable to assume a relatively smooth polyp motion, even when the polyp is partially occluded.
A bayesian strategy estimates and tracks the polyp, modeling the probability $p(\widehat{\mathbf{X_t}}\ Z_t)$ of the state of the polyp delination $\widehat{\mathbf{X_t}}$ at time t, given the spatio-temporal observations $\mathbf{Z}_t = (Z_1, \ldots, Z_N)$. This model is assumed markovian, i.e., the current state of the system stores the relevant information. Such Bayesian strategy requires a model of the dynamics $p(\mathbf{X_t}\ \mathbf{X_{t-1}})$ and a likelihood function $p(Z_t\ \mathbf{X}_t)$ that maps the estimated polyp to the spatio-temporal space. Once this information is avaible, the polyp delineation at a particular state, is calculated in two steps:

- *Prediction.* A particular state $\mathbf{X}_t$ is computed by updating the previous belief $\widehat{\mathbf{X}}_{t-1}$, after a prediction given by the Chapman-Kolmogorov equation: $\widehat{\mathbf{X}}_t = \int p(\mathbf{X}_t\ \mathbf{X}_{t-1}) \widehat{\mathbf{X}}_{t-1} d_{\mathbf{X}_{t-1}}$
- *Update.* The predicted belief $\mathbf{X}_t$ is adjusted after observations: $\mathbf{X}_t \quad \widehat{\mathbf{X}}_t p(Z_t\ \mathbf{X}_t)$

For the sake of computational efficiency, a second order kalman filter models the polyp delineation $\widehat{\mathbf{X}}_t$ by using the first and second statistical moments as $\mathbf{X}_t \sim \mathcal{N}(\mu_t, \Sigma_t^2)$, where μ_t is the mean distribution and Σ_t^2 is the covariance matrix of the state t. This kalman estimator is computationally optimal because it linearizes the system with a first order taylor series expansion.

2.3.2.6 Polyp size estimation

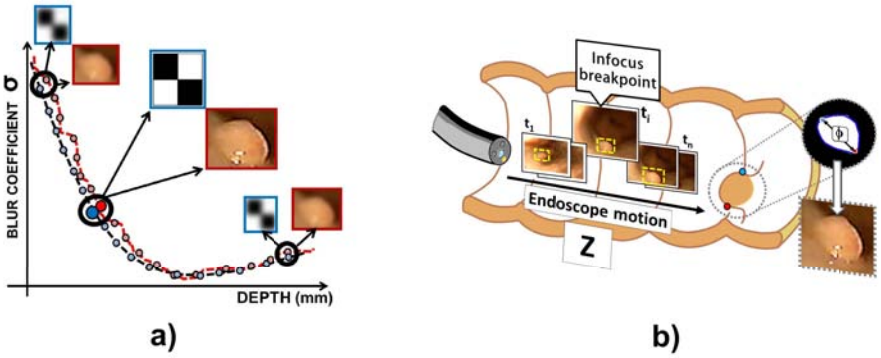


Figure **2-6**: The off-line depth defocus learned model is presented (a) an artificial grid image with known defocus-depth relationship. Using this learned representation, a depth estimation in endoscopy images finds an optimal depth distance (infocus breakpoint) and computes the maximum polyp size (b).

A polyp size is estimated from the obtained temporal segmentation at a fixed depth position of the camera, as a function of the focused image [118] [2] and the pinhole camera parameters.

The depth was herein estimated by a defocus strategy [159] that assumes each frame is contaminated with an unknown gaussian blur with standard deviation σ^o, proportional to the object distance to the camera. This unknown Gaussian blur is estimated by convolving the image with a known Gaussian blur and computing the difference between gradients of the original (unknown blur) and blurred (known blur) images. This gradient ratio R_t is proportional to the unknown standard deviation as $\sigma_t^o = \frac{1}{R^2-1}\sigma_t^b$, where σ^b is a known standard deviation of a blurred gaussian.

In an off-line posterior training step, the blur coefficients, a set of correspondences between the blur levels of a phantom image[3] and actual camera depths (see in Figure

[2] a well known psychophysical theory states that the distance to the camera is a function of the image blur [118]

[3] An artificial grid phantom was adapted for this task

2(a)), are computed. A single blur coefficient is then associated to the infocus breakpoint image $IB_{off-line}$ (the clearer cut-frame) and serves as a reference depth since this is the minimum estimated blurring with a unique depth correspondence. Such relationship - the blur coefficients - was herein used within the endoscopic RoI_t to estimate the unblurred polyp by computing the corresponding infocus ROI breakpoint. The unblurred polyp is then estimated from the segmentation previously obtained.

2.3.2.7 Dataset

The dataset herein used is composed of a set of videos captured using an *Olympus EVIS EXERA (GIF-1TQ160)* gastrovideoscope device, provided with a field of view of 140° and a focal length of $(357.3, 325.5)$. Each sequence was recorded in color, with a spatial resolution of 720 per 480 pixels and a temporal resolution of 30 frames per second. Two different groups of videos were captured for training and evaluation. The first dataset was captured under controlled conditions using an artificial phantom grid superimposed to a set of images captured at different angles, estimating thereby the intrinsic and extrinsic camera parameters. The depth function was trained with captures of the artificial phantom grid, as illustrated in Figure **2-5**(c). The grid is placed at different depth distances, using a custom platform that is displaced in steps of 1 *mm*, with a maximum distance of 30 *mm*. Additionally, as shown in Figure 3, a tubular structure emulated the digestive tract while a set of spheres of known size, the polyps. Four navigations within this structure were recorded to test the proposed approach in controlled conditions. The second dataset included real endoscopic procedures and presence of polyp lesions. Ten videos were captured and four gastroenterologists annotated the videos, segmenting the polyps and estimating their size.

2.3.3 Evaluation and Results

The performance of the proposed approach was assessed in two different tasks: segmentation and estimation of the polyp size. Four expert gastroenterologists delineated and estimated the polyp size in phantom and real endoscopy sequences.

2.3.3.1 Polyp segmentation

Figure **2-8** illustrates the segmentation results in actual endoscopic videos. The yellow contour stands for the ground truth delineation. In the second row, the red polyp delineation is computed using an alternative tracking strategy introduced as a baseline, the classical exponentially weighted moving average EWMA [147], for which an actual polyp delineation is propagated along the sequence using the bhattacharyya coefficient

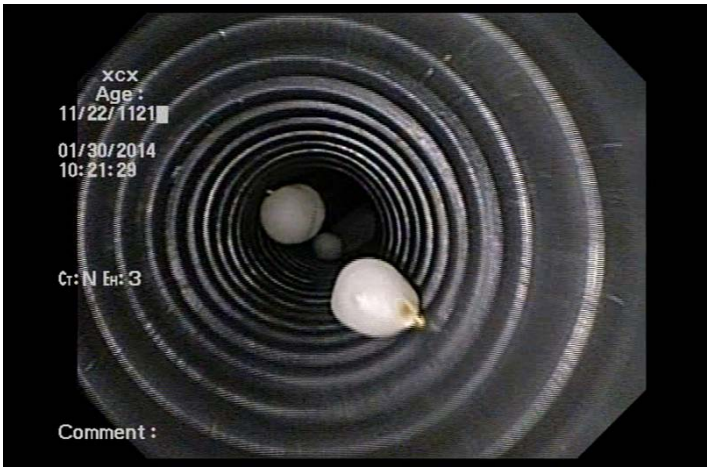


Figure **2-7**: Custom tubular structure that emulates the tract with spheres of known sizes that simulate polyps.

and a set of exponentially decreasing weights obtained from previous frames. Despite this strategy takes into account the motion histoy, and the polyp is relatively well localized within the analysis RoI, the method misses polyp changes resulting from abrupt camera movement, leading to a wrong segmentation. In contrast, in the third row, the blue delineation is obtained with the proposed approach, showing a reliable overlap during those periods with a relatively slow motion. When the camera abruptly moves, the polyp appearance and size result modified, but the proposed approach follows the lesion more accurately than the tracking observed with the EWMA.

Two quantitative metrics were used for assessing the segmentation task: the Dice coefficient (DSC) and the Hausdorff distance (HD). The $DSC(A, b)$ is $\frac{2(A \quad B)}{A+B}$ [78], where A and B represent the obtained polyp area and the expert ground truth delineation, respectively. The Hausdorff measure $H(A, B)$ computes the maximum distance between two sets of points as $\max(h(A, B), h(B, A))$ and $h(A, B) = \max_{a \ A} \min_{b \ B} \ a - b \ {}^{2}_{2}$. In this case, each set of points represents the polyp delineation at each frame. This measure allows to indirectly assess the compactness of the segmentation since outliers are penalized. In videos captured within the artificial tubular structure (see in Figure 4.12(b)) a DSC of 0.96 was obtained when the phantom polyps were segmented, under controlled illumination conditions. Table **2-5** shows the performance obtained by the proposed approach and the EWMA tracking when segmenting 1040 frames.

An additional comparison with the Hausdorff Distance allows to assess the compactness of the polyp delineation, since this measure penalizes those pixels far from the ground truth, reporting in such a case a small value. Overall, the proposed approach outperforms the baseline in terms of overlapping and compactness (small Hausdorff dis-

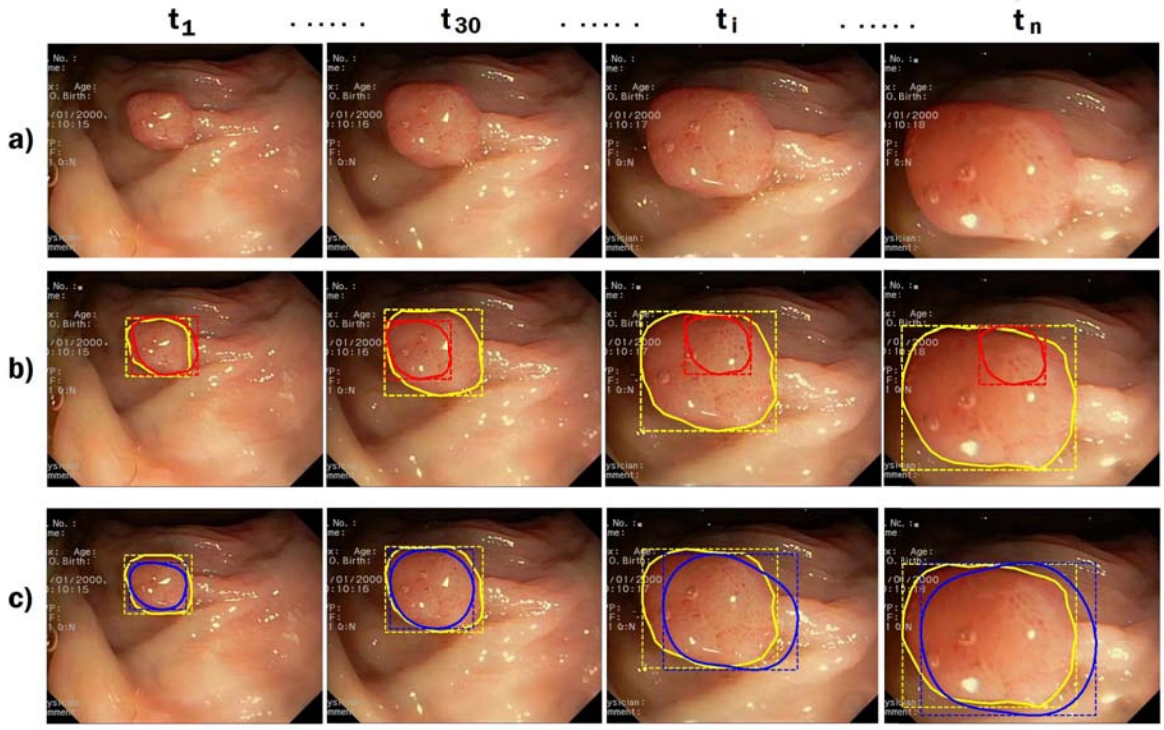


Figure **2-8**: Illustration of a polyp segmentation in a real sequence. First row, the original sequence, second row: EWMA and third row, the proposed method.

Score	EWMA tracking	Proposed approach
DSC	0.52 ± 0.05	0.71 ± 0.12
HD	0.55 ± 0.08	0.38 ± 0.14

Table **2-5**: Performance of the proposed approach using Dice Score (DM) and Hausdorff distance (HD)

tance). Some segmentation errors may be caused a certain polyp occlusion is present or in cases in which abrupt motions may change the appareance, size and shape of the polyp.

2.3.3.2 Polyp size estimation

There exists a well documented high intra and inter expert variability [121], a reported kappa coefficient of 0.4 and an agreement of only a 50.0% [211] in series with three gastroenterologists. In consecuence, a second experiment aimed to evaluate the accuracy of the estimated polyp size. This task is much more challenging because of the

multiple sources of distortion, but also more useful from a clinical standpoint since the gastroenterologist usually has no reference to establish an actual polyp size. Results are shown in Figure **2-9**, the blue lower and upper boxes stand for the spread of the estimated sizes reported by four experts (interquartile range), while the maximum and minimum values are shown as the vertical dotted lines. A total of four endoscopy phantom sequences, with 4 spheres whose size varied between 5 and 15 mm, were evaluated. A part of the experiment required the gastroenterologist to simulate a procedure with similar gestures to an actual endoscopy, the camera moved abruptly and the navigation patterns were complex. In spite of the controlled conditions, experts showed a large variability in their estimation, as illustrated in Figure **2-9**, where yellow diamonds correspond to the ground truth measure per video. Interestingly, results evidence a very large variability of the obtained measures with respect to the reference. In average, the standard deviation was about ± 5.4 mm, confirming the high inter expert variability. In contrast, the proposed metod (green circles) systematically achieved estimations much closer to the actual value. In this case, the method accomplished an accurate segmentation of the phantom polyps and also a proper estimation of the break focus frame.

Overall, it has been traditionally acknowledged that the expert estimation is the most reliable information source in real endoscopy procedures and therefore the ground truth reference. Figure 6 shows the obtained estimations by the proposed approach (green circles) and the gastroenterologists (interquartile range). In average, the gastroenterologists showed a variance of ± 3.63 mm with respect to polyps that varied between 5 and $20\,mm$. In case of real endoscopies, the mean expert estimation -the red line- amounts to the ground truth. As illustrated in figure **4-16**, the estimated size of the proposed approach is within the range of variability observed for the the group gastroentelogists and no significant differences were found when statistically evaluated (Anova test with $p < 0.01$).

The real gain of any method is always contaminated by a particular noise at measuring. The quality of the estimation was herein weighted by the noise as the SNR-like measurement, using a logarithmic scale and measuring the difference between the expected control data and the predicted values. This SNR-like measurement was defined as $SNR-like = 10\log_{10}\frac{\sigma^2}{RMSE}$, where σ^2 is the largest delineation variance among the group of experts and $RMSE$ is the root mean squared error, the computed error of the proposed approach w.r.t. the ground truth estimation. Table 2 shows the results obtained by the proposed approach in terms of this SNR-like measure. In summary, the proposed approach achieves a gain of 37.48 dB, indicating that the proposed approach estimates the polyp, with a high degree of confidence, within the interval defined by the estimations of the experts. Table **2-6** also reports the mean and the standard deviation

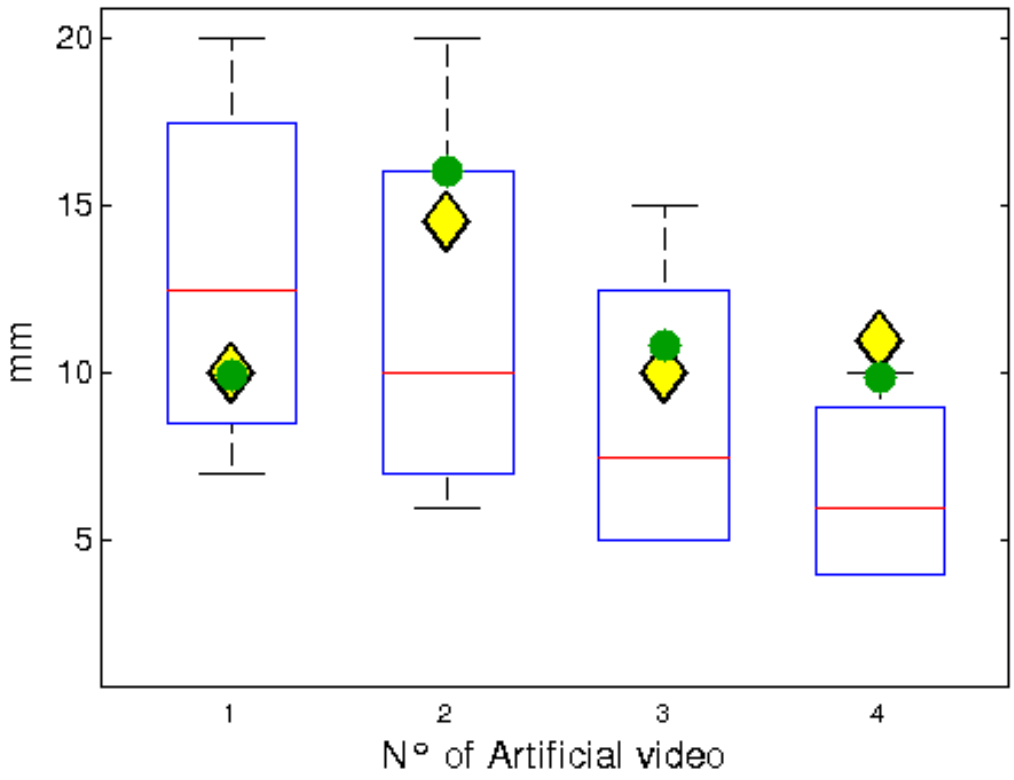


Figure **2-9**: Size estimation of phantom polyps in an artificial tubular tract. The yellow diamonds represent the real measure of each recorded phantom polyp. The box plot represents the obtained measures by a group of four expert gastroenterologists. The green circles show the estimation with the proposed approach in each video.

of the Error (RMSE), indicating a high accuracy of the estimation in artificial videos and a size estimation within the interval defined by expert variability, in real videos.

Dataset	**RMSE** (mm)	**SNR-like (dB)**
Artificial	0.89 ± 0.56	53.8 ± 16.4
Real	4.7 ± 3.2	37.48 ± 8.06

Table **2-6**: Performance of the proposed approach using Root mean square error (RMSE) and the SNR-like measurements

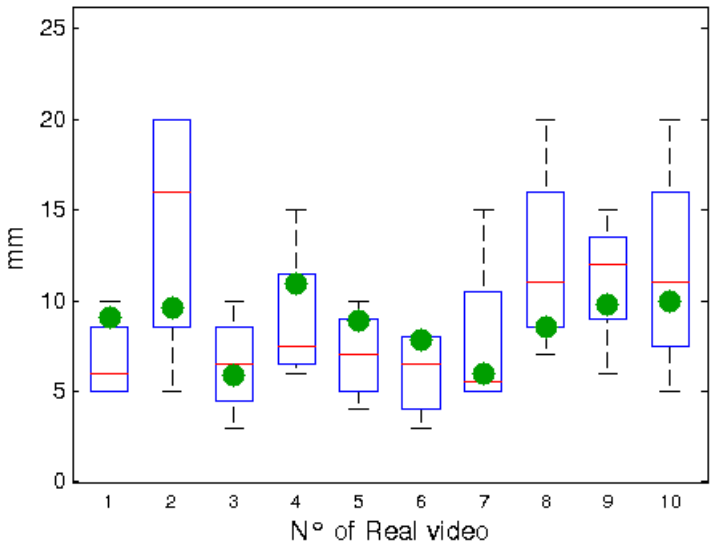


Figure **2-10**: Size estimation by the experts is summarized in the statistical box plots, where the upper and bottom blue lines represent the quartiles (μ= 9.25 mm; $\sigma = \pm 3.63$ mm). The green circle stands for the size estimation obtained by the proposed approach.

2.3.4 Discussion

This work introduces a novel approach that segments polyps and estimates their sizes during video-endoscopy procedures. The method starts with an initial expert delineation that is warped along the sequence by using information obtained from both a motion and an appearance per-frame analysis. The resultant coarse shape is then refined by a second order Kalman filter, a bayesian strategy that uses the motion history as observation. From such segmentation, the maximum polyp size in pixels is computed and then transformed to real-world coordinates by a combination of camera parameters and computation of an optimal depth distance.

Every polyp, found during a colonoscopic procedure, must be extirpated, but those polyps whose size exceeds the 10 mm are sent for further pathological examination [7, 172, 243]. In spite of the demonstrated importance of quantifying the polyp size, the colonoscopic measure so far consists in comparing the observed lesion with micro-scales introduced within the endoscopy tube, a very highly expert-dependent procedure. Hence, reliable, efficient and reproducible measurements that estimate the polyp size are required. As mentioned before, this problem is particularly challenging since the procedure is by nature the result of abrupt motions, the exploration is carried out un-

der uncontrolled illumination conditions and the complex anatomy introduces a huge variability of the polyp appearance and shape. Several computational strategies have been proposed to cope with identification, characterization and measure of polyps. Regarding polyp delineation, Ganz et al. [107] used a multiespectral endoscopic imaging to highlight the region that bounded the polyp, according to certain expected histological properties. Then, the boundary is detected using a prior shape term as regularizer. In terms of overlapping, this strategy achieved an average Jaccard index of 0.52 for the segmentation task. This method requires a special device to characterize the polyp, that is to say, to define the set of histological characteristics that might be associated to the lesion. Hence, this approach results dependent on a very large database that can store the high shape variability. In contrast, the presented approach characterizes the polyp in standard endoscopic sequences, without whatsoever prior shape, achieving an overlapping score of 0.71 over 1040 frames in 10 videos. On the other hand, for estimating the polyp size, Chadebecq et al [53]. proposed a semi-automatic method that started by manually placing a bounding box surrounding the polyp, followed by a conventional affine registration that propagates such initial guess to the whole sequence and estimates the best focused region by a depth learning procedure. Unlike our approach, this break focus is determined on the entire bounding box, with the consequent error coming from calculating the polyp distance as a linear function of the estimated depth within the box. Likewise, the rigid registration is very limited to follow abrupt changes of the camera view.

The proposed approach has presented a framework that performs polyp segmentation and size estimation, during colonoscopy procedures. One of the main advantages of the proposed approach is the adaptability to different polyp shapes, with different appearances, depending only of the required initial characterization. Additionally, the deep estimation was computed only within the polyp regions, obtaining more accurate approximations. The proposed approach was evaluated on phantom and real endoscopy videos, showing in general an appropriate performance in both tasks, polyp segmentation and size estimation, i.e., an overlapping average of 0.82 ± 0.09 for segmentation and a RMSE of $0.89 \pm 0.56\, mm$ in phantom polyps varying from 5 to $20\, mm$ and a RMSE of $4.7 \pm 3.2\, mm$ in real sequences, when estimating sizes. In summary, the proposed approach shows adequate sensitivity and reproducibility so that it may potentially be useful in clinical applications as a tool to support the diagnosis.

2.3.5 Conclusions and Perspectives

This work has introduced a novel approach to segment and estimate the polyp size in colonoscopy sequences. The proposed method uses a combination of local motion

information and appearance polyp features. The results show a reliable segmentation and tracking of the polyp along the video sequence, capable of dealing with video artefacts and high appearance variability. Additionally, the polyp size was estimated in millimetres. In a future work, a strategy to automatically initialize the segmentation will be integrated, based on machine learning strategies while the segmentation will be refined with alternative polyp appearance descriptors.

2.4 A Novel Right Ventricle Segmentation Approach from Local Spatio-Temporal MRI Information

This paper presents a novel method that follows the right ventricle (RV) shape during a whole cardiac cycle in magnetic resonance sequences (MRC). The proposed approach obtains an initial coarse segmentation by a bidirectional per pixel motion descriptor. Then a refined segmentation is obtained by fusing the previous segmentation with geometrical observations at each frame. A main advantage of the proposed approach is a robust MRI heart characterization without any prior information. The proposed approach achieves a Dice Score of 0.62 evaluated over 32 patients. *The complete content of this section has been published as a research article in the proceedings of* ***CIARP-2013*** *(see [17])*

2.4.1 Introduction

Cardiovascular diseases (CVDs) are world wide one of the principal causes of death and disability [239]. An accurate quantification of the right ventricular structure and function has become important to support the diagnosis, prognosis and evaluation of several cardiac diseases and also to complement the typical analysis of the left ventricular function [48, 124]. Currently, most common methods, for assessing the heart, are based on quantification and characterization of patterns during a Cardiac Magnetic Resonance Imaging (CMRI) [24]. Such methods are widely used to analyze, diagnose and even prognose certain heart diseases. Among the evaluated heart variables, the most common are the ventricular chamber sizes, volumes and masses at each cardiac phase, ventricular function and correlation flow [125]. A proper RV analysis demands an accurate 3D temporal segmentation, specifically endocardial and epicardial contours. Typically such task is carried out by expert cardiologists who perform manual delineations that may take $18.9 \pm 4 min$ [226] per case, introducing high inter-and-intra observer variability [49]. Therefore, automatic segmentation methods are appealing to obtain more accurate RV temporal-segmentation. Nevertheless, several challenges arise because of the complex RV geometry shape and high non-linear heart motion during

diastole and systole transition. In addition, RV fuzzy edges and random acquisition noise make more challenging the RV segmentation [226].

Several state-of-the-art methods have been proposed for automatic RV segmentation, most of them based on the use of strong structural and appearance priors that adjust the shape w.r.t a set of samples. In this sense, these methods use mainly statistical shape models, multi-atlas strategies and deformable approximations [226]. These strategies are strongly dependent on how data is learned to build up the prior. However, accurate quantification of certain variables like the ejection fraction depends on the shape changes, particularly important in pathological cases. In addition, such approaches pay a high price when mapping the prior to the MR since the metrics is usually noisy because of the dependency of intensity variations or the pixel spatial distribution to represent the heart, facts that may lead to inconsistent segmentations [226].

On the other hand, methods with no prior are based on appearance and temporal MRI information. Cocosco et al. [57] describe the segmentation of both the left ventricle (LV) and right ventricle (RV), by a simple temporal RoI estimation of major motions and then a voxel classification is performed between RV and LV using morphological operations. However, the simplicity of the temporal descriptor, a simple subtraction between consecutive frames, turns out to be very noisy. In addition, Wang et al. [275] capture information that is shared during the sequence and merge it with a spatial within-frame descriptor, based on a classical isodata algorithm. Nonetheless, RV segmentation may easily overflow the actual borders.

The main contribution of this work is a fully automatic method that uses no prior at all and delineates the RV endocardium contour in 4D MR sequences. The strategy uses both a heart motion descriptor and an estimation of RV shape for each frame of the sequence. Firstly a robust per pixel motion model is introduced to highlight the edges with major changes along the sequence, under the hypothesis that heart is the organ with larger motion. Afterwards, a conventional isodata strategy estimates the heart shape which is superimposed to the edges computed from the motion estimation. The final delineation is set to the intersection between those edges and estimated heart shape. The following section describes the proposed segmentation approach. In section 3, the evaluation and results. Finally in section 4 is presented the discussion on the results obtained and some conclusions.

2.4.2 Methodology

The strategy herein proposed is capable of capturing the temporal RV contours from a spatio-temporal MRI characterization. As widely acknowledged, heart motion is the main biomarker in cardiology, allowing by itself an appropriate assessment of cardiac

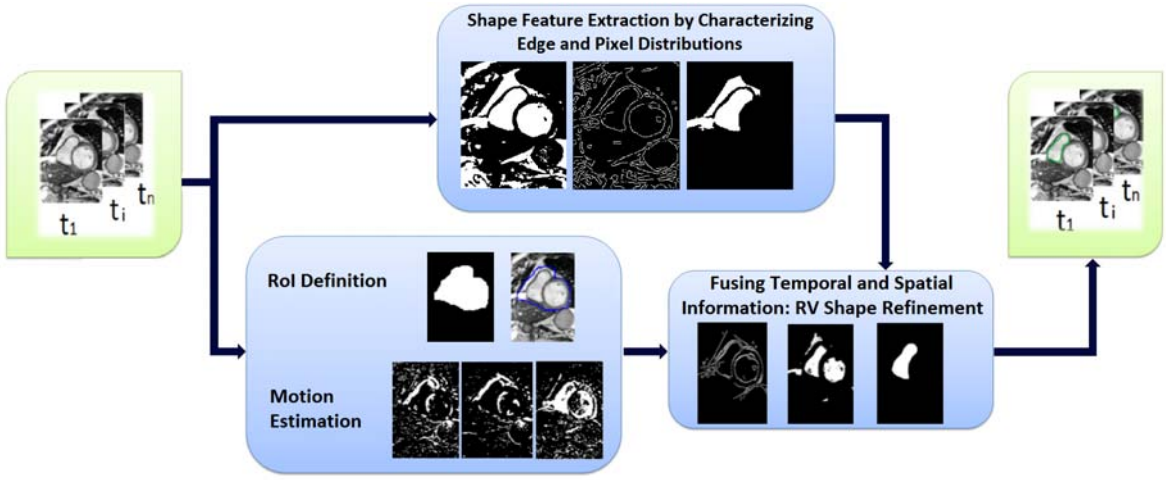


Figure **2-11**: The proposed method. A motion descriptor is computed for the whole MRI cardiac cycle, which is then adjusted to the edged and spatial estimation found at each frame in the estimated shape.

function [230]. Hence, the approach starts by coding temporal MRI information with a bidirectional per-pixel motion descriptor [197]. A coarse heart segmentation is initially obtained from that estimated cardiac motion. This segmentation is corrected using geometrical observations from the estimated shape. The pipeline of the proposed approach is illustrated in Figure **2-11** and described in the following subsections.

2.4.2.1 Motion Estimation

The heart is the organ whose vital function amounts to a constant motion. The proposed strategy starts by estimating the cardiac movement with a bidirectional local motion descriptor. For doing so, a temporal median sets the elements with less motion during the sequence by recursively updating the frame median, as follows $M_t(x) = M_{t-1}(x) + sgn\left(I_t(x) - M_{t-1}(x)\right)$, where $M_t(x)$ represent the median and $I_t(x)$ the frame at time t for each pixel x. Using such recursive median, a likelihood measure Δ_t sets those pixels in movement at each instant t as $\Delta_t = M_t(x) - I_t(x)$. This last term is in due turn regularized by a cumulated variance of the motion history, as: $V_t(x) = V_{t-1}(x) + sgn(N \times \Delta_t(x) - V_{t-1}(x))$. This descriptor is highly noise robust and computes the per-pixel temporal variation that allows to classify the RV. Specifically, At the End of the Diastole, when the heart is maximally expanded, pixel candidates should meet two conditions: the pixel motion is larger than an accumulated temporal variance under the restriction that the movement must span an important percentage of the cardiac cycle. Such relationship is represented by a simple thresholding as

$\widehat{BS_t}^{(D)}(x) = \Delta_t(x) \geq V_t(x)$. In contrast, at the Systole, the heart contraction is maximum and the motion is practically null so that this phase constructs a very steady history of the cardiac flow. After the semilunar valves open, blood flows out the ventricle with an important change that is very likely detected by the motion estimation algorithm. The heart edges are thus calculated from pixels with major motion by comparing the likelihood measure with a learned scalar parameter τ as: $\widehat{BS_t}^{(S)}(x) = \Delta_t(x) \geq \tau$.

Classically, local motion descriptors [197] are usually unidirectional recursive algorithms, but in this case the first iteration yielded a very blurry estimation of the heart contour at the End-of-Diastole. As the recursive accuracy depends on the captured motion history, the descriptor is herein bidirectionally run, i.e., forward and backward as $BS_t(x)^{(D,S)} = \alpha\widehat{BS_t}^{(D,S)}(x) + (1-\alpha)\widehat{BS_{N-t}}^{(D,S)}(x)$, where α is an important parameter defined as $\frac{t}{N}$ and N is the number of frames. Once motion has been thresholded, morphological operators groups up pixels associated with movement [197].

2.4.2.2 Shape Feature Extraction by characterizing edge and pixel distributions

The previous motion estimation produces a coarse shape segmentation and serves also to define a Region of Interest (RoI). The aim of this second phase is to construct another complementary shape approximation, using exclusively spatial observations. A first approximation to such heart shape consisted in finding a RoI that consistently surrounded the heart, as the spatial region with larger temporal motion at each time step. Within such RoI, heart ventricles are estimated from two complementary measurements: a global shape clustering and an edge extraction.

Firstly a global shape description of the ventricles was herein obtained by a classical isodata algorithm [235] that is used to separate the intensities corresponding to the myocardium and the cardiac chambers. The cardiac wall or myocardial tissue is segmented and therefore the right and left heart chambers serve as a reference frame of the right and left ventricles.

On the other hand, ventricle edges are estimated from the MRI RoI by using a conventional Canny filter [43]. In the apical slices however, while the LV is still differentiable, RV edges are very blurred (as shows Figure **2-12**). Overall, edges in apical slices are considered as part of the LV. Estimations of RV edges are performed from the previous motion estimation provided that such edge is not already part of the previously defined LV edges.

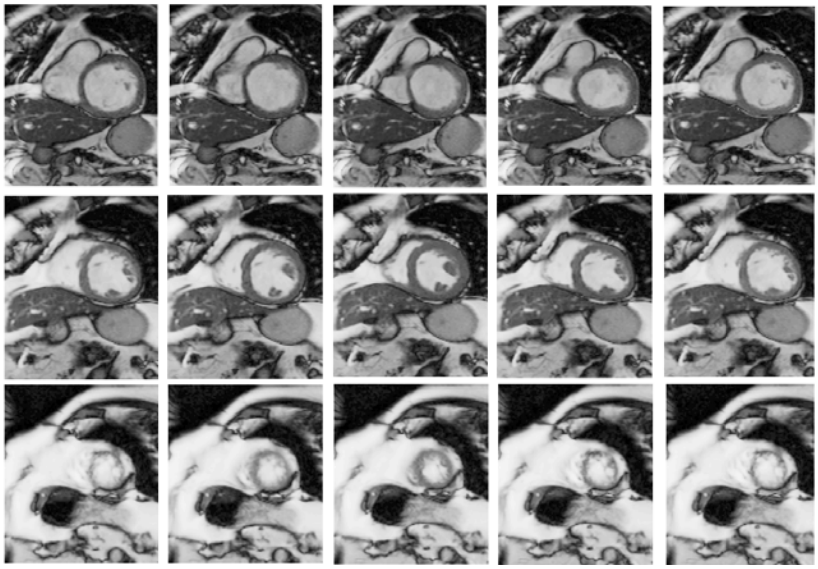


Figure **2-12**: The variability of the RV shape, from basal (top row) to apical (bottom row), and from left to right for the whole cardiac cycle, being the first column the End-of-Diastole and the mid column the End-of-Systole.

2.4.2.3 Fusing temporal and spatial information: RV shape refinement

During certain phases of the cardiac cycle, some boundaries of the heart were not properly segmented. Two fusion strategies were herein implemented to cope with such issue: 1) a first approach fused the spatial information obtained from the temporal information with the edge estimation and the isodata algorithm 2) a second strategy fused the temporal and isodata informations, but using exclusively the left ventricle isodata information. This second approach was particularly useful to segment the right ventricle at the apex level. For the first fusion strategy, the two edge representations are fused by simply summing and normalizing. The final shape is in this case outlined by intersecting both the RV shape estimated from the isodata information and edges. For the second strategy, it was applied a simple difference between the temporal heart segmentation and the spatial LV segmentation obtained by the isodata algorithm so that the remaining pixels then correspond to the RV. Finally, isolated pixels are always removed by simple opening and closing operators.

2.4.2.4 Data

The evaluation of the proposed approach was performed over a public Cardiac MRI dataset [48,84] with 32 patients split into two subsets: training and test data set, which are specified by the authors of the dataset. For evaluation, the obtained segmentation

was submitted to the RVSC [90] which sends back the results. Training data consisted in a set of 16 cardiac MRI, half split into men and women, with an average age of 51 ± 12 years. For test data was split into 3 women and 13 men cases, respectively, with and average age of 48 ± 18 years. The recorded patients were diagnosed with several cardiac pathologies like myocarditis, ischaemic cardiomyopathy, arrhythmogenic right ventricular dysplasia (ARVD), dilated cardiomyopathy, hypertrophic cardiomyopathy, Aortic stenosis, cardiac tumour, left ventricular and ejection fraction assessment. Each MR sequence was captured in the short-axis with 1.5 Tesla, in a plane resolution of 1.3 mm and a between-slice distance of 8.4 mm. The epicardium and endocardium of 32 MR sequences were manually delineated by an expert cardiologist. Trabeculae and papillary muscles were included in the RV cavity.

2.4.3 Evaluation and Results

Figure **2-13** illustrates the good performance of the method in cardiac MRI sequences. The green contour corresponds to the result obtained by the presented method, while the ground truth is drawn in red. As expected, failures are mainly present in apical images because of the fuzzy borders and small RV area.

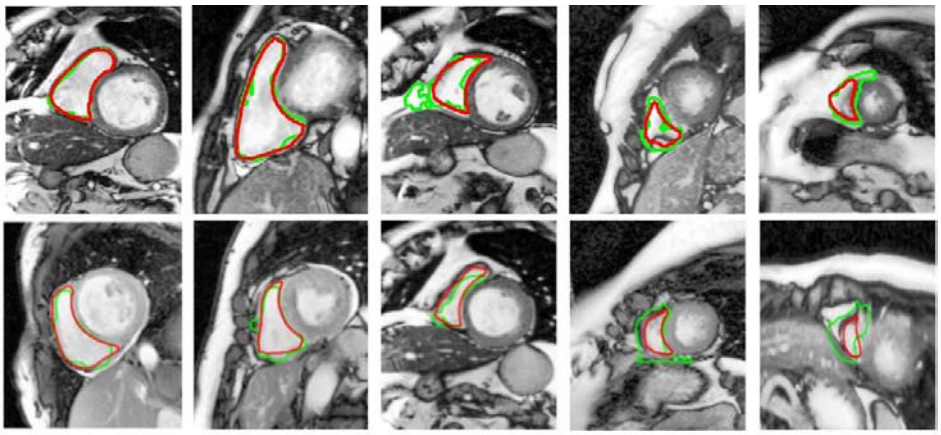


Figure **2-13**: Example of RV segmentations with several cases, including the End-of-Diastole (firts row) and End-of-Systole (second row). The ground truth is the red line and the green line is the automatic segmentation. As expected, better results were observed at the basal slices (first column).

Quantitative technical evaluation was performed using the most classical metrics described in the literature: Dice Score (DSC) measure [77] and Hausdorff distance (HD) [149]. An overlap DSC measure is defined as: $DSC(A, B) = \frac{2(A \quad B)}{A + B}$, where A

and B represent the obtained area and the expert ground truth, respectively. On the other hand, the Hausdorff measure $H(A, B)$ computes the maximum distance between two sets of points as $\max(H(A, B), H(B, A))$ and
$H(A, B) = \max_{a\ A}\min_{b\ B} \|a - b\|_2^2$. In this case, each set of points represents the organ surface. This measure allows to indirectly assess the compactness of the segmentation. A clinical performance was also assessed as the ejection fraction (EF).

	DSC	**HD (in mm)**
End-of-Diastole (ED)	0.66 (0.22)	20.66 (13.00)
End-of-Systole (ES)	0.54 (0.26)	27.72 (23.45)

Table **2-7**: Performance of the proposed approach for training data using Dice Score (DSC) and Hausdorff distance (HD) over the Endocardium contour.

Quantitative results were only evaluated at End-of-Diastole (ED) and End-of-Systole (ES) since these two states are the most important to clinical quantification [113]. As baseline it was taken the work proposed by Wan et. al [275], which until now represents the best strategy to segment the RV without prior. Table 2.4.3 summarizes the obtained performance for training data sequences in ED and ES times. The proposed approach clearly outperforms the baseline method in terms of overlapping and compactness in both cardiac states. As expected a much better segmentation is obtained at the ED because the MRI frame quality allows a better quantification. Although, at the ES many times the poor MRI contrast leads to a quite fuzzy RV edges, the proposed approach outperforms the state-of-the-art approach. Table 2.4.3 summarizes the performance obtained by our approach over the test data. Although the obtained score errors are slightly larger for the RV segmentations, the proposed approach properly captures the shape variability and is easily adapted to new RV shapes since it only depends on the particular MRI observations.

	Proposed approach		**Baseline**	
	DSC	**HD (in mm)**	**DSC**	**HD (in mm)**
ED	0.72 (0.29)	16.17 (16.48)	0.63 (0.32)	22.89 (25.01)
ES	0.51 (0.31)	27.47 (27.96)	0.50 (0.34)	27.99 (24.97)

Table **2-8**: Performance of the our RV segmentation method for the Test data set using Dice Score (DM) and Hausdorff distance (HD) over the Endocardium contour.

Finally, it was calculated the mean error for the ejection fraction, defined as $error =$

$\sum_{p=1}^{N} EFp_{auto} - EFp_{manual}$, where an error of 0.36 was obtained over the whole data set (32 patients). Although the error index shows an acceptable performance, some important noise sources, such as the inter-and-intra high variability of RV shape, the fuzzy edges and the complex heart motion, are not properly captured by our method. Nevertheless, the approach herein presented shows appropiate RV segmentations using an strategy based principally in temporal characterization. This approach outperform state-of-the-art methods that use only appearance and temporal observations for each sequence [57, 275].

2.4.4 Conclusions

In this paper it was introduced a new strategy to segment the right ventricle in MR sequences. The proposed mixed approach uses spatio-temporal observations and produces reliable RV segmentations. A great advantage of the proposed approach is its independency of any prior heart shape, facilitating the capture of dynamic and shape heart variability, which could be associated to specific cardiac pathology. In future work, the method could extend to 3D processing and further validation with a larger data set will be performed.

2.5 Automatic right ventricle (RV) segmentation by propagating a basal spatio-temporal characterization

An accurate right ventricular (RV) function quantification is important to support the evaluation, diagnosis and prognosis of several cardiac pathologies and to complement the left ventricular function assessment. However, expert RV delineation is a time consuming task with high inter-and-intra observer variability. In this paper we present an automatic segmentation method of the RV in MR-cardiac sequences. Unlike atlas or multi-atlas methods, this approach estimates the RV using exclusively information from the sequence itself. For so doing, a spatio-temporal analysis segments the heart at the basal slice, segmentation that is then propagated to the apex by using a non-rigid-registration strategy. The proposed approach achieves an average Dice Score of 0.70 evaluated with a set of 48 patients. *The complete content of this section has been submitted for presentation in* ***SIPAIM-2014***

2.5.1 Introduction

An accurate assessment of the right ventricular function has become crucial for improving diagnosis and prediction of several cardiovascular diseases as well as for complementing the left ventricular evaluation [48]. Currently, this assessment is carried out by the quantification of the right ventricular function during a Cardiac Magnetic Resonance Imaging (CMRI) [48]. Such modality characterizes both the right ventricle (RV) structure and function with a good spatial and temporal resolution [25]. A suitable RV analysis depends on an accurate 3D temporal Endocardium and Epicardium segmentations, usually performed by manual delineation, a time-consuming and experience dependent task. Expert cardiologists may take 18.9 ± 4 min delineating a case [227] while introducing high inter-and-intra shape variability [51]. Several semi-automatic and automatic approaches have then been developed for RV segmentation [21, 192, 274, 300]. Evaluation of the cardiac function is a very challenging task, dependent not only on the particular cardiac motion but also on the RV anatomy, i.e., some portions of the ventricle are really difficult to distinguish, particularly those slices near to the apex which results in blurred edges, already fuzzy, become more blurred also on the heart relations with its neigh going from basal to apical slices. In addition to the complex RV geometry, there exists a well documented variability dependent on factors such as age, cardiac pathology or gender, among others.
Several strategies have been proposed to automatically segment the RV, most of them based on some kind of prior information like the statistical shape models, multi-atlas approaches and deformable strategies. Although these strategies report a relatively accurate segmentation, these methods are computationally expensive and most of them require dedicated hardware, a very difficult issue in a clinical environment. Likewise, the success of atlas or multi-atlas approaches depends on a number of samples that counts for the large RV variability. For instance, Zuluaga et. al [300] propose a multi-atlas strategy based on an exhaustive non rigid registration of a set of templates to the new image, under a similarity criterion of convergence. However, this method is limited by the variability of the different shapes in the set of atlases and the high computational cost. Additionally, Bai et. al [21] proposed an atlas-based approach with a prior introduced as a probabilistic fusion of previously segmented hearts that is used to regularize the rigid non registration of any heart in the database. The obtained model is of course dependent of the type of organs stored in the database. Grosgeorge et. al [117] propose a semi-automatic approach that uses a statistical PCA shape model to select the best representative of the database. The obtained delineations are then used to generate the best segmentation by using a graph cut method. Besides, semi-automatic methods introduce intrinsic expert variability, for instance, Maier et. al [192] proposes a method based on the propagation of an initial segmentation at the

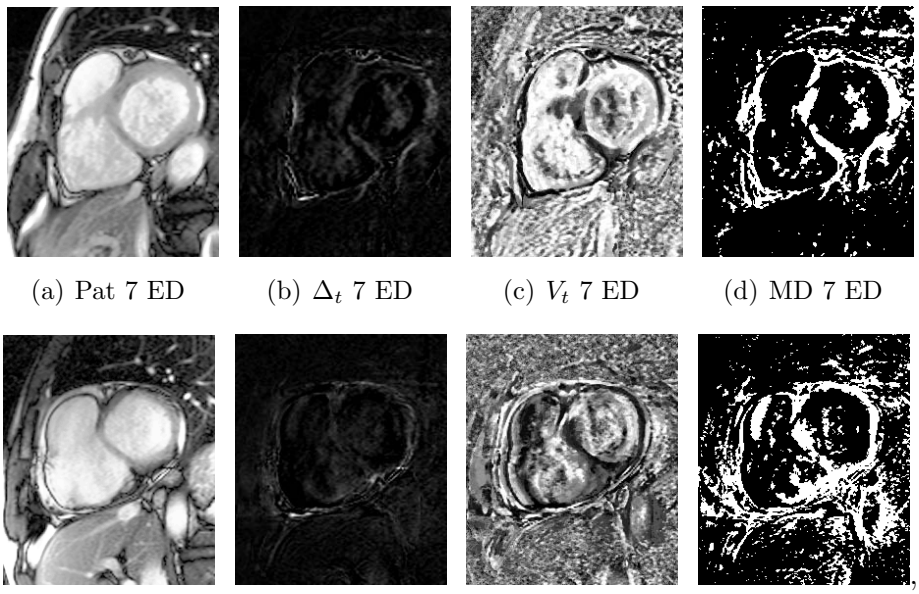
ED to the ES using a region-merging graph cut that is sequentially eroded.
In contrast, Wang et. al [274] have segmented the target using an isodata algorithm. This region-based segmentation is complemented with a basic motion estimation by intersecting consecutive frames. This basic motion estimation nevertheless fails at the apex regions. Ringenberg et. al [238] proposed an appearance-based segmentation as a difference of gaussians combined with a window-constrained accumulator thresholding technique. In this work, a large number of parameters must be tuned and is also affected in the presence of variable pathologies like fallot tetralogy, where the RV shape significatively changes. The method by Mahapatra et. al [191] extracts semantic information from a trained Random Forest classifier, which is also dependent on the training dataset.
The main contribution of this work is a fully automatic method that segments the RV from a spatio-temporal characterization. Unlike typical statistical and atlas (multi-atlases) approaches, the proposed method need not any prior knowledge. For so doing, the most basal heart slice is firstly segmented from a spatio temporal descriptor that tracks those objects with more motion. Such segmentation is propagated, slice by slice, towards the apex by using a conventional non rigid registration, a process that is independently performed for the End of the Diastole and Systole. The rest of this paper is organized as follows: section 2 describes the proposed segmentation approach. Section 3 presents the experimental results and Section 4 concludes the paper.

2.5.2 Method

The proposed method estimates the RV contour by a per-pixel temporal characterization of the basal slice, followed by a local propagation to the apex using a classical non-rigid registration. The following subsections completely describe the proposed approach.

2.5.2.1 Spatio-temporal Characterization of the basal heart

In cardiology, the heart motion is considered as the principal biomarker of the cardiac function [231], specially at the basal region. Under such assumption, a basal cardiac motion is estimated by a per-pixel motion descriptor, which is fused with low level features such as edges and classical shape features to obtain an accurate RV segmentation.
This motion descriptor starts by finding a set of motionless MRI pixels, i.e., a relatively static background, defined as: $M_t(x) = M_{t-1}(x) + sgn(I_t(x) - M_{t-1}(x))$, where $M_t(x)$ represents the accumulated temporal average and $I_t(x)$ the frame at time t for each pixel x. A function $\Delta_t(x)$ measures the instantaneous motion between two consecutive frames as $\Delta_t(x) = \ M_t(x) - I_t(x)\ $ and a coarse ED basal segmentation is defined as the



(a) Pat 7 ED (b) Δ_t 7 ED (c) V_t 7 ED (d) MD 7 ED

(e) Pat 30 (f) Δ_t 30 ED (g) V_t 30 ED (h) MD 30 ED

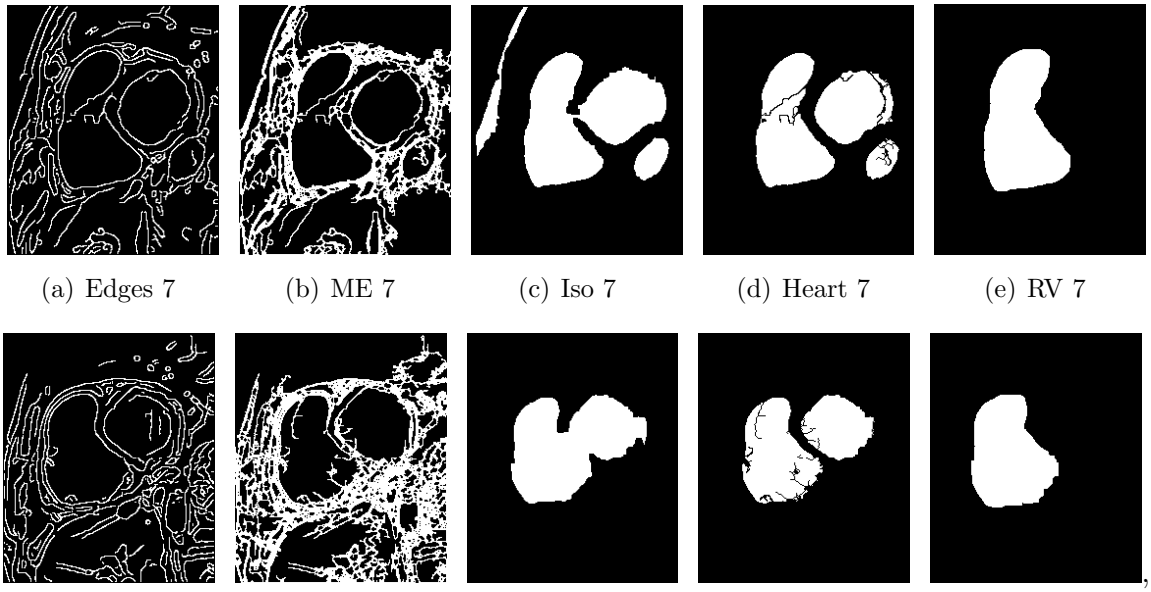
Figure **2-14**: Basal ED motion descriptor from two patients. Upper row illustrates patient 7 and bottom row, patient 30. MD: Motion descriptor.

set of pixels for which $\Delta_t(x)$ is close to a cumulated variance $V_t(x) = V_{t-1}(x) + sgn(N * \Delta_t(x) - V_{t-1}(x))$. Provided that at the ES, the heart contraction is maximum and the motion variance is minimal, a coarse ES basal segmentation is estimated by a simple comparison of the instantaneous motion $\Delta_t(x)$ with a learned scalar threshold τ. Such motion descriptor is bidirectionally calculated [196], while isolated regions are clustered and corrected by using classical morphological operators [196]. This segmentation allows also to define a Region of Interest (RoI) that consistently surrounds the heart (see Figure **2-14**).

At this point, the motion segmentation is fused with basal edge information obtained by filtering the original basal slice with a classical Canny filter [44]. The obtained edges, from the previous step, are then intersected with a third basal slice segmentation computed from a conventional isodata algorithm [236]. Finally, isolated pixels are removed by simple opening and closing operators (see Figure **2-15**).

2.5.2.2 The whole ventricle segmentation

Near to the apex the cardiac motion is minimal and mainly passive so that any motion estimation makes no sense. Provided that the blood flow is also small in this region, segmentation is really difficult because of the complex geometry and fuzzy edges [227]. However, there exists an intrinsic coherence between the basal and apex shapes, i.e., the



(a) Edges 7 (b) ME 7 (c) Iso 7 (d) Heart 7 (e) RV 7

(f) Edges 30 (g) ME 30 ED (h) Iso 30 (i) Heart 30 (j) RV 30

Figure **2-15**: Spatial-temporal information of the basal ED from two patients. Upper, patient 7 and bottom row, patient 30. ME: Heart contour combined with the motion descriptor; Iso: Heart shape from Isodata algorithm.

spatial change of the heart from basal to apex is herein considered as smooth. Under this assumption, the apex is obtained from a series of deformations of the initial basal segmentation by non rigidly registering (NRR) consecutive slices, propagating thereby that initial basal segmentation. Two consecutive slices are non rigidly registered by maximizing the mutual information defined as: $I(S, S+1) = \sum_{a,b} p(a,b) log \frac{p(a,b)}{p(a)p(b)}$, where S and S+1 are two consecutive slices and a S and b $S+1$. The degree of dependency between the slices is weighted by the joint intensity distribution. This metric is adaptively computed under a stochastic gradient descent method [165] that finds the transformation model, defined as: $\hat{T} = \underset{T}{argmax} f(S+1, S)$. This metric has been widely used in many applications, because its relative time efficiency and the robustness in terms of intensity variations [299].

The resulting transformation model $\hat{T}$ is used to propagate the heart segmentation $L(S)$ by B-spline interpolating, defined as $L_{S+1}(x) \equiv L_S(\hat{T}(x)), x$ Ω, where x stands for each labelled pixel. The NRR was implemented in *elastix* software package [164]

2.5.2.3 Data

The proposed approach was evaluated on a public Cardiac MRI data set of 48 patients, supplied by the organizers of the right ventricle segmentation challenge in MICCAI'12 [51]. The data set was acquired from the short-axis with 1.5 Tesla, in a plane resolution of 1.3 mm and a between-slice distance of 8.4 mm, with 20 heart phases for each patient. Such data set have been split in three groups *training*, *test 1* and *test 2*. The experimental group was composed of patients with several cardiac pathologies, namely myocarditis, ischaemic cardiomyopathy, arrhythmogenic right ventricular dysplasia (ARVD), dilated cardiomyopathy, hypertrophic cardiomyopathy, aortic stenosis and cardiac tumour, as well as left ventricular ejection fraction assessment. For each patient, ED and ES endocardium and epicardium contours have been delineated by an expert cardiologist, for which trabeculae and papillary muscles were included in the RV cavity.

2.5.3 Evaluation and results

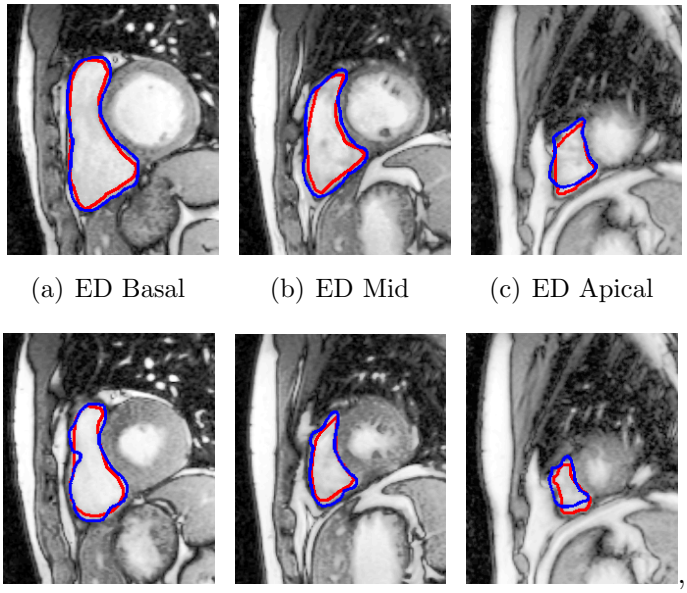
The obtained RV segmentation at ED and ES were evaluated using the classical Dice Score (DSC) measure [79] and the Hausdorff distance (HD) [150]. The DSC defined as $DSC(A,B) = \frac{2(AB)}{A+B}$ measures the overlapping between the obtained segmentation (A) and the manual ground truth (B). On the other hand, the HD establish the compactness of the obtained segmentation as $HD(A,B) = \max_{a\ A}\min_{b\ B} \|a-b\|_2^2$, where A is the evaluated volume segmentation and B the ground truth.

Figure **2-16** illustrates the performance of the proposed method for different slices at different axial locations. The segmentation obtained with the proposed approach is drawn in blue, while the ground truth is in red. Segmentation at Basal slices show a good RV correspondence with the expert, while the apical segmentation shows some slight differences with the expert.

The RV segmentation at the basal slices achieves an average DSC of 0.91. Table 1 shows quantitative results at the ED and ES. As expected, a more accurate segmentation was obtained at the ED. Nevertheless, the obtained HD demonstrates a better segmentation, i.e., without outlier voxels labelled as RV heart, a typical problem in atlas-based approaches.

In addition, quantitative parameters, the ED and ES Volumes (EDV and ESV), were computed and compared with the ground truth using a regression coefficient (R). Table 2. shows the clinical parameter R for sub set Test 1 and compared to State of the Art approaches.

One main contribution of the proposed approach is a method with a very low computational cost. In average, the RV segmentation takes 70 seconds using a Matlab



(a) ED Basal (b) ED Mid (c) ED Apical

(d) ES Basal (e) ES Mid (f) ES Apical

Figure **2-16**: Blue line represents the obtained delineation while the red line is the ground truth at ED (top row) and ES (bottom row) from basal (left) to the apical slice (right).

	Test1 - Training		**Test 2**		**Total**	
	DSC	HD (mm)	DSC	HD (mm)	DSC	HD (mm)
ED	0.74 (0.16)	15.71 (13.6)	0.75 (0.19)	14.78 (9.79)	0.75 (0.17)	15.41(9.0)
ES	0.64 (0.20)	17.0 (8.85)	0.64 (0.26)	18.17 (13.54)	0.65 (0.22)	17.40(10.42)

Table **2-9**: Performance of the proposed method for the dataset. segmenting the ED and ES. Two metrics were used to test the performance: the Dice Score (DSC) and Hausdorff distance (HD).

	OA	**WPC**	**GPR**
EDV	0.95	0.95	0.94
ESV	0.95	0.83	0.82

Table **2-10**: Evaluation of several methods from a clinical standpoint, computing the regression coefficient. Other methods are: WPC - spatio-temporal information [274]; GPR - semi-automatic method using graph cut [117].

implementation, running in a computer with a RAM of 6.0 GB and a 2.4 GHz intel core i7 processor. This issue is crucial for this technology to be implemented in real

scenarios. In contrast, classical atlas based methodologies or statistical methods may take hours to obtain reliable segmentations, a time that grows exponentially, depending on the number of cases in the atlas. On the other hand, although our results are smaller than the state of the art method (0.70 against 0.83 [238]), in terms of the relevant measure, the RV volume evaluation, our method is very competitive. It should be strengthen out that comparison was performed against a single expert and we have not counted the inter-observed variability that has been reported in other problems as about a 30 % [50, 51].

2.5.4 Conclusion

In this paper, a simple but efficient RV automatic segmentation, is proposed. The method takes advantage of a cardiac spatio-temporal characterization. The proposed method captures the shape variability without depending on the nature of the cardiac pathology. The performance of this method will be improved in the future by extending the motion descriptor to the segmentation of the apical slices.

3 Spatio-temporal descriptors from optical flow patterns

Visual systems capture dynamic patterns by integrating different levels of processing, characterization and analysis. Along the dorsal pathway there exist mechanisms charged of processing visual stimuli by representing the motion as optical velocity flow field maps. This representation captures apparent motion patterns, object orientations and stereo reconstruction. This chapter approaches the optical flow representation, capturing dynamic patterns in diverse applications such as: the recognition and classification of human activities (see in section 3.1), the computation of hummingbird dynamic patterns during flight (see in section 3.3), and the global characterization of the heart in cine-MRI sequences (see in section 3.4)

3.1 Representing activities with layers of velocity statistics for multiple human action recognition in surveillance applications

A novel action recognition strategy in a video-surveillance context is herein presented. The method starts by computing a multiscale dense optical flow, from which spatial apparent movement regions are clustered as Regions of Interest (RoIs). Each ROI is summarized at each time by an orientation histogram. Then, a multilayer structure dynamically stores the orientation histograms associated to any of the found RoI in the scene and a set of cumulated temporal statistics is used to label that RoI using a previously trained support vector machine model. The method is evaluated using classic human action and public surveillance datasets, with two different tasks: (1) classification of short sequences containing individual actions, and (2) Frame-level recognition of human action in long sequences containing simultaneous actions. The accuracy measurements are: 96.7% (sequence rate) for the classification task, and 95.3% (frame rate) for recognition in surveillance scenes. *The complete content of this section has been published in* ***SPIE Electronic Imaging.*** *(see [201]).*

3.1.1 Introduction

Human action recognition is the base of many surveillance applications [4, 229]. The challenge is to automatically detect an action occurring in a recorded sequence and to classify selected actions. The great difficulty arises with the extreme variability, in terms of geometry of the scene, people appearance and poorly controlled conditions. Previous action recognition methods are based on motion global descriptors that use optical flow strategies. For instance, Ikizler et al [152] used orientation histograms of a pre-computed optical flow combined with contour orientations. This method can distinguish simple periodic actions but the contour-optical flow integration is too limited to address more complex activities. Chaudhry et al [55] described human activities using histograms of oriented optical Flow (HOOF) with vertical symmetry (i.e. ignoring the difference between motions to-the-left and to-the-right). Such symmetry results in certain invariance but also in limitations to distinguish some actions, for instance antagonist motions of the limbs. Other methods use sparse optical flow, reducing the representation to certain salient features that may not be representative enough to describe a particular action [45, 248]. On the other hand, strategies based on local patch relationships highlight and summarize the motion, for example, $3d$ spatio temporal Haar features, proposed in pedestrian applications [161]. Such descriptors are less sensitive to the quality of the action segmentation but remain strongly dependent on the subject appearance and recording conditions.

The aim of the present work is to recognize human actions using motion descriptors from temporal series of orientation histograms, collected within a multi-layer structure, each layer being a potential human action (temporal RoI). This approach starts by computing a dense optical flow that is then spatially clustered into RoIs which are described as orientation histograms. The histograms are gathered within a multi-layer structure that allows to handle and store temporal information. A motion descriptor can then be extracted at any moment by computing temporal statistics from a particular layer. Finally, the descriptors are labelled as actions using SVM classification. Evaluation was performed with Weizmann [115] and ViSOR video-surveillance [23] databases. This paper is organized as follows: Section 2 details the proposed approach, Section 3 demonstrates the effectiveness of the method and the last section presents the conclusions and possible future works.

3.1.2 Proposed method

In Figure **4-13** the whole method is summarized.

Figure **3-1**: The proposed method starts by computing a dense optical flow, which is clustered as individual motion RoIs. An orientation histogram for each RoI is then calculated and stored in a multi layered data structure. At any time, a vector of characteristics is extracted from every layer and classified using a SVM model to label the action.

3.1.2.1 Optical flow characterization

The computed multiscale dense optical flow estimation [195] consists in projecting every pixel to a feature space composed of spatial derivatives of different orders, at several scales (the local jet). Then, for each frame and every pixel, the apparent velocity vector is estimated by searching the pixel associated to the nearest neighbor in the space of local jet vector at the precedent time. The dense optical flow is firstly used to coarsely segment potential human actions by morphologically closing those pixels whose velocity norm is above a certain threshold and spatio-temporally connecting the resulting regions, according to a distance criterion (see Figure **4-13**, first row). Afterward, for each RoI, a frame-level descriptor is built, based on the distribution of the instantaneous motion orientations. For a non-null flow vector $\mathbf{V}$, let $\phi(\mathbf{V})$ its orientation, quantized to N values. The motion orientation histogram of each RoI frame is computed as the relative occurrence of flow vectors with a given orientation, weighted by their vector norm [68] (see Figure **4-13**, second row) :

$$H_t(\omega) = \frac{\sum\limits_{x;\phi(\mathbf{V}_t(\mathbf{x}))=\omega} \mathbf{V}_t(\mathbf{x})}{\sum\limits_{x;\ \mathbf{V}_t(\mathbf{x})\ >0} \mathbf{V}_t(\mathbf{x})}$$

where $\omega \qquad \omega_0 \ldots \omega_{N-1}$. N is the number of orientations.

3.1.2.2 Multi-layer data structure

The RoI histograms are stored in a FIFO multi-layer data structure, the $x-$ axis being the computed histogram, the $y-$ axis the temporal dimension and the $z-$ axis the actions (layers) present in the video (Figure **4-13**, third row). If a data layer collects more than three consecutive histograms, it is considered as potential motion and then a motion descriptor can be computed. When a layer does not show activity for more than three consecutive frames, it is voided and the corresponding motion is eliminated.

Figure **3-2**: Example of motion descriptors. The blue and gray lines represent the maximum and mean values, respectively. The red square, yellow triangle and green circle represent the mean values for the beginning, middle and end portion of the n histograms respectively

3.1.2.3 Motion descriptor

For each activity layer of the data structure, a motion descriptor is computed. If the layer contains n histograms $H_t(\omega)$, a set of temporal cumulated statistics are calculated for every orientation ω, as follows:

1. Maximum: $M(\omega) = \max_{0 \leq t < n} H_t(\omega)$

2. Mean: $\mu(\omega) = \sum_{0 \leq t < n} \frac{H_t(\omega)}{n}$

3. Standard deviation: $\sigma(\omega) = \sqrt{\sum_{0 \leq t < n} \frac{H_t(\omega)^2}{n} - \mu(\omega)^2}$

Afterwards, the histograms stored in the multi-layer structure are also split into 3 intervals of equal durations, and the corresponding means are calculated. Examples of human action descriptors are shown in Figure 3.3.4. For the initial computation of the motion descriptor, different activities may show similar patterns, because some actions are composed by two or more simple motions. However after 10 frames, the activities are usually well labeled.

3.1.2.4 SVM Classification and Recognition

Finally, the recognition of each potential motion stored in a layer is performed using a Support Vector Machine (SVM) classifier, a *one-against-one SVM multiclass classification* [143] (see an example in Figure **3-3**). As will be shown later, taking into account all the vote values instead of the winner label only is useful to perform time filtering of the recognition. A Radial Basis Function (RBF) SVM model was trained with a set of motion descriptors [54], extracted from previously labeled human activity sequences. A sensitivity parameter analysis (γ, C), was performed under a grid-search using a cross-validation. Additionally, a simple rule was introduced to detect complex activities. If the system detects two simple human actions and they are grouped as a single region, a new activity is defined and tagged as complex.

3.1.3 Evaluation and Results

Experimentation was carried out with different public dat asets. In the first experiment, we tested our approach in an action classification task, using a leave-one-out cross validation scheme. Firstly, we used the Weizmann dataset [115], which is composed of 9 subjects and 10 actions recorded in 93 sequences. The corresponding confusion matrix for the Weizmann dataset is shown in Table 1. The proposed approach achieves an average accuracy of 95%.

Category	bend	jack	jump	pjump	run	side	skyp	walk	wave1	wave2
bend	100	0	0	0	0	0	0	0	0	0
jack	0	100	0	0	0	0	0	0	0	0
jump	0	0	100	0	0	0	0	0	0	0
pjump	0	0	0	89	0	0	11	0	0	0
run	0	0	0	0	80	0	20	0	0	0
side	0	0	0	0	0	100	0	0	0	0
skyp	0	0	0	0	0	20	80	0	0	0
walk	0	0	0	0	0	0	0	100	0	0
wave 1	0	0	0	0	0	0	0	0	100	0
wave 2	0	0	0	0	0	0	0	0	0	100

Table **3-1**: Confusion matrix for the Weizmann dataset.

The proposed approach was also tested with the ViSOR dataset (Video Surveillance Online Repository dataset) [23]. This dataset is composed of 5 activities, recorded in 150 real world surveillance videos. Each video was divided into two parts, to get more examples. A k- fold cross validation scheme was used: for each split, 60 % of the data were used for training and 40 % for testing, obtaining an averaged accuracy of, 96.7%. Results are shown in the confusion matrix (Table 2, top). Performance was also evaluated in terms of classical statistical indices (Table 2, bottom). The obtained results demonstrate both good performance using a very compact action descriptor.

Figure **3-3**: Multiclass SVM voting example for action recognition in a long video. The upper band represents the ground truth for this sequence.

In a second experiment, we evaluated the accuracy of our approach in an action recognition task for 5 long videos from the ViSOR dataset(each one $\sim$ 400 frames). Figure **3-14** shows the performance of our approach in a long video example w.r.t. ground

Category	gc	lo	w	r	h
get car	100	0	0	0	0
leave Object	0	96.67	0	0	3.33
walk	0	0	91.65	8.35	0
run	2.38	0	0	97.62	0
hand shake	0	0	0	0	100

Action	Acc	Sen	Spec	PPV	NVP
get car	98.6	100	96.5	97.7	100
leave Object	98	96.7	100	100	95.2
walk	95	91.7	100	100	88.9
run	94.3	97.6	90.4	92.1	97.1
hand shake	97	100	92.2	95.2	100
Average	96.7	97.2	95.8	97	96.2

Table **3-2**: Top: Confusion matrix for ViSOR dataset. w: walking, r: running, gc: get into a car, lo: leave an object, h: handshake. Every row represents a ground truth category, while every column represents a predicted category.

truth (red line) for two actors. A first raw prediction (blue line), at each frame, achieves an average accuracy of 90.81% with a delay detection time of three frames. Looking at Figure **3-14**, as expected, the major part of mistakes occurs when the motion descriptors is computed with few samples, specially during the transition between actions. Then, a time smoothing of the prediction (green line) was useful to improve the recognition rate. It consists in averaging the SVM votes for each class in a non causal interval $\Delta_t = [t-1, t+4]$ to get a more stable prediction. To take advantage of the stabilizations of the motion descriptor, the filter puts more weight in the future, a strategy whereby we achieved an average accuracy of 95.3%.

Figure **3-4**: Action recognition prediction for a long video sequence with two actors (left: actor 1, right: actor 2). The red line represents the ground truth. The blue line stands for the raw prediction. The green line is the time filtered prediction. The "complex activity" corresponds to the moment when the 2 actors meet.

3.1.4 Conclusions and Perspectives

This paper presented a novel approach for multiple human action recognition, using segmentation of the video flow in individual actions, multiple action representation using a multi-layer data structure, and action classification based on velocity orientation statistics. The motion descriptors can be computed on line and represented with a moderated amount of memory. Using a 96 dimension action descriptor, we achieved an averaged accuracy of 96.7% in human action classification of video sequences and 95.3% in frame level human action recognition. One advantage of the proposed motion descriptor is that it can be extended to mobile camera videos. In future works we will try to extend the recognition algorithm to mobile scenarios, and also build strategies that allow the recognition of more complex actions.

3.2 An overcomplete spatio-temporal multi-scale motion descriptor for human action recognition in video-surveillance from an optical flow characterization

This work presents an overcomplete spatio-temporal motion descriptor that is used in classification, recognition of human activities and interactions for video-surveillance applications. The method starts by computing a dense optical flow that serves to detect human actions which are then marked as Regions of Interest (RoIs). Each of these RoIs is then sequentially partioned, up to obtain an overcomplete spatial representation of small subregions with a different size, which are then used to compute an associated temporal multi-scale descriptor. A particular RoI is then transformed into a set of parts that are characterized along the time. The final descriptor is composed of the temporal estimation of the flow orientation histograms for each part of the overcomplete representation. At any time, the whole motion descriptor can be extracted and labeled by a previously trained support vector machine. The method was widely evaluated using 2 different public datasets with two different tasks: (1) Global classification of short sequences containing individual actions and (2), Frame-level recognition of long sequence human actions, containing more than one action. The ViSOR datasets evaluated task (1), and the method achieved an average accuracy of 95% (sequence rate) while the ViSOR and the UT-datasets were used for evaluating the recognition task, obtaining an average accuracy of 80% (frame rate). *A complete version of this chapter has been submitted for publication to the* ***IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*** .

3.2.1 Introduction

Current video surveillance applications are aimed to automatically segment, identify or recognize normal and abnormal human activities [270]. The recognition task is particularly difficult because of the high variation at capturing and recording, usually under not controlled illumination or occlusion conditions. Characterization of Human activities introduces additional challenges, namely: (1) human activities such as "leaving an object" or " get into a car" share very similar gestures, (2) interactions with other humans or objects may occur, occluding the capture or producing very different dynamic patterns and (3) group activities should be analyzed as a whole [5, 37, 270].
These challenges have been previously approached by characterizing either local or global dynamic patterns of human activities. From a local analysis standpoint, in-

dependent spatio-temporal patches have been proposed to represent human activities. Pedestrian applications have used spatial descriptors, based on a special $3D$ space-time Haar feature with automatic scale selection [162, 180]. Likewise, a $3D$ cuboid descriptor [276] represents activities as a set of patches with maximum gradient responses. An additional Hessian descriptor detects relevant spatio-temporal blobs from saliency measures [286]. The selected patches, processed with these descriptors, are commonly used to classify human activities, using machine learning strategies such as the extended $2D$-bag-of-words [46, 82]. Yet such methods reduce the computed features to a small set of salient points, making this analysis much more efficient, these descriptors are intrinsically dependent on the object appearance and the recording conditions. Actually, the performance of these methods is highly limited by a very variable saliency in real scenes [171]. Global approaches include silhouette based methods, in which temporal variations of the human shape are associated with specific activities [115, 286]. These approaches have reported high scores in datasets with static camera and controlled conditions, but they need accurate shape human segmentation.

On the other hand, dense spatio-temporal and appearance flow motion representations have been widely used in human recognition approaches. Such descriptors have been popular since they are relatively independent of the visual appearance and allow to capture complex patterns of human actions [45, 87, 249]. For instance, Ikizler et al [151] used block-based histograms of optical flow orientations combined with contour orientations. This method can distinguish simple periodic actions, but the motion characterization is still too limited as to deal with more complex activities. Guangyu et al [162] use dense optical flow and histogram descriptors, but their invariant representation, focused on human spatial pattern variations, limits its applications. Chaudhry et al [180] proposed histograms of oriented optical Flow (HOOF) to describe human activities, invariant under vertical symmetry. This approach is yet limited at describing local differences, for instance the relationships between limbs. Currently, convolutional deep models have been applied to constructing features used in action classification, achieving high accuracy in standard datasets [154]. Nevertheless, the well-know high computational cost along with an intrinsic time-delay constitute a critical limitation in real action recognition applications. Additionally, this method is dependent on a bias parameter that has to be learned together with the spatio-temporal kernel weight parameters, a task for which the whole video sequence is required.

The main contribution of this work is an overcomplete spatio-temporal multi-scale motion descriptor, consisting of a set of global statistics that collect information from the temporal history of orientation histograms of a multiresolution and overcomplete representation of a Region of Interest. Such descriptor is capable of determining locations of human actions with potential interest. Each RoI is spatially divided in several over-

lapped sub-regions with a different size that constitute an overcomplete representation. Each of these subregion was temporally characterized by computing flow orientation histogram, weighted by the norm of the velocity, during variable intervals of time. A complete dynamic description is then achieved by recursively computing statistics of the histograms spanning different temporal intervals. The resultant descriptor is used as input to a trained SVM classifier. Evaluation is performed using two human action recognition databases in video-surveillance and real scenarios [23], obtaining a performance that is competitive with respect to the state-of-the-art. This paper is organized as follows: Section 2 introduces the proposed method, section 3 presents tyieldshe evaluation and results about the effectiveness of the method, section 4 present a discussion and the last section concludes with a discussion and possible future works.

3.2.2 The Proposed Approach

Visual systems are naturally entailed with the ability of optimally detecting, recognizing and interpreting visual information, in many cluttered scenarios, using practically the same evolved mechanism [112, 216]. In general, the visual system explores an overfragmented environment and constructs a valid world representation by recognizing a relevant motion when there exists a sort of temporal coherence during a time interval. Hence, a major challenge at analysing any human action is then the optimal duration, during which such analysis should be carried out. This interval is obviously task-dependent on the action sparsity, for instance a walking action may be characterized during very short periods while composed actions, such as getting into a car, may require longer times. This work introduces a novel strategy that integrates several temporal scales of an overfragmented representation of a Region of Interest, which in due turn is determined from a dense optical flow estimation. The method starts by computing a dense optical flow, using a local jet feature approach, from which a spatial coherence during a certain time helps to discover potential RoIs. Each RoI is then divided to obtain an overcomplete representation that allows to integrate local and global dynamic information. Any of these subregions is basically characterized by a set of statistics, computed from the history of the orientation histograms, a complex dynamic structure that, at each time, stores the information flow between consecutive frames.

3.2.2.1 Dense optical flow estimation using Local Jet features

The computation of an apparent velocity flow field has been succesfully applied to recover motion patterns from object orientation [37, 270]. The herein proposed strategy firstly characterizes motion by computing a velocity flow field that is used to determine

potential actions in the scene. Many dense optical flow algorithms might cope with such task, but a multiscale approach is usually more robust to local noise, in particular the nearest neighbor search in the local jet feature space [195]. It consists in projecting every pixel to a feature space constructed with the spatial derivatives of different orders, computed at several scales (the local jet). For each frame t and every pixel $\mathbf{x}$, the apparent velocity vector $\mathbf{V}_t(\mathbf{x})$ is estimated by searching the most similar feature in the local jet space at the frame $t-1$. This method provides a dense optical flow field without explicit spatial regularization and an implicit multi-scale estimation by using regional spatial characteristics. In our experiments, we used 5 scales, with $\sigma_{n+1} = 2\sigma_n$, and three first order derivatives, resulting in a descriptor vector of dimension 15 (Figure **3-5**, first row).

3.2.2.2 Segmentation of Potential RoIs and overcomplete representation

Once a dense optical flow is estimated, the regions with more motion are highlighted, obtaining a coarse spatial segmentation of potential human actions per frame. Specifically, motion regions are connected by applying a simple binary closing operations to those pixels whose velocity norm is above a certain threshold. The resulting spatially connected components are then temporally grouped by a inter-frame distance criterion. The result of this step is then a sequence of RoIs with minimum size and coherent motion (Figure **3-5**, second row).

Afterwards, each of the selected RoIs is partioned as follows, first the RoI is split in two parts, then in four parts, then in six, nine and so on, as illustrated in figure **3-5**, obtaining a set of overlapped subregions with different sizes. Perceptually inspired, such representation captures global and local appareance motion patterns, i.e, high statistical dependence of these subregions [127].

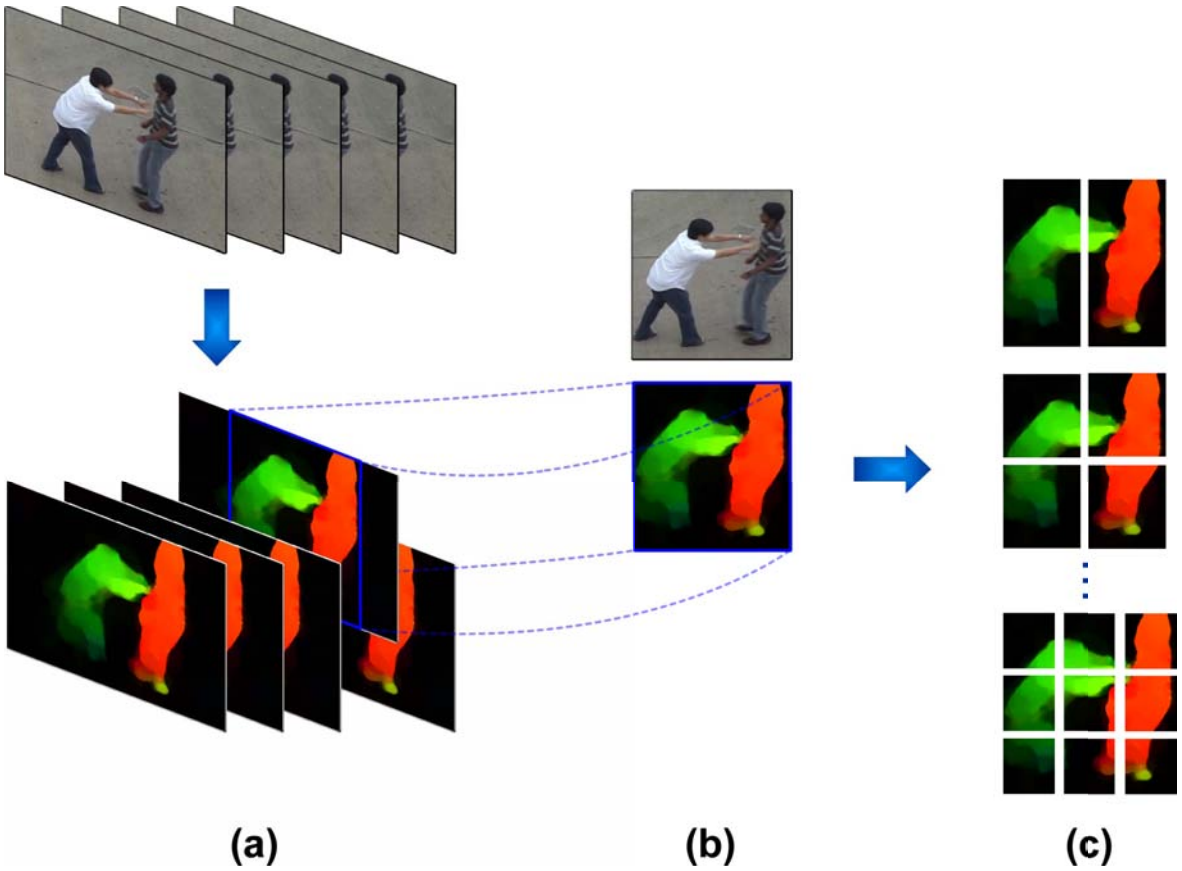


Figure **3-5**: Motion RoI segmentation and overcomplete spatio-temporal representation. In (b) is bounded the region with coheren movement. In (c) the selected RoI is sequentially partioned, up to obtain an overcomplete spatial representation of small subregions with a different size

3.2.2.3 Histogram of Velocity Orientations

For each of these RoI subregions, a per-frame temporal descriptor is built up using the distribution of the instantaneous motion orientations. For a non-null flow vector $\mathbf{V}$ let $\phi(\mathbf{V})$ its orientation or phase. As in the Histogram of Gradients, HOG [68], a per-frame motion orientation histogram is computed as the occurrence of flow vectors with similar orientation, weighted by their norm. A preferred direction may then be the result of many vectors or a single vector with large norm, such histogram reads as:

$$H_t(\phi) = \frac{\sum\limits_{x;\phi(\mathbf{V}_t(\mathbf{x}))=\phi} \mathbf{V}_t(\mathbf{x})}{\sum\limits_{x;\ \mathbf{V}_t(\mathbf{x})\ >0} \mathbf{V}_t(\mathbf{x})}$$

where $\phi \qquad \phi_0 \dots \phi_{N-1}$, being ϕ_N the number of orientations, herein set to 64 (see Figure **3-6**, first row)

3.2.2.4 multi-scale motion descriptor

Visual systems are in general capable of recognizing activities by somehow integrating simple primitives during different intervals of time [216]. It is well known that most of the retina cells respond to transients and basic information of edges. Overall, this information is organized in terms of space and time by cells that are topographically connected to the visual field and that are triggered by variable time stimuli. This characterization allows, among others, to filter specific noise out, to analyze complex dynamics and to recognize objects during variable intervals of information. In surveillance applications, human activities may correspond to very simple periodic motions like some gait actions or more complex actions as the composition of a series of simpler periodic motions. From an optical flow standpoint, periodic activities are repetitions of smooth trajectories that can be characterized during small temporal intervals. In contrast, activities like "get into a car" or "leave an object" are much more complex and involve more than a single periodic activity. In such a case, the computation of global features over the flow trajectory may result insufficient and, in many times, with an important loss of relevant information, as for instance the transition between simple actions. Hence, a successful temporal descriptor should combine the analysis spanning different temporal windows.

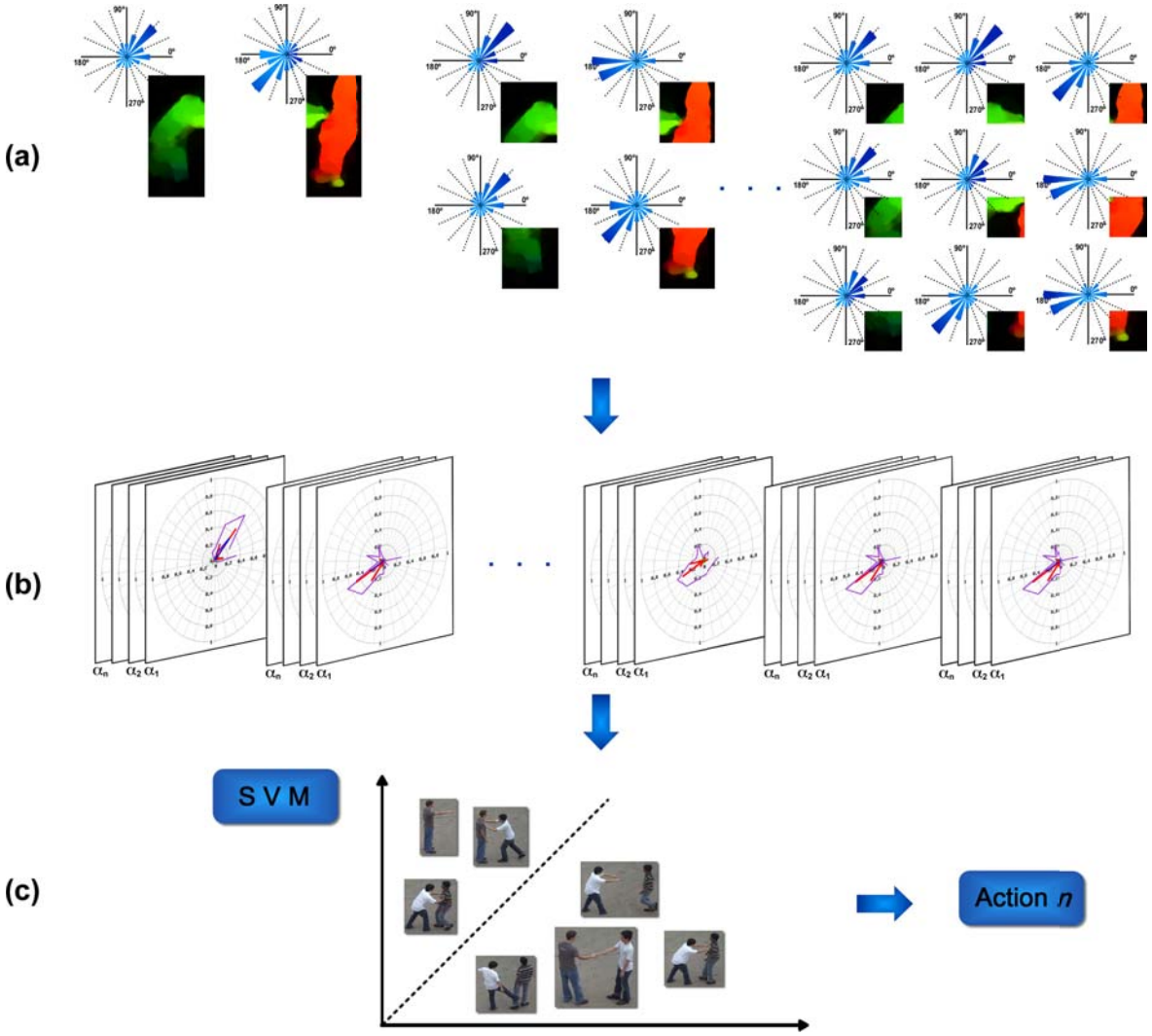


Figure **3-6**: Computation of the multiscale motion description from an overcomplete RoI representation. In (a) the potenatial action bounded in a RoI is sequentially partitionated several times and for each sub-region a motion orientation is computed. In (b) it is computed several recursive statistics for each bin and changing an α parameter to achieve several time intervals of analysis. Finally, in (c) each multiscale-motion descriptor is mapped over a previously trained support vector machine to predict each action.

The new descriptor herein introduced is supposed to fuse information from different time periods. For so doing, a set of relevant motion features are computed during variable time intervals (temporal scales) using recursive estimations as the cumulated statistics from the orientation histograms. Firstly, the mean and standard deviation

are computed from iterative operators as follows:

1. Recursive Mean : $\mu(\phi_j)_t = \mu(\phi_j)_{t-1} + \alpha_i(H_t(\phi_j)_t - \mu(\phi_j)_{t-1})$

2. Recursive variance : $v(\phi_j)_t = \alpha(H_t(\phi_j)_t - \mu(\phi_j)_t)^2 + (1-\alpha_i)v(\phi_j)_{t-1}\text{rw}$

where $H_t(\phi_j)_t$ is the j_{th} histogram bin, computed at time t, $(\mu_{t-1}, v(\phi)_{t-1})$ are the cumulated statistics and α_i is a constant in $[0,1]$ that tunes the time interval. For each histogram bin, two additional statistics were recursively calculated: the maximum and minimum of the temporal histogram, using forgetting morphological operators [234] as follows:

3. Forgetting Max : $M(\phi)_t = \alpha H_t(\phi)_t + (1-\alpha)\max(H_t(\phi)_t, M(\phi)_{t-1})$

4. Forgetting Min : $Min(\phi)_t = \alpha H_t(\phi)_t + (1-\alpha)\min(H_t(\phi)_t, Min(\phi)_{t-1})$

These features require the computation of non-linear dynamic information, complementing the information captured by the mean and the standard deviation. The proposed descriptor combines a set of morphological and linear features recursively computed, for different α constants, describing composed and periodic human activities. Furthermore, the motion descriptor shows interesting properties: (1) little sensitivity to the impulse noise because of the forgetting term, corresponding to the exponentially decreasing weights attached to the past values; (2) periodic and composed motions are characterized by a set of global features computed at several time scales, including both recent and old dynamic information; (3) the recursive computation achieves an efficient me of the memory. Provided that some actions are commonly composed of different simpler activities, the statistics and the multiscale estimation facilitate discrimination of periodic and aperiodic motions.

3.2.2.5 SVM Classification and Recognition

Finally, the recognition of each potential activity is carried out by a Support Vector Machine (SVM) classifier, using the set of recursive statistics in the whole RoI partition, as a spatio-temporal multiscale motion descriptor. This classifier is well known by being successfully applied to many pattern recognition problems, given its robustness, generalization aptness and its efficient use of machine resources. The present approach was implemented by using the *One against one SVM multiclass classification* [144] with a Radial Basis Function (RBF) kernel [54]. For k motion classes, the bank of $\frac{k(k-1)}{2}$ classifiers uses a majority voting strategy to assign the test sample to a particular class. A sensitivity parameter analysis, (γ, C) , was performed with a grid-search using a cross-validation scheme and selecting the parameter with the largest number of true positives.

3.2.3 Evaluation and Results

Experimentation was carried out with two public datasets, commonly used for assessing human action recognition tasks: 1) the ViSOR dataset (Video Surveillance Online Repository), captured by a real world surveillance system [23] and 2) UT-Interaction dataset (High-level Human Interaction Recognition Challenge) which is oriented to identify complex human activities in real world scenarios [247]. The ViSOR dataset is composed of a set of videos showing 5 different human activities: walking, running, getting into a car, leaving an object and people shaking hands. These videos were captured by a stationary camera and contain a different number of actors and activities, occurring simultaneously (examples in actions are shown in Figure 3.2.3, first row). The challenge is related with recognizing individual activities performed by several actors with different appearance, when the scene background differs and also the motion direction may be variable during the video sequence. Experiments with the ViSOR dataset consisted in classifying the different actions in 150 videos with individual human activities. The test was performed under a k-fold cross validation scheme, each fold using 60 % for training and 40 % for testing. The performance herein reported was obtained with 64 directions while the multi-scale motion descriptor was calculated with the α parameter set to 2^{-i}, where i takes the values $4, 5, 6$ for 3 scales and $5, 6, 7, 8, 9$ for 5 scales.

Figure **3-7**: Top row (a) illustrates different examples of the human activities recorded in the ViSOR dataset. Bottom row (b) shows different examples of activities in UT-interaction dataset

The herein proposed strategy has succeeded about recognizing complex human actions in real scenarios. Such spatio-temporal descriptor has been able of popping out the most

Category	gc	lo	w	r	h
get car	100	0	0	0	0
leave Object	0	100	0	0	0
walk	0	0	83.3	16.7	0
run	0	0	14.29	85.71	0
hand shake	0	0	0	0	100

Table **3-3**: Confusion Matrix obtained with the ViSOR dataset, using the proposed motion descriptor with 3 temporal scales. In average, the proposed approach achieves an accuracy of 93.3. Some misclassified actions are walking and running

salient regions in terms of dynamic information by tracking the statistical dependency among the different partitions of the potential RoIs. This strategy was crucial when discriminating composed actions such as getting into a car or leave an object, since the descomposition into motion primitives captured the differences between similar actions.

Table 1 and 2 show the confusion matrices obtained with the ViSOR dataset using a motion descriptor parameterized with 3 and 5 temporal scales, respectively. In general, these results demonstrate a good performance of the proposed descriptor in real-surveillance applications, obtaining an average accuracy of 93.3 and 96.7. As expected, the motion descriptor may confuse similar actions like walking and running, basically because their representation in terms of the dynamic primitives is quite alike. Likewise, table 2 reports some mistakes regarding the composed action getting into a car, probably because it may be associated to the identification of a single simple action, while the action of crouching is not well characterized or is developed during a short interval w.r.t the walking primitive.

Because the recursive nature of the descriptor, action prediction at any time of the sequence, making this descriptor a good candidate to be an online descriptor. For doing so, the motion is computed at each frame and mapped to a previously trained SVM model. Therefore, in a second experiment, we evaluated the accuracy of our approach in an action recognition task for 5 long videos (each $\sim$ 400 frames long) of the ViSOR dataset. In this evaluation, the online prediction perfomed by the herein proposed approach achieved an average accuracy of 92.3%.

In a second experiment, the proposed spatio-temporal motion descriptor was evaluated using the UT-interaction dataset. This dataset contains around six different human interactions, such as: shake-hands, point, hug, push, kick and punch (example of these actions are illustrated in Figure 3.2.3, second row) [247]. The actions included in this

Category	gc	lo	w	r	h
get car	85.76	0	0	14.24	0
leave Object	0	100	0	0	0
walk	0	0	100	0	0
run	0	0	0	100	0
hand shake	0	0	0	0	100

Table **3-4**: Confusion Matrix obtained with the ViSOR dataset using the proposed motion descriptor with 5 temporal scales. In average the proposed approach achieved an accuracy of 96.67

Category	hs	hg	ki	po	pun	pus
Hand Shaking	80	10	0	10	0	0
Hugging	0	80	0	0	0	10
Kicking	0	0	90	0	0	10
Pointing	10	0	0	90	0	0
Punching	0	10	10	0	80	0
Pushing	20	10	0	0	0	70

Table **3-5**: Confusion matrix for UT-interaction dataset No-1.

dataset are much more complex since they show a high interaction level between two people, with a high variability regarding the human appareance and motion patterns during each of these tasks. A total of 120 videos of this dataset were used for assessment, with a 720×480 spatial resolution and 30 fps temporal resolution. A ten-fold leave-one-out cross-validation was performed, as described in [247]. The best configuration of the motion descriptor for the UT-interaction dataset was a RoI configuration of 4, 12, and 20 divisions while three temporal scales were used to characterize each computed sub-region. The scales were computed with an α parameter determined as 2^{-i}, where i takes values of $7, 7.5$ and 8 .

Table 3 and 4 show the confusion matrices obtained when assessing with the UT-interaction, using its two different datasets. In average, it was obtained an accuracy of 81.6 and 78.3 for datasets one and two, respectively. In summmary, the proposed approach achieves a proper dynamic characterization of the different human interaction activities. However, these interaction activities many times are the result of the composition of many complex motion primitives that may occur during a short time interval and whose differentes are not properly captured. Additionally, the complexity

Category	hs	hg	ki	po	pun	pus
Hand Shaking	90	0	0	10	0	0
Hugging	0	80	10	0	10	0
Kicking	10	0	80	0	0	10
Pointing	10	0	0	80	10	0
Punching	0	0	20	0	80	0
Pushing	10	20	0	10	0	60

Table **3-6**: Confusion matrix for UT-interaction dataset No-2.

Approaches	Accuracy UT-dataset 1	Accuracy UT-dataset 2
Propagative voting	93	91
selection of negative	86.7	86.7
Propossed approach	81.6	78.3
Cuboid + svm	75	70
daysy	71	51
SIFT 3D	63	55

Table **3-7**: Average accuracy for different reported state-of-the-art strategies.

of the task makes different sources of noise may easily mislead the descriptor, among others the large number of dynamic actors and their variability when performing a similar task or the actor appareance and the scene variations .
Finally, table 5 reports the results of comparing the proposed motion descriptor with other state-of-the-art strategies. In general, the proposed approach achieves competitive results w.r.t the other stragies, with the side advantage of being easily used for on-line action recognition surveillance applications, since the motion descriptor can be computed at any time during the sequence.

3.2.4 Discussion

This article has introduced a novel overcomplete regional representation for human action recognition based upon the computation of recursive orientation primitives in different temporal intervals. Inspired in visual sytems, the proposed approach carries out a spatio-temporal analysis of the enviroment by defining salient information as those regions that show coherent patterns along the time, thereby achieving the quantification of diverese activities with large variability. The approach, assessed in surveillance

applications, demonstrated to be capable of predicting actions at each time, whereby it results a good candidate to be used in sequences captured with a mobile camera while it is also easy to implement and efficient in terms of accuracy and time.

Action recognition is in general a very active research domain, with potential applications such as human computer interaction, biometrics, heath care assistance, survilleance among others [5]. Specifically in surveillance, the recognition task is oriented to continously identify and categorize abnormal events and generate alarms [270]. In the litetarure, different recognition strategies have been reported, coarsely categorized as appareance based methods, motion fetures based strategies and complex learning approaches. The appearenace based strategies consider the intesity or geometrical characterization of each frame of a sequence which is then mapped to predefined model. For instance in [180], a set of histograms of gradients (HoG), computed at each frame, defined a particular motion. This approach is limited to the characterization of individual actors and sensitive to illumination variations during the sequence. Other approaches use Viola and Jones algorithm [268], and Haar-like features [218] in which the shape is also represented as distribution of edges, with the main drawback of being not robust to orientation changes, issue that is critical in surveillance applications. Overall, these methods describe static shapes, lossing dynamic information that may be useful when discriminating among similar activities. Additional appareance characterization approaches have used spatio-temporal patches computed along the whole sequence. These strategies are however dependent on the detection of salient patches, a high computational cost process. On the other hand, many approaches have taken advantage of the motion modeling by quantifying temporal changes during a video-sequence. Likewise, assuming that human movements can be represented as a continuous progression of the body geometry [35, 145], actions may be classified using any approximation to the silhouettes. These approaches are however limited in surveillance applications since the dynamic quantification is subjected to a proper computation of silhouettes during the time, while motion captured in real conditions is in general contaminated by a dynamic background. Additionally, these approaches may fail when undersegmentation or oclussion occurs, a very likely situation in real video-sequences. To overcome occlusion challenges, in [39], a geometrical shape is tracked using moving average models (ARMA), obtaining smooth estimations of the silhouette during the sequence. Nevertheless, this approach may lose the non-linear information of the geometrical temporal chamges because the proposed tracking acts as a temporal smoothing filter.

Motion description, from dense and sparse optical flow primitives, have been also widely used to characterize and recognize individual and interactive activities since this is relatively invariant to appareance variations and more noise robust [270]. For instance,

in [180] it was proposed a histogram descriptor of dense flow features, invariant to vertical symmetry which was applied to recognize human gestures. This approach however misses local details that might define a particular activity in surveillance applications, for instance the motion relathionship between the limbs. In contrast the herein proposed strategy is easily adaptable to recognize individual and interactive human activities thanks to the overcomplete representation which allows to recover local details that can differentiate similar complex actions. In [205, 237], several recognition strategies have been proposed, based on the computation of different dynamic relathionships using an optical flow and combined with bag-of-features models that aid to determine the occurrence of motion patterns. These methods in general highlight the most frequent dynamic patterns but lose temporal and spatial probabilistic feature distributions. The herein proposed approach in conytrast characterizes the occurrence of dynamic patterns that are spatio-temporally localizated. Currently, some approaches have used learning algorithms to compute dynamic patterns and recognize different human activities from large training sets [20, 29]. In [169], it was proposed a propabilistic hidden markov strategy that models structured social human behaviours and interactions in video-sequences. This approach is however highly dependent on the learned scene context. Deep convolution networks have been also used [154, 259] in real video-sequences, containing different types of actions and achieving a proper performance for the classification and recognition tasks. These methods correlate, at different levels, the appareance and temporal information but at the price of a highly increased computational complexity and processing time. The proposed approach is in contrast flexible to represent from atomic gestures to simple and interaction activities, exploiting the same representation and being relative independent of the scenario. In fact, for simple gesture recognition, the temporal descriptor requires few temporal scales, while for interaction and behavioural activities, the different scales completelly cover the dynamic description. Different temporal scales can be applied by adapting the hyperparameter α during K different periods as: $\alpha_1 < \ldots < \alpha_K$. Finally, the presented framework gains by its simplicity while it is also potentially used in real time applications because of its adaptability to different gestures, a clear advantage in surveillance applications.

3.2.5 Conclusions and Perspectives

This paper presented a novel approach for recognition of multiple human actions and for classification of simple and interactive human activities. The descriptor is based on multiscale velocity orientation statistics computed from an overcomplete RoI representation. This algorithm can run under any of the main stream architectures and

also can be computed on line because of the recursive nature. The proposed descriptor achieved an average accuracy of 95 and 80 in real surveillance dataset (visor) and UT-imteraction dataset, respectivelly. In future works we will try to extend the recognition algorithm to mobile scenarios, and also to build strategies that allow recognition of more complex actions such as group human activities.

3.3 Automatic markerless analysis of the hovered hummingbird flight from a dense optical flow

A new method for automatic motion analysis and characterization of the torsion and deformation of the hummingbird wing is proposed. The method starts by computing a multiscale dense optical flow field, which is used to segment the wings, i.e., those pixels with larger velocities. Then, the whole flight is characterized as a set of global and local temporal measures: an angular wing acceleration as a time function and a local acceleration profile that approximates the dynamics of the different wing segments. Additionally, while the variance of the local velocity orientation estimates those wing foci with larger deformation, a local measure of the orientation highlights those regions with maximal torsion. The approach was evaluated in a total of 93 flight cycles, captured under three different acquisition setups. The proposed measurements follow the hovering hummingbird flight dynamics, with a strong correlation of all computed paths, reporting a variance of 0.31 and 1.96 for the global angular acceleration and the global wing deformation respectively. *A complete version of this chapter has been submitted for publication to the journal of* ***Bioinspiration & Biomimetics*** .

3.3.1 Introduction

Emulation of flying animal movements may facilitate the design of smarter micro air vehicles (MAVs), allowing them to perform more complex maneuvers and expend less energy [198]. Among these MAVs, hovering machines have succeeded in executing elaborated movements within very limited spaces [198]. Designs of such new hovering MAVs may be inspired in animals such as the Hummingbird [10], a bird species able to hover in mid-air. Their maneuverability is based on their ability to easily turn in very restricted spaces, a skill developed to access flowers with sugar rich nectars. A proper design of flapping-wing flight machines requires a precise comprehension of the many interactions between kinematic and morphological wing parameters. Temporal torsion or wing deformation patterns constitute the fundamental variables to estimate the nonlinear relationships between the different wing segments. Moreover, rotations and velocities are related not only with kinematic information but also with forces and torques produced by muscles and tendons [13, 198]. Nevertheless, in spite of extensive studies, many of the morphological, local dynamic, kinematic and even aerodynamic details are poorly understood [198] because of the still limited tools used to capture flight patterns, like for instance flapping frequency, stroke amplitude, angular velocity and even spatio-temporal wing deformation.

Many questions about the flight mechanisms still remain open, e.g., how the humming-

bird accomplishes yaw turns during the hovering flight or how the bird is able to keep turning once he starts this maneuver. In many cases, the only available information about these turns comes from other species like the gliding or flight patterns observed in insects [130,132,157,232,279,280]. However these animals are actually very different since they are entailed with simple wings that operate over a larger range of speeds [258] and are supported by an outer skeleton. In addition, while birds roll their body to orient the wing forces, insects can easily turn by asymmetric left-right wingbeats. The hummingbird flight has been broadly studied using customized analysis systems that require a set of reference points, somehow highlighted and superimposed on the animal structure. These salient anatomical points are usually drawn by an expert [215], adding new variables to the analysis since such delineation is a frame-per-frame task which inevitably leads to high inter and intraobserver variabilities. Furthermore, these tools introduce artifacts that alter the natural hummingbird motion gestures since they need markers that must be placed on the wing bird, markers that in addition can be easily occluded during a wing cycle. In particular, wing deformation has been estimated by following up the temporal patterns of several interest points, manually selected, in the recorded video. However, these measures are indirect and are performed under very strong simplifications [272].

Classical approaches may miss many important details of the real dynamic patterns since they rely completely in both an expert user and an invasive tracking strategy. Recently, some approximations have automatically estimated the dynamics of flight trajectories, most of them using machine-learning methods that require a considerable number of videos for training and finding global flight patterns [257]. Overall, these kinematic methods model the wing as a stiff airfoil [290], but in reality the wing is a very flexible structure, able to asymmetrically produce non uniform movements [290]. Some works have extracted patterns to characterize the hovering hummingbird dynamics [281], supporting the importance of performing analyses of the whole structure rather than following some isolated reference points. A very useful tool has been the Particle image velocimetry (PIV) [3, 36], a strategy that permits to estimate aerodynamic variables by indirect measurement of the resultant flow produced in the air by the wing motion. In practice the set of particles, near the wing boundary, is supposed to closely follow the animal motion. The whole procedure consists then in illuminating such particles with a pulsed light laser and capturing this motion with a camera. A flow field is obtained by cross-correlating consecutive image patterns formed by particles highlighted by the laser [3, 36]. A main advantage of this method is its relative non-invasiveness, but a limitation is that the particle illumination may produce interference phenomena near the wing boundary. Besides, the whole dynamics can not be estimated with the PIV method since the captured flow is related to the external

motion and hence the analysis of forces and torques is only indirect. In other words, this method may improve understanding of the hummingbird movement from an aerodynamic standpoint, but not from a kinematic angle, whose importance is mostly associated to the local wing deformation and temporal rotation. Lately, several researchers have introduced mechanical and mathematical wing models, aiming to exploit some basic physical relationships to simulate different types of wing motions [12,38,183,278]. However, these models are very restrictive, for instance wings are represented as rigid structures and therefore only very linear approximations are possible, whereby the obtained conclusions are limited and the predictive model capabilities are largely reduced [250].

In this work, a novel description of the wing dynamics is introduced, providing previously unavailable information about rotational and deformation wing patterns. This spatiotemporal description is based on a set of velocity vectors obtained from a dense optical flow. Four principal temporal estimations are computed: (1) the global angular acceleration of the hummingbird, (2) the angular acceleration profiles computed over different wing segments, (3) the variance of the spatial orientation of the velocity field for each of the two wings and (4) a distribution map of the local orientation variation within the wing. These four complementary measures allow not only to follow up the wing but also to determine the instants of maximal and minimal deformations, as well as the regions with larger torsion. This paper is organized as follows: Section 3.3.2 introduces the proposed method, section 3.3.3 describes the dataset, section 3.3.4 demonstrates the effectiveness of the method and section 3.3.5 presents a discussion of the proposed work. The last section concludes and presents some possible future works.

3.3.2 Materials and Methods

Overall, a dense optical flow provides a motion description by estimating the instantaneous displacement of each pixel between consecutive frames. Such vector field allows to compute four complementary motion descriptors at each time step, all of them further described hereafter.

3.3.2.1 The Dense Optical Flow Computation

First, a dense optical flow is estimated by matching all the points from frame to frame, using a nearest neighbor search in a projected feature space [195]. The multidimensional feature space is herein obtained using the local jets [177], an approximation to the image spatial structure provided by derivatives at different orders and scales: $f_{ij}^{\sigma} = f * \frac{\partial^{i+j} G_{\sigma}}{\partial x_1^i \partial x_2^j}$, where G_{σ} is the 2d Gaussian function with standard deviation σ. Each pixel is associated then to a feature vector defined as the collection

$\left\{f_{ij}^{\sigma}; i+j \leq r, \sigma \quad S\right\}$, where r is the derivation order and $S = \quad \sigma_1, \ldots \sigma_q \quad$ the selected scales. The set of features, associated to a frame, is represented in a kd-tree structure which facilitates efficient distance search [30]. Finally $\mathbf{V}_t(\mathbf{x})$, the apparent velocity vector for pixel $\mathbf{x}$ at time t is calculated by coupling the feature vector associated to $\mathbf{x}$ to its nearest neighbor in the feature space calculated at time $t+1$. Thanks to the multiscale representation, which implies the matching of that extended spatial structure, this method provides a spatially consistent dense optical flow, without using explicit spatial regularization.

3.3.2.2 Hummingbird motion patterns

Quantifying kinematic flight patterns and their structural relations is crucial for optimal design and development of flapping-wing flight prototypes. These complex relationships may be primarily determined from the analysis of variables like the angular acceleration and temporal wing deformation, herein estimated from the dense optical flow. Such variables are temporal functions, some of them being global scalar functions of t and other local spatial maps indexed by t. Local dynamic patterns are determined by firstly segmenting the calculated apparent velocity flow and finding the two largest connected regions, corresponding to the right and left wing and denoted $\mathbf{V}_{\mathbf{r}t}$ and $\mathbf{V}_{\mathbf{l}t}$. The isolated remaining flow areas are assigned to the nearest region, according to the spatial Euclidean distance.

3.3.2.2.1 Angular Acceleration patterns Rotational wing patterns are crucial to understand complex flight mechanisms, like: particular rotational patterns, an optimal angle of attack, a large downstroke span ratio and a high wingbeat frequency that together maintain the achieved equilibrium and facilitate complex maneuvers [80, 263]. In this work, the angular acceleration was estimated by associating the orientation of the computed apparent velocity optical flow and the wing, under the assumption that the rotation axis is perpendicular to the optical axis of the camera [138, 267]. Two configurations were herein proposed to analyze the hummingbird wing motion from an angular acceleration standpoint: a global angular acceleration of both wings and a temporal map of regional angular accelerations, constructed by estimating the angular acceleration profiles for a set of wing segments which correspond to a wing anatomical partition.

The Global Angular acceleration: A main global description of the hummingbird flight is the mean angular acceleration. For a non-zero flow vector $\mathbf{V}$, let $\phi(\mathbf{V})$ be the apparent velocity flow orientation, the averaged spatial orientation $\overline{\phi(\mathbf{V})_t}$ at each time t is calculated. Under the assumption that there exists a dominant

orientation of the velocities, the rate of change of the velocity angle is then computed as $\Omega_t = \frac{d\overline{\phi(\mathbf{V})_t}}{dt}$. This variable describes the turn angle during the flight, and highlights the relaxation and effort phases during the flight.

Angular acceleration spatial profiles: A local angular estimation was herein obtained from a temporal acceleration profile that follows the wing region $\mathbf{s_t}^{r,l}$ along a flight cycle, where $\mathbf{s_t}^{r,l} \quad \left\{ {s_t}^{r,l\ 0}, {s_t}^{r,l\ 1}, \ldots, {s_t}^{r,l\ i}, \ldots, {s_t}^{r,l\ N-1} \right\}$ stands for the set of N wing segments (Figure 1 (d)). Each of the two $\mathbf{s_t}^{r,l}$ regions, associated either to $\mathbf{V_t^r}$ or $\mathbf{V_t^l}$ flows, is bounded within a bounding-box at a particular time t. The bounding-box is rotated by an angle $\theta_t^{l,r}$ which corresponds to the main direction defined by the wing and then split into N spatial segments. The temporal series is constructed by associating a wing region to the closer region in the precedent frame. For so doing, the metrics is defined as the minimal distance $d_{t+1}^{r,l}$ between the centroids of consecutive wing regions $(\overline{x}, \overline{y})_{\mathbf{s_t^{r,l}}}$, as follows: $d_{t+1}^{r,l} = \left|\left|(\overline{x}, \overline{y})_{\mathbf{s_t^{r,l}}} - (\overline{x}, \overline{y})_{\mathbf{s_{t+1}^{r,l}}}\right|\right|_2^2$. Finally, a local angular acceleration for each segment s_i^t was computed and its temporal map $\Omega(s_i)$ was obtained. Such representation is:

$$\Omega(s_i(\mathbf{x}, t)) = \left\{ \frac{d\overline{\phi(\mathbf{V}_t(\mathbf{x}))}}{dt}; x : \mathbf{x} \quad s_i \right\} \tag{3-1}$$

3.3.2.2.2 Wing deformation patterns Although hummingbird kinematic patterns are generally similar to the insect flight, the morphological structure of the wings are very different given the mass and structural relationship, further considering that the wing muscle mass is largely determined by the need of producing efficient mechanical power. This is why the impact of the wing deformation during the flight must be quantified for understanding how the wing geometry supports forces and produces motion. For instance, it has been documented [281] that the hummingbird upstroke and downstroke are not symmetrical. Such observation is strongly related to the local loads supported along the wing.

Wing deformation is herein estimated by computing the variance of the local velocity along the time. This estimation is carried out by assuming that every flow vector should be equally distributed along a rigid wing, while deformable structures produce a non homogeneous distribution, whereby the local motion variance is proportional to the physical deformation.

Local map of the wing deformation: The wing flow and its averaged orientation are used to compute a per-frame deformation map, estimated as a squared difference between each non-zero flow vector and the average orientation:

$$Q_t^{l,r}(\mathbf{x}) = \left|\left|\phi(\mathbf{V}_t(\mathbf{x})) - O_t^{l,r}\right|\right|_2^2 \tag{3-2}$$

This analysis provides a stress map of the wing, highlighting the forces shearing the wing [1].

Global wing deformation: For each wing, a global deformation $W_t^{l,r}$ is estimated [1] by averaging the local wing deformation map $Q_t^{l,r}(\mathbf{x})$. This measure allows to detect accurately the largest deformation instants during a flight stroke.

3.3.3 Dataset description

Four annas hummingbirds (*Calypte Anna's*) were recorded. Each hummingbird was placed inside a flight chamber that contained a wooden perch in the corner and a feeder, in which a 1 ml syringe was mounted at the end of a metallic arm ($0.22m \times 0.19m \times 0.15m$). The feeder arm was connected to a stepper motor (MDrive 23 Plus, Intelligent Motion Systems, Inc.) placed at the centre of the cage roof. This setup allowed the bird to feed from the spinning feeder, while maintaining its centre of gravity in the same spot, executing a pure yaw turn. A white cardboard square was placed above the hummingbird to offer a contrasting background to the hummingbird body. The hummingbirds were trained to follow the feeder at 30 rpm. A stationary hovering flight was recorded using a high speed camera (Fastec Imaging, Troubleshooter) with spatial resolution of (640×480) and a temporal resolution of 1000 $\frac{frames}{s}$.
In this study, three different sets of videos recorded the hummingbird flight from three complementary perspectives (see Figure 3.3.3). A total of 93 flight cycles were herein considered. Additionally, two high spatial resolution videos (1024×1024), capturing a single wing, were recorded.

Figure **3-8**: In (a), (b), (c) is shown the three complementary configuration perspectives used to record the hummingbird flight. In (c) is illustrated the computation of the N wing segments using as reference a rotated bounding box

[1] where l stands for the left and r for the right

3.3.4 Results

The set of motion measures, herein proposed to analyze the hummingbird hovering flight, provides a valuable spatio-temporal wing description from local and global perspectives, highlighting the maximal and minimal moments of the angular acceleration and the wing deformation. The introduced analysis framework, allows for a time-resolved analysis of wingbeat kinematics, promoted by a direct wing kinematic analysis, promoted by a dense optical flow that facilitates a spatio-temporal tracking of a particular wing region. In addition, the proposed work is a markerless strategy so that the natural flight gesture is not altered.

(a) hummingbird flight video sequence

(b) Dense optical flow computation during the hummingbird flight. The optical flow field in each frame it is represented by a set of green arrows.

(c) De-

An accurate estimation of the apparent velocity field facilitates the extraction of any motion pattern. However, understanding of the dynamics performance along the flight requires a proper balance between information collected at the global and local levels, a task for which a conventional optical flow is insufficient, because it performs pixel matching at a very local level, obtaining density through explicit spatial regularization. In contrast, the multidimensional dense flow herein implemented, turned out appropriate because it performs matching at the object level, and no explicit spatial regularization is needed. A typical multidimensional dense flow estimation of the hummingbird flight is shown in Figure 3.9(b), illustrating that the apparent velocity vectors are mainly within and around the wings. For the experiments presented hereafter, the dense optical flow was parameterized as follows: 7 scales with a $\sigma_0 = 0.3$ and $\sigma_{n+1} = 2\sigma_n$, while up to the second order derivatives were used, resulting in a descriptor vector of dimension 42.

Figure 3.9(c) shows the temporal deformation map $Q_t(\mathbf{x})$ as a function of time and space $(\mathbf{x}, t)$ obtained during a stroke cycle. The color scale represents the magnitude of the local deformation along the wing. Such deformation maps, estimated from the velocity fields, are totally coherent with the classical bio-mechanical descriptions of the musculoskeletal hummingbird structure: during the first stroke phase, the shoulder and pectoral muscles support the main energetic expenses to drive the wing forward [13]. This action produces an initial wing torsion, basically controlled by the proximal wing, as illustrated in the first panel of the deformation map(Figure 3.9(c)). During the intermediate stroke phase, the wing deformation is propagated to the distal segments, which lack of any skeletal structure and whose maximal deformation occurs at about 60 % of the stroke, as observed in the mid panel of the Figure 1. The final cycle is completed by backwardly extending the wings, as illustrated in the last panel of figure 1, the passive part of the movement which is may generated by the elastic wing component.

The estimated global angular acceleration Ω_t correlates with the wing deformation and was calculated as the derivative of the averaged apparent velocity orientation of the apparent velocity flow. All the computed wing deformation paths, computed from each video, were normalized between 0 % and 100 % of the stroke cycle. Each stroke cycle in the video was manually selected and phase-aligned by an expert. The pattern of a cycle was estimated by determining the correlation among the experimental group: for the whole set of 90 recorded sequences and for each instant t, a statistical boxplot was computed, within which the red line represents the median of the angular acceleration, the blue lower and upper boxes stand for the spread of data (interquartile range) and the maximum and minimum values are drawn as the vertical dotted lines, as illustrated in figure 3.

Figure **3-10**: Global dynamic hummingbird patterns computed from the videos with the **C** camera configuration.

Overall, the temporal series follows a very similar pattern for each of the recorded wings, with minimal variance in the passive phases and larger differences for the active parts of the cycle. The observed acceleration peak between the 40% and 60% percent of the stroke, represents two complementary states: an initial relaxation, when the wings are forwardly driven to join, and a second phase characterized by an important effort to separate them. During each stroke cycle, it has been well documented that there exist three complementary phases, namely upstroke, downstroke and the transition between them, as observed in Figure 3. Determination of both the energetic cost and the dynamic patterns for each one of these phases demands an independent analysis of the three phases. Upstroke and downstroke phases correspond to the $[10\,\% - 40\,\%]$ and $[60\,\% - 90\,\%]$ of the temporal series, respectively. In such intervals, the flight is basically passive and smooth because motions are mainly driven by the inertia and forces are typically reactive. The resultant variance in these intervals is about 0.1871 $\frac{rad}{frame^2}$, a figure that can be considered as negligible, i.e., a high correlation is observed among the whole experimental group. In contrast, the mid part of the temporal series reveals a larger variance because of the transition between the upstroke and downstroke phases. During such interval, the anatomical differences as well as the capturing noise contaminate the observations, which results in larger differences between the wing dynamics amongst the whole experimental group. In other words, during a cycle the amount of expended energy is different.

Figure **3-11**: Local angular acceleration profile by the subdivision of the wing in eight segments.

A local angular acceleration profile was computed by dividing the wing into segments, the angular acceleration being regionally estimated for each segment. Figure 3.3.4 shows a typical 3D-map acceleration profile using eight wing segments. Each computed map reveals further details about how the angular acceleration moments are distributed along the wing, from the proximal to distal regions along the stroke flight. As expected, in the first phase of the stroke cycle, the acceleration profile is very similar for the proximal and distal parts. At about a 30 % of the cycle, this pattern changes and the distal part is fully responsible for the observed motion changes. The proximal wing simply leverages the distal motion whose amplitude varies in several mini-cycles to a larger extension at the end of the cycle. From such acceleration 3D-map, it was also

possible to observe the non linear distribution of the angular acceleration in both axes: wing position and temporal cycle.
Finally, a global wing deformation, during a stroke cycle, was estimated from the whole experimental group W_t^l (resp. W_t^r), and shown in Figure 5. In general, every series of the wing deformation follows a very similar pattern, with a variance whose amplitude is smaller than a 10% of the spanned wing amplitude. As for other deformation estimations, a statistic boxplot was computed at each time of the cycle, using every recorded video, and again the red line represents the sample median while the spread of data is represented by the blue lower and upper quartiles. The larger stroke deformations correlate with the maximum wing effort moments, when motion abruptly rises and there exists a change of direction.

Figure **3-12**: Global wing deformation from the videos with the **C** camera configuration.

An independent analysis for the three flight stroke phases was also performed for the global deformation patterns. As expected, during the downstroke and upstroke phases, when there exists a relatively relaxed state and motion is almost constant, the wings act like rigid structures with small deformations, principally at the wing borders. The pattern is systematically reproduced by the different wings, with a very small variance (around 1.96 $\left(\frac{rad}{frame}\right)^2$). On the other hand, during the transition stroke phase, the deformation increases, as well as the variance. During this phase, the deformation is large because of the rapid wing adaptation to the new change of acceleration direction.

3.3.5 Discussion

This work introduced a novel approach to perform a complementary kinematic analysis of the hummingbird wing motion, using a multidimensional dense optical flow and without any prior information or any invasive sensor that could alter the natural flight motion gesture. The optical flow herein applied aims to extract global and local relevant information rather than following up the apparent location of individual pixels, and therefore is a suitable and robust tool to determine the fundamental motion wing patterns and to estimate the local wing deformation. The whole picture is reconstructed by means of a set of kinematic measurements computed from the apparent velocity vector field, which was in due turn used to spatially segment and cluster the different wing flow patterns. Angular acceleration and temporal wing deformation measures were herein estimated from local and global perspectives over a set of 93 videos. Quantitative metrics and results demonstrate both a remarkable coherence with previously

described biomechanical and dynamic hummingbird patterns, and a low variance of the computed kinematic patterns among all the evaluated samples. This method, of very easy implementation, is capable of establishing the kinematic flight lift generated during the upstroke and downstroke phases.

Flight analysis by means of physical approximations are often based on the study of some temporal sparse biomarkers, and therefore only provide temporal information for very simplified structures, or are too sparse to be statistically significant. These approximations in consequence result insufficient or unsuitable to understand complex maneuvers, especially when such motions are directly related with temporal non-linear structural deformation. Some computational models partially overcome such simplicity, allow for simulation of several flight patterns by subdividing the wing into simpler parts which are steered by independently parameterized physical models [183]. However, such simulations are usually unrealistic and far from being interpretable, they would hardly discover complicated relationships among forces, interactions, neuromotor commands and joint movements that might produce complex motions such as the hang up flight. In contrast, the proposed approach is able to describe convoluted kinematic patterns that are spatially spread out along the wing, using exclusively the video information, without prior hypotheses. In this sense, the presented strategy not only can find out reliable motion patterns, but it also may reveal local deformation patterns of the wing .

Rotational patterns are crucial in flight analysis since they allow to study the source of circulation air which is straightforwardly related to the upward and downward forces during the stroke cycle. Furthermore, the angular acceleration, the angle of attack and the span ratio are responsible for controlling the upstroke circulation, the wing orientation and the maintenance of the forward flight [80, 263]. The Global angular acceleration herein estimated presents large gradients during the upstroke and downstroke transitions, a result of the upward and downward wing forces which are responsible for the circulation changes. From a local angular acceleration standpoint, the proposed approach also allows to estimate a regional angular acceleration profile along the wing, thereby discovering not only the phases where the sources of circulation change but also determining which particular wing regions are more involved in the generation of these forces. As illustrated in Figure 3.3.4, the main angular acceleration modes are initially concentrated in the proximal wing regions because the shoulder and pectoral muscles drive the forward wing motion, a pattern that is displaced to the distal wing regions when the passive accumulated energy rules the downstroke phase. Although the proposed method is able to capture local motion patterns from the dense optical flow, the spatial accuracy is restricted by the number of scales and the standard deviation used to convolve each frame. For instance, when the scales are obtained by

convolving Gaussians with large sigmas, results would be more robust to noise motion but less accurate spatially. A proper balance between the noise control and the spatial accuracy is part of the method tuning as to obtain coherent results.

On the other hand, the stunning deformations of the wing are responsible for much of the salient abilities of hummingbirds such as their stability and relative conservation of the orientation, even if they are forced to change constantly their acceleration patterns [11]. In addition, localized wing deformations may produce relative point circulation vortices that induce a non linear pattern along the wing structure and therefore completely alter the hummingbird flight smoothness [131]. In the proposed strategy, deformation wing patterns were locally estimated along the wing as the relationship between each local orientation w.r.t an average orientation, resulting in a wing map of deformation for every stroke. Main wing deformations are commonly localized along the proximal and distal segments while the largest deformation is observed when the wing acceleration is maximum, i.e., during the downstroke and upstroke transitions (see Figure 3.3.4).

Recent methodological advances have improved our comprehension of the most complex hummingbird maneuvers and their intrinsic mechanisms, for instance, aerodynamic studies using PIV have shown an asymmetric distribution of the lift forces during the stroke, probably caused by unknown micro-mechanisms like the stall and leading-edge vortices [281]. New analysis instruments are then required to approach such local and non linear phenomena whose role is yet an open problem. The proposed approach herein introduced is a global-to-local analysis tool that complements the existent methods, with the benefit of methodological simplicity. Likewise, the two variables herein estimated, angular acceleration and wing deformation, were locally and globally computed, opening thereby the possibility of conducting more subtle studies that discover the wing regions that contribute to every step of the hummingbird hovering. The proposed approach is also able to locally compute force and torque patterns at the different wing segments by directly relating the angular acceleration with real anatomical values. Finally, this method might be used to automatically segment the hummingbird flight strokes in sequences of video, highly reducing inter and intra expert variability.

3.3.6 Conclusions

This work has presented a novel approach to quantify the hovering hummingbird motion, using a dense optical flow computed from video sequences. A main contribution is the markerless flight analysis framework that estimates two relevant kinematic hummingbird variables i.e., the angular acceleration and wing deformation. The proposed method opens up an actual possibility of understanding more complex hovering phe-

nomena. Future works include the study and development of a robust method that temporally relates the kinematic dense flow description with force measurements generated by flapping, under actual anatomical and mass parameters.

3.4 Characterization of motion cardiac patterns in magnetic resonance cine

One of the most important tasks in Cardiac Magnetic resonance Cine (CMRC) consists in identifying and describing normal and abnormal dynamic heart patterns, a task usually performed by physicians. Segmentation and tracking may support decisions during a particular treatment, but their performance is dependent on the quality of the video. The acquired signal, on the other hand, is contaminated with noise coming from physiological movements and devices, resulting in cardiac blurred boundaries. This paper presents a novel method that automatically identifies flow heart patterns by establishing similarities between two consecutive frames to which a local jet feature analysis has been applied. Once a vector motion field is calculated, spatially connected regions with minimal variance are found as the sources of movement and different statistics objectively estimate movement patterns of these regions. The utility of this method is illustrated by comparing the temporal series of these regions between normal and abnormal patients. *The complete content of this section has been published as a research article in the proceedings of IEEEXplore (see [199])*

3.4.1 Introduction

Cardiovascular disease (CVD) is an important worldwide health concern that amounts to a 29.2 % of the total global deaths, according to World Health Report in 2003 [213]. Magnetic Resonance Cine (MRC) is a non invasive video technique that provides cardiac anatomic information with good spatial and temporal resolution, allowing the analysis of the heart dynamic patterns. These analyses aim to estimate, globally, the myocardial function, and specifically, the wall motion information. Therefore they are used as an indicator of the pathological and normal movements [135]. Overall, these analyses are carried out by expert physicians, whereby results are highly subjective. Likewise, in the actual clinical routine there is no objective measurement upon the cardiac flow, a fundamental limitation of this analysis. This may be especially useful in early disease detection, case in which the boundary between normal and abnormal is quite diffuse, even for expert physicians. The point is that any objective measurement must take into consideration different variation sources, namely, changes of pose and reflection properties of the object, acquisition device noise and similar tissue properties

of the neighboring heart tissues. For instance, in the video, it is really difficult to establish a neat border between heart and liver [190, 219].

So far different approaches have been proposed to obtain dynamic and structural heart patterns. Some of them, based on region growing or edge-detection, present difficulties when dealing with noise, grey scale variations and low gradients. Therefore, a high degree of supervision is required [189]. On the other hand, other works have used MR tagged images, a technique in which the tissue motion is related to the tag line locations or intersections [170]. Active contours or statistical models that attempt to track the ventricle movements have been also used. These active contours are computationally expensive and their performance in noisy conditions is very poor. The statistical models are not correlated with any physiological phenomenon, a very important issue from the clinical standpoint since the objective is not uniquely to follow a movement but rather to establish patterns with physiological meaning [123, 254]. Finally, even if the tagging technique offers appropriate quantitative analysis for some dynamic parameters, its clinical use has so far remained limited mainly because its performance is very poor when the acquisition process is noisy. Besides, the tagging resolution is so low that the physiological meaning of the found flow patterns is really difficult to establish [254].

The main contribution of this work is the design of a fully automatic method which determines the fluid vector field that represents the apparent heart motion. The method allows to define objective measurements upon the main FOE (Foci of Expansion), defined as those locations with minimal motion from which line fields are coming out from one of these foci to another. This paper is organized as follows: We first briefly present the dataset, then Section III introduces the proposed method, Section IV demonstrates the effectiveness of the method and the proposed metrics, a useful tool to help the expert physician. The last section concludes with a discussion and possible future works.

3.4.2 Dataset description

Ten different cases were herein included to assess the proposed method, five normals, three diagnosed with Fallot tetralogy and the last two, with acute myocardial infarction. The images were acquired in a 1.5T Intera scanner (Philips Medical Systems, The Netherlands), equipped with 30 mT/m gradients, using a dedicated 5 elements phased array cardiac coil for every patient, The 4D sequences follow the conventional views: short axis, two chambers and four chambers.

A balance gradient echo sequence (balanced TFE) has been used in all cases. The most relevant parameters for the short axis view are: slices = 13, slice thickness = 8 mm, heart phases = 30, TR/TE = 3.15 ms, flip angle = 60°, matrix = 256 × 192, SENSE factor = 1.5, turbo factor = 8. For the two chambers view: slices = 2, slice thickness =

8 mm, heart phases = 20, TR/TE = 2.7/1.3 ms, flip angle = 60°, matrix = 256×186, SENSE factor = 1.5, turbo factor = 24. Finally, for the 4 chambers view: slices = 2, slice thickness = 8 mm, heart phases = 20, TR/TE = 3.2/1.6 ms, flip angle = 60°, matrix = 256×175, SENSE factor = 1.5, turbo factor = 15.

3.4.3 The proposed Method

Figure **3-13**: Local jet features from MRC slices are calculated and stored in a kd-tree by similarity. The motion is estimated by searching a particular pixel in the precedent kd-tree. Once the apparent heart motion is obtained, some metrics define the dynamic patterns.

The whole method starts by calculating the Jacobian and Hessian matrices for different scales of a slice, using a Local Jet Feature approach [98]. The resulting local jet representation is then grouped by similarity i.e., foci with minimal variance. A kd-tree structure stores the values of local jet for every pixel. The motion is estimated by searching a particular pixel in the precedent kd-tree (the kd-tree at time $t-1$). Such search is firstly addressed to the pixel with the closer local jet feature Euclidean distance. The kd-tree structure allows orientating this search to the pixels with similar local feature description. Once this most similar pixel is found, the two associated coordinates are used to calculate the motion vector. One main contribution of the present work was the design of an objective measurement on the obtained vector field: spatially connected regions with minimal variance were firstly grouped, i.e. regions that maximize the inter-group variance regarding the vectors magnitudes and angles.

3.4.3.1 The Multiscale Local Jet Similarity Space

Nowadays it is well accepted in the computer vision community that relevant information can be determined by different strategies that measure correlation or coherence through different scales [185, 244]. The strategy herein applied is known as local jets and basically consist in obtaining an homogenous partition of the frequential spectrum by analyzing the image at different scales, which are then characterized by a map of partial derivatives of different order (we used here a first and second orders but the analysis can be extended to higher orders when one might want a finer spectral partition). In consequence, the first step in our approach was to obtain a collection of spatial derivates at different scales from independent slices, as follows:

$$f_{ij}^{\sigma} = f * \frac{\partial^{i+j} G_{\sigma}}{\partial x_1^i \partial x_2^j} \tag{3-3}$$

where G_σ is the 2d Gaussian function with standard deviation σ. The Gaussian kernel is here justified because this is the unique kernel with an homogeneous scale-space representation (linearity and shift invariance in both frequency and space) [185]. The multiscale local jet is given then by the collection $\left\{ f_{ij}^{\sigma}; i+j \leq r, \sigma \quad S \right\}$, where r is the derivation order and $S = \sigma_1, \ldots \sigma_q$ the selected scales.
Local jets as descriptors have been already used on natural images, in which it has been shown that the first eigen (singular) vectors obtained by PCA or SVD are quite similar to the first derivatives of a 2d Gaussian function [194,212], whereby this representation is quite sparse. Therefore, information can be automatically estimated by calculating the energy concentrated at each scale.
Then, local jets are normalized [185]:

$$F_{ij}^{\sigma} = \frac{\sigma^{i+j}}{i+j+1} f_{ij}^{\sigma} \tag{3-4}$$

where σ^{i+j} is the scale normalization factor and $i+j+1$ is the number of $(i+j)$ order derivatives. The following single scale distance is used:

$$d_f^{\sigma}(x,y) = \sum_{i+j \leq r} \left(F_{ij}^{\sigma}(x) - F_{ij}^{\sigma}(y) \right)^2 \tag{3-5}$$

3.4.3.2 Data structure

Once a MRC slice is represented by a multi-scale set of features (local jet collection F_{ij}), the next step consists in storing the local jet in a kd-tree structure F. This data structure, a binary space partition, organizes the whole set of features and allows rapid access to them [30], performing nearest neighbor (NN) searches in the feature space. This efficient search is extensively used to perform operations based on similarity between pixels that are not connected in the image space.

3.4.3.3 Heart Motion Estimation

The apparent heart motion is estimated for every pixel x at frame t by using the similarity cluster established before. At frame t, slice f_t, and pixel x, we compute $u(f_{t-1}, f_t, x)$, the nearest neighbor of the feature vector associated to x in the f_{t-1} feature space:

$$\boldsymbol{u(f_{t-1}, f_t, x) = \arg \min_{v \epsilon \mathcal{F}^{-1}_{f_{t-1}}} d^F(\hat{x}_{f_t}, v)} \tag{3-6}$$

where $\hat{x}_{f_t}$ is the x projection onto the feature space of f, i.e., $\hat{x}_{f_t} = \ f_{i,j}(x)$ and d^F denotes the feature space distance. Then we compute $y(f_{t-1}, f_t, x)$, the pixel from f_{t-1} which is the most similar to x from f_t:

$$\boldsymbol{y(f_{t-1}, f_t, x) = \arg \min_{z \epsilon \mathcal{F}^{1}_{f_{t-1}} u(f_{t-1}, f_t, x)} d^I(x, z)} \tag{3-7}$$

where d^I denotes the distance in the image space. We obtain the pixel from the set of pixels associated to the feature vector u which is the closest to x in the image space. Finally, the velocity vector is computed as the difference:

$$c(f_{t-1}, f_t, x) = x - y(f_{t-1}, f_t, x) \tag{3-8}$$

Our method was implemented in C++ language on a standard PC (AMD with 2.66 GHz CPU and 3 GB of RAM) takes about 10-15 sec to process all slices of a plane.

3.4.3.4 Computed Measures

Our fundamental hypothesis is that those places with minimal movement constitute a stable source of information of the whole flow pattern. Once the vector flow was determined, the next measurements were proposed:

Flow motion clustering: Similar flows are grouped as the sets of points with minimal variance regarding the vector magnitude and angle. The number of sets depends on the type of description one might want to obtain, for instance with $k = 2$ the two circulatory circuits are repaired, i.e., left and right. With $k = 4$ we observe the principal fluid directions, mostly in the left part of the circulatory circuit. A larger number produces more clusters but with no physiological meaning.

Global and local statistic measures: computing simple statistics either globally or in a small neighborhood, results in a saliency map of the flow pattern. This saliency map corresponds to heart segmentation, when this is locally considered, and to a tracking of the Center of mass, when the analysis is global. In both cases, these patterns (as it will be illustrated later) show very different appearance.

3.4.4 Experimental Results

The method parameters were adjusted to obtain the best performance and then set to the following values: $NN = 1$ with a local jet order of 2 and $\sigma = 5$. Typical results of computing motion estimations for the three planes are shown in Figure **3-14** (upper sequence). After visual evaluation, the expert determined that the proposed method achieved a very realistic motion simulation during a cardiac cycle. Figure **3-14** (lower sequence) shows the flow determined by the different selected clusters on every slice of a cardiac cycle. On the other hand, it is possible to use the saliency map built up from the computed flow (magnitude and norm) and determine a center of mass of this flow or global heart trajectory, as showed in Figure **3-15**. This trajectories can be used as dynamic patterns to identify or grouping different kinds of movement. The RMSE computed from global trajectories demonstrated a minimal error between similar movements as illustrated in Table **3-8**, for instance, the difference among normal flows is 0.04 in average, while the difference with Fallot tetralogy is 20.51 and the difference with infarction movement is 23.023. This difference suggests that this measure could be used as indicator of some abnormal movements.

Figure **3-14**: The flow determined by the different selected clusters on different slices of a cardiac cycle.

Mean Signal	**RMSE-Normal**	**RMSE-Fallot**	**RMSE-Infarction**
Mean-Normal	0.04	20.57	23.16
Mean-Fallot	20.51	3.89	10.27
Mean-Infarction	23.023	8.81	3.82

Table **3-8**: Root Mean Squared Error for the three evaluated movements. These averages have been calculated from the differences between the signals and the mean representative signal to every movement of the cardiac cycle.

Figure **3-15**: Mean trajectories of CoM heart movement computed since the saliency map. In this figure it is possible to observe the characteristic of global motion computed from saliency map and the strong differences between the three kinds of movements.

3.4.5 Conclusions

We have introduced a new method to analyze the heart dynamics by estimating the apparent motion. We obtain a robust representation which captures some important heart dynamic and static features, for instance the most important topological elements that represent both places with minimum coherent movement (fluid sources) and with largest change rate (they represent the fluid lines). Experiments demonstrated that this can be used to characterize fluid patterns and establish a base normal line, as confirmed by an expert in the domain. Measures presented in this work constitute a first approximation to understanding the complex dynamic heart. From this kind of analyzes, we expect that it is possible to set up a collection of descriptors which allow to accurately describe motions patterns and quantify its semantics. The presented procedure could be extended to other type of medical imaging problems, or may be used to segment the wall of the heart with better accuracy by taking advantage of the temporal information. On the other hand, the combination of our method with statistical approaches would allow tracking of specific heart areas.

4 Motion analysis and simulations from prior knowledge.

At a more sophisticated level, visual systems exploit the learned relationships between motion observations and prior concepts and analyze complex patterns, accomplishing a remarkable discrimination of normal and abnormal behaviors. This chapter aims to approach such analysis at characterizing, following and simulating normal and pathological dynamic gait patterns (see in section 4.1 and 4.2). As a contribution, the prior knowledge and observations were herein related in a classical segmentation task: a prior shape was iterative evolved to adjust noisy observations and segment the prostate in the context of radiotherapy planning (see in section 4.3).

4.1 A Kinematic Method for Computing the Motion of the Body Center-of-Mass (CoM) During Walking: A Bayesian Approach

The gait pattern of a particular patient can be altered in a large set of pathologies. Tracking the body centre-of-mass (CoM) during the gait allows a quantitative evaluation of these diseases at comparing the gait with normal patterns. A correct estimation of this variable is still an open question because of its non-linearity and inaccurate location. This paper presents a novel strategy for tracking the CoM, using a biomechanical gait model whose parameters are determined by a Bayesian strategy. A particle filter is herein implemented for predicting the model parameters from a set of markers located at the sacral zone. The present approach is compared with other conventional tracking methods and decreases the calculated root mean squared error in about a 56% in the x-axis and 59% in the y-axis. *The complete content of this section has been published as a research article in the journal of* ***Computer Methods in Biomechanics and Biomedical Engineering*** *(see [92]).*

4.1.1 Introduction

The gait pattern can be altered in a large set of pathologies, such as diabetes, brain palsy, cerebral vascular accidents, and neuromuscular dystrophies or from any kind of accident. The study of the human body movement or visual gait analysis is a modern tool that allows to objectively assess any of these pathologies. Examination is based on the follow-up of dynamic variables, whereby the disease severity can be quantified, and the gait compared with normal patterns [104, 105, 193].
The musculoskeletal dynamics, obtained from a gait analysis, is evaluated from the kinetic and kinematic perspectives. The kinematic analysis describes patient displacements in terms of the system components and its fundamental relationships, using variables such as the trajectory of the CoM, some specific angle joint variations or the step length, among others [74]. The kinetic analysis quantifies the needed energy for the human movement to be produced, measuring electro-physiological states during displacement [47]. A correct extraction and quantification of these variables is an open question since they are highly non-linear.
The CoM constitutes a fundamental descriptor for the clinical gait analysis because through its movements it is possible to describe both central nervous system and musculoskeletal disorders [76, 122, 265]. The more accurate CoM is usually estimated using a force plate, where a double integral of the ground reaction force in the time define the CoM displacement [86, 110]. This relationship is a simple dynamic equilibrium equation:

$$d_{CoM} = \int\int \frac{F_{GR} - mg}{m} dt^2 + v_0 t + d_0 \tag{4-1}$$

with F_{GR}= ground reaction force, m= body mass, g= gravity, t= time, v_0 and d_0 the integration constants of the initial velocity and position. This method is the gold standard for the CoM calculation. However, this method requires a patient steps upon the force plate, a difficult task in many musculo-skeletal disorders. Moreover, conventional gait laboratories have only two force plates, which results insufficient when gait analysis demands data from the whole gait cycle. This method is only used for validation of others techniques and its accuracy depends on the integration constants so that its utility is still limited in real clinical scenarios [86, 110].
In the clinical routine, it is common to use the optic kinematic methods for estimation of the CoM because of its versatility and control. However, accurate location of the CoM is impossible because of the high inter-patient anatomical variability and its within-the-body location [76]. Usually, the CoM is tracked the closest marker in the video [85]. This CoM can also be estimated from the body segments [86], but its extraction

is a complicated task that requires many markers and changes the natural motion expression. These methods do not accurately estimate the CoM trajectory [86, 110], since this is the result of a complex interaction of forces, neuromotor commands and joint movements of the lower limbs, and therefore it shows a high non-linear dynamics [203].

This article presents a precise and efficient strategy for estimating the temporal CoM location using a non-linear gait biomechanical model whose parameters are recursively adjusted by a Bayesian strategy, herein implemented as a particle filtering. The Main contribution of this work is to use a very simple methodology to follow an actual non linear dynamic. The whole strategy allows a natural and accurate tracking of the non-linear gait patterns with a high degree of noise robustness.

4.1.2 Materials and Methods

The strategy for tracking the CoM requires that the set of video frames, within two alternating heel strikes (gait cycle), are segmented. Tracking is carried out under a Bayesian framework that defines the trajectory of the CoM as a sequence of hidden states. At each step of the gait, the more probable parameters are calculated from a set of observations and the previous step parameters.

The Bayesian approach for tracking aims to estimate the hidden states of a system from a set of observations, or at least to extract useful information under the fundamental hypothesis that the observation process do not completely destroy the link between the true and the observed variables, or that the observed and true data are somehow close together under a particular metric. A Bayesian strategy starts at defining the system state as a random variable x_t and its associated probability density function (*pdf*), i.e., the uncertainty level of the occurrence of the state x_t. Bayesian filters estimate such *pdfs* upon a state space following the sequence of observations. The belief $Bel(x_t)$ is defined as a posterior probability density function of x_t, conditioned to all the observed available data $z_1, z_2, \ldots, z_t$ at time t. This *pdf* addresses the question: *what is the probability that the system state is in x_t if the history provided by the measurements of the process are $z_1, z_2, \ldots, z_t$?* Since the number of observations increases through the time, the complexity of such *posterior density* grows exponentially. This estimation becomes computationally tractable by assuming a Markovian hypothesis: the current state of the system contains all the relevant information. Under this assumption, $Bel(x_t)$ may be computed efficiently without information losses. In practice, it is a required a system model $p(x_t\ x_{t-1})$, that represents *how the system states change in a time step* and a likelihood model $p(z_t\ x_t)$, which describes the *probability of making the observation z_t if the system state is in the x_t* state. Finally, the initial system state

$Bel(x_0) = p(x_0)$ is also needed. Once this information is available, the belief $Bel(x_t)$ is calculated in two recursive steps:

- *Prediction.* The belief in the state x_t is computed by updating the previous belief $Bel(x_{t-1})$, according to the prediction given by the system model $p(x_t\ x_{t-1})$, through the Chapman-Kolmogorov equation:

$$\widehat{Bel}(x_t) = \int p(x_t\ x_{t-1}) Bel(x_{t-1}) dx_{t-1} \tag{4-2}$$

- *Update.* The predicted belief $Bel(x_t)$ is adjusted after the system observations:

$$Bel(x_t) = \frac{\widehat{Bel}(x_t) p(z_t\ x_t)}{\int \widehat{Bel}(x_t^*) p(z_t\ x_t^*) dx_t^*} \tag{4-3}$$

Bayesian filters provide a probabilistic framework for recursive and sequential estimations of the system state. This representation is important for obtaining good estimators in non-linear/non-Gaussian dynamics as it is the case of the CoM trajectory, herein tracked through a particle filtering which approached the belief function by discrete sampling.

4.1.2.1 Modeling the CoM movement

The human gait is a locomotion process which involves both lower limbs to help the body to keep the balance while it gains support and propulsion [284]. This process comprises a cyclic set of movements, where one foot acts as a fixed point which supports the body swings, whereas the free foot moves forward until it reaches the floor and becomes the new fixed point.

Figure **4-1**: Sequence of events of a complete gait cycle extracted from the humanEva database (Sigal and Black 2006).

Schematically, the upper part of the body is represented by a mass which moves forwards with respect to each fixed point, describing a harmonic oscillating trajectory, similar to the inverted pendulum [40]. At the same time, the free foot swings with respect to this mass, such as a simple pendulum. Given that these processes are coupled together, the human gait is modeled by a double-articulated pendulum (see Figure

4.1.2.1). This model properly represents the gait trajectory and it has been widely used [83, 167, 174]. Based on the double articulated pendulum model, MacGeer [203] formulated the theory of passive dynamic motion, which describes the movement without a complex control mechanism, i.e., it is more important the body's structure for understanding the gait rather than its control or muscular activity. Afterward, based on this theory, Garcia et. al. [109] formulated a simplified gait model. Accordingly, the feet are relatively small with respect to the trunk and the heelstrike is subjected to a restriction rule. This model has been used for tracking other body's structures [116,153,297]. The model corresponds to two coupled non-linear differential equations:

$$\begin{aligned} &\ddot{\theta}(t) - \sin\theta(t) = 0 \\ &\ddot{\theta}(t) - \ddot{\phi}(t) + \dot{\theta}(t)^2 \sin\phi(t) - \cos\theta(t)\sin\phi(t) = 0 \end{aligned} \tag{4-4}$$

where θ is the angle of the stance leg at particular time t with respect to the slope and ϕ is the angle between the stance leg. The model also defines a transition rule that simulates the swing foot when it hits the ground at the heelstrike, this moment corresponds to $\phi(t) - 2\theta(t) = 0$.

4.1.2.2 Online Adaptation with a Bayesian Filter

The biomechanical model computes the trajectories for the CoM from temporal functions (θ_t, ϕ_t), which are hidden, and the set of markers, which are the only direct measurements. We shall focus on the markers located within the sacral zone, more specifically the $z_t = (M_t^1, M_t^2, M_t^3, M_t^4)$ located around the hip (two at the back and two at the front), as illustrated in Figure 4.1.2.2. Provided that locations of these markers are indirect measurements of the CoM movement, their location in time t corresponds to the two hidden parameters $x_t = (\theta_t, \phi_t)$ which better match the observed marker locations $z_1, z_2, \ldots, z_t$ at time t. The most probable state x_t can be recursively found using a Bayesian filter. For doing so, we need to set an initial estimation of the CoM location, a prior of CoM evolution during the gait and the likelihood function which associates states and observations (marker locations).

Figure **4-2**: Marker locations in the sacral region.

An estimation of the initial CoM location is herein calculated from a Gaussian distribution whose mean and covariance are computed from the set of four positions located

closest to the actual CoM. These positions are marked by an expert in the first frame of the video gait recording.
The dynamic model of the CoM movements indicates how the angular functions (θ_t, ϕ_t) change during a gait cycle. Let us suppose that the gait satisfies Garcia's model (equation 4-4) including the heel strike rule. This model can be modified for including the gait variability so that gait dynamic reads as

$$P\left(x_t \; x_{t-1}\right) = G(x_{t-1}, \Sigma_1^2) \tag{4-5}$$

where G corresponds to a multivariate Gaussian distribution, whose mean was experimentally set [109] and σ describes an inherent gait variability. The covariance matrix Σ_1^2 was calculated using a maximum likelihood estimation [160] from a data set given by the difference between the ideal CoM signal (drawn from [86]) and the signal obtained from the prior model [109]. On the other hand, the likelihood function should yield a maximal probability when the states closely follow the observations, i.e:

$$P\left(z_t \; x_t\right) = \qquad G\left(\frac{1}{4}\left[\bar{z}_{x,t}, \bar{z}_{y,t}\right] - Le[\sin(x_{\theta_t,t}), \cos(x_{\theta_t,t})], \Sigma_2^2\right) \tag{4-6}$$

where $\bar{z}_{x,t} = \frac{1}{4}\sum_{i=1}^{4} M_{x,t}^i$, $\bar{z}_{y,t} = \frac{1}{4}\sum_{i=1}^{4} M_{y,t}^i$, $M_{x,t}^i$ and $M_{y,t}^i$ are the marker coordinates in the x and y directions. The CoM coordinates defined by $(Le[\sin(x_{\theta_t,t}), \cos(x_{\theta_t,t})])$ are directly computed from the geometric representation of Garcia's model, where Le is the leg length and $x_{\theta_t,t}$ is the angle of the stance leg with respect to the slope, at a particular time t. Adittionally, Σ_2^2 is a predefined covariance matrix calculated as a maximum likelihood estimation from a data set, given by the difference between the ideal CoM signal and observations. The mean of the Gaussian distribution stands for a measure of how well the hidden angles meet the observed marked positions. The use of this Gaussian is fully justified since the noise associated to the captured positions is independent of the anatomical details.

4.1.2.3 Particle Filter Implementation

A different approach to represent the Belief is a discrete approximation of the *pdf* using Monte Carlo techniques. These methods approximate the probability density function $p(x)$ using a large number of samples,

$$p(x) \approx \sum_{i=1}^{n} w^{(i)} \delta(x - x^{(i)}) \tag{4-7}$$

where $x^{(1)}, x^{(2)}, \ldots, x^{(n)}$ are a set of n discrete independent and identically distributed variables (i.i.d), $w^{(i)}$ are the weights which stand for the probability of occurrence of the sample x_i, and $\delta(x)$ is the Dirac function. Clearly, the larger the number of samples, the closer the description is to the *pdf*. This approximation allows useful quantities, such as the discrete expected value of any function $f(x)$:

$$\int f(x)p(x)dx = \lim_{n} \sum_{i=1}^{n} w^{(i)} f\left(x^{(i)}\right) \tag{4-8}$$

and the samples are generated using methods such as the rejection or importance sampling [16], better known as Particle Filters.
A particle filter is the discrete version of the Bayesian filter obtained when the belief is approximated using a Monte Carlo (MC) method. The belief is estimated with the point mass distribution defined in equation 4-7, when replacing the Chapman-Kolmogorov equation by the approximation defined in equation 4-8, and discretizing the update equation with the belief previously calculated. Since each particle corresponds to an independent state of the system, with a number of particles which is function of the desired precision, the method allows to simulate the evolution of the complete system with no restrictions regarding linearity or noise. Additionally, the belief can be modified for each of the simulated system states as to adapt to real observations. Finally, classic estimators such as the expected value of the state or the maximum a posteriori can be calculated from the simulation while the system is evolving for predicting future states.
The Particle filter is constructed as follows. First, the Belief is approximated by an empirical point-mass function:

$$Bel(x_t) \approx \sum_{i=1}^{n} w_t^{(i)} \delta(x_t - x_t^{(i)}) \tag{4-9}$$

where the weights $w_t^{(i)}$ are chosen using the principle of sequential importance sampling (SIS) [16], which states that samples can not be directly sampled but rather found through a density importance function. Hence a weighted aproximation is given by:

$$w_t^{(i)} \quad \frac{P(x_0 \ldots x_t \; z_1 \ldots z_t)}{q(x_0 \ldots x_t \; z_1 \ldots z_t)} \tag{4-10}$$

where $P(x_0 \ldots x_t \; z_1 \ldots z_t)$ is exactly the very same belief function, as defined in section 2 and $q(x_0 \ldots x_t \; z_1 \ldots z_t)$ is an importance density function, chosen to meet:

$$q(x_0 \dots x_t \; z_1 \dots z_t) = q(x_t \; x_{t-1}) q(x_0 \dots x_{t-1} \; z_1 \dots z_{t-1}) \tag{4-11}$$

Now, particles evolve by dependences which are always of first order, so we can suppose that $\left\{ x_{t-1}^{(i)} \right\} \sim q(x_t \; x_{t-1})$. We can then assume that $q(x_t \; x_{t-1})$ represents the system CoM dynamics as $P(x_t \; x_{t-1})$, so:

$$q(x_t \; x_{t-1}) = G(x_{t-1}^{(i)}, \Sigma_1^2) \tag{4-12}$$

Particles need then to be updated after the Bayes rule defined in equation (3), which can then be approximated as follows:

$$P(x_0 \dots x_t \; z_1 \dots z_t) \quad P(z_t \; x_t) P(x_t \; x_{t-1}) P(x_0 \dots x_{t-1} \; z_1 \dots z_{t-1}) \tag{4-13}$$

By substituting (4-11) and (4-13) into (4-10), the weight updating equation reads as:

$$w_t^{(i)} \quad \frac{P(z_t \; x_t) P(x_t \; x_{t-1}) P(x_0 \dots x_{t-1} \; z_1 \dots z_{t-1})}{q(x_t \; x_{t-1}) q(x_0 \dots x_{t-1} \; z_1 \dots z_{t-1})} \tag{4-14}$$

$$w_t^{(i)} = w_{t-1}^{(i)} \frac{P(z_t \; x_t) P(x_t \; x_{t-1})}{q(x_t \; x_{t-1})} \tag{4-15}$$

note that the left term is just the weight in the precedent time and the rest of the expression corresponds to the system dynamics, previously defined. Therefore, this entire expression can be re-written as:

$$w_t^{(i)} = w_{t-1}^{(i)} G \left(\frac{1}{4} \left[\bar{z}_{x,t}^{(i)}, \bar{z}_{y,t}^{(i)} \right] - Le[\sin(x_{\theta_t,t}^{(i)}, \cos(x_{\theta_t,t}^{(i)})], \Sigma_2^2 \right) \tag{4-16}$$

After some iterations, all but one particle will have negligible weight, a recurrent drawback broadly documented in the literature and known as the SIS degeneracy phenomenon [16]. To cope with, we introduce a step of resampling that generates a new set of particles at each iteraction. Finally, The algorithm is explicitely described hereafter:

Algorithm 2 Particle Filter implementation

$\left[\ x_t^i, w_t^i \ {}_{i=1}^{N} \right] \sim Bel(\ x_t^i, w_t^i \ {}_{i=1}^{N}, z_k)$

FOR $i = 1 : N$

set $\left\{ x_{t-1}^{(i)} \right\} \sim G(x_{t-1}^{(i)}, \Sigma_1^2)$

update weights according to $w_t^{(i)} = w_{t-1}^{(i)} G \left(\frac{1}{4} \left[\bar{z}_{x,t}^{(i)}, \bar{z}_{y,t}^{(i)} \right] - Le[\sin(x_{\theta_t,t}^{(i)}, \cos(x_{\theta_t,t}^{(i)})], \Sigma_2^2 \right)$

END FOR

normalize weights $\sum_{i=1}^{n} w_t^{(i)} = 1$

resample $x_t^i, w_t^i \ {}_{i=1}^{N}$

4.1.2.4 Gait data

Validation was carried out on a set of gait cycles segmented from the humanEva dataset [251]. The humanEva data consists of a set of videos, captured from four subjects in different activities, using a calibrated marker-based-motion-capture-system and multiple high-speed video capture systems. Every video is provided with an associated motion data in C3D format that describe the accurate 3D marker position. For evaluation purposes, we assume this capture has an associated Gaussian noise, an statement fully justified since the marker position is independent of the anatomical location and the capture process, as described before in subsection 2.2. The proposed strategy was assessed on twenty gait cycles from three different subjects ,i.e., 8400 frames which corresponded to a total of sixty cycles. Each cycle corresponds to the set of frames within two alternating heelstrikes. The initial and final points of each cycle were selected by an expert in the domain, who also verified that the extracted sequences corresponded to normal gait patterns. The locations of four hip markers were extracted for each frame, as well as the CoM location. For doing so, the evaluated subjects have a set of attached makers, following the VCM protocol (Vicon Clinical Manager), from which 4 were extracted and defined as:

- $LPSI$: Placed directly over the left posterior superior iliac spine
- $RPSI$: Placed directly over the right posterior superior iliac spine
- $LASI$: Placed directly over the left anterior superior iliac spine
- $RASI$: Placed directly over the right anterior superior iliac spine

Finally, a marker placed in the sacral region was set and defined as the CoM.

4.1.3 Evaluation and Results

Evaluation was carried out by comparing the accuracy of the presented strategy with two standard tracking approaches: the location of the closest marker to the CoM and an exponential weighted moving average (EWMA) [65,146], a typical method for tracking time series. Control data were obtained from normal patterns, captured from force plates and reported in the literature [86] which thereby fully describe coordinates x and y of a normal cycle. The particle filtering algorithm was written in java (JDK 1.6.0_07) and run under an AMD turion 64 processor of 1.59 GHz and 3 GB in RAM. The Garcia's model was solved using a fourth order Runge-Kutta method, also implemented in Java.

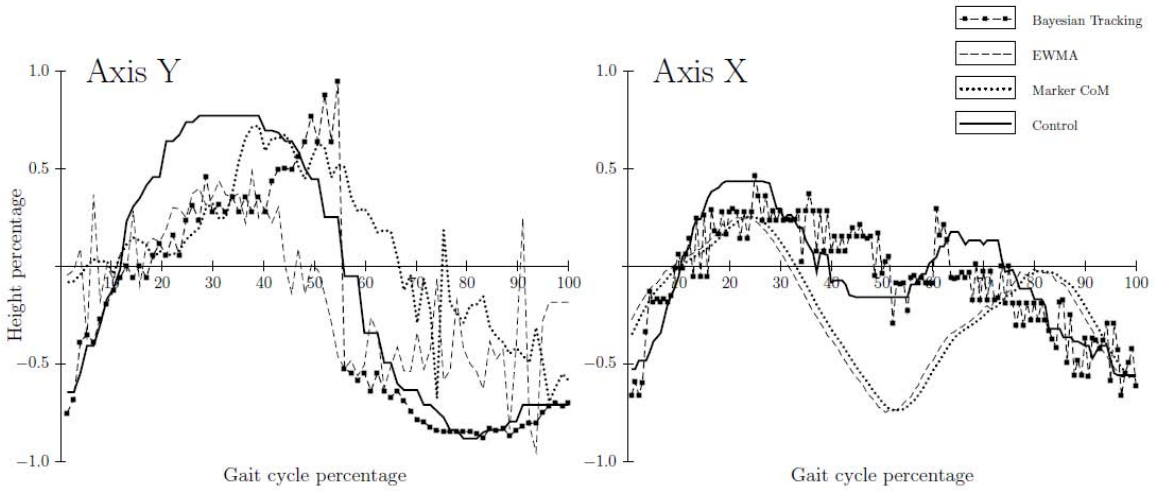


Figure **4-3**: The gait trajectory has been divided into its y (left) and x (right) components for better analysis. The y-pattern is characterised by a non-linear periodic path (thicked line) which is closely tracked by the Bayesian strategy (squared dots) and hardly matched by the other two methods (dashed lines). The x-plot complements the whole picture, the marker CoM and EWMA methods are shifted away from the control path, which is closely followed by the Bayesian strategy. The Bayesian method highly outperfoms the two other methods, regarding no linear adaptation.

Figure 4.1.3 shows the gait trajectory decomposed into the y-coordinate at the left, and the x-coordinate at the right, during a complete gait cycle (control). The y-

axis displays in both cases the CoM change, in percentage, weighted by the body height, while the x-axis is weighted by the entire cycle duration and also expressed in percentage. The two panels superimposed upon the same plot, the control gait path and the predicted trajectory of three different strategies: the closest marker to the CoM, a EWMA computed using the actual CoM observations along the recorded gait cycle and the approach herein presented (Bayesian tracking). The vertical movement or the y-axis spans the control pattern during $300\,ms$, along which the periodic movement is smoothly increasing within the first $40\,\%$ of the entire cycle. At this point, the height percentage decreases in a non-linear manner, because of the heelstreak, a task quite difficult to follow by any of the methods used. Importantly, the Bayesian tracking matches this non linearity much better than the other two, i.e., the Bayesian presents a little oscillation in this phase but rapidly decreases and closely follows the control pattern. On the contrary, the other two methods are highly oscillating after this phase and remain distant from the actual pattern until the cycle ends up. Regarding the x-axis displacements, there appear two different trajectories, a first one formed by the control and the Bayesian method and a second one by the other two methods, which follow nearly the same path but are shifted away from the control trajectory. These results demonstrate that the proposed method is more stable than the EWMA after the heel strike and follows much better a control pattern which is quite noisy. Interestingly, the particle filtering approximates better the non-linear plot of the control trajectory, a result which holds our original hypothesis about the necessity of using more robust tracking methods.

Tracking Method	x-axis	y-axis
Marker CoM	0.414 $\pm$0.05	0.4 $\pm$0.081
EWMA	0.38 $\pm$ 0.05	0.41 $\pm$ 0.079
Bayesian tracking	0.232 $\pm$ 0.021	0.236 $\pm$ 0.028

Table **4-1**: Root Mean Squared Error and its tracking deviation for the three evaluated methods. These averages have been calculated from the differences between the control and any of the predicted data, expressed as percentage height.

Table **4-1** presents the mean and the standard deviation of the Root Mean Squared Error (RMSE) between the tracked data and the control series for the three methods during twenty gait cycles, which correspond to about 30 s of gait, distributed among four different subjects. The higher tracking error is observed for the COM marker, a result that may be attributed to the fact that this is not actually placed close to the CoM as well as to the high noise levels in the acquisition process. Indeed, the

EWMA decreases these error rates at smoothing these noise levels, but the error is still unacceptable for many applications, such as the prostheses fabrication, in which the prostheses mean life is absolutely dependent on the accuracy design [155, 228]. The particle filtering presents the lower errors, approximating better the control data since this model is not only capable to adapt to the marker position variability but also to the non linearities in the trajectory. These results strongly suggest this method is steadier and more accurate for the tracking task.
Since the whole problem consists in following temporal series which are highly non linear and whose dynamics is therefore very difficult to determine, it is important to establish a metric on which it would be possible to measure the level of agreement between two trajectories. Herein we have measured this concordance level using the correlation coefficient of the temporal differences between two series. The correlation coefficient measures the degree to which two things vary together or to which two things draw apart together. In this case, the concordance level was evaluated as the correlation coefficient from the temporal differences between series.

Tracking Method	x-axis	y-axis
Marker CoM	0.54 $\pm$0.05	0.7849 $\pm$0.081
EWMA	0.56$\pm$ 0.05	0.7584 $\pm$ 0.079
Bayesian tracking	0.9 $\pm$ 0.021	0.91$\pm$ 0.028

Table **4-2**: Correlation factor of the differences between the control and the three assessed methods.

Table 4.1.3 shows the correlation factor of the temporal differences between the kinematic data of the three methods and the control series during twenty gait cycles, in the x and y axes. Again, while the Marker CoM and the EWMA methods hardly follow the control data, the particle filtering tracking outperforms these two methods and definitely correlates with the control data (Correlation factor of about 0.9 in the two cases). A statement which is easily inferred from the correlation coefficients presented in table 2. Furthermore, figures in this table show a high correlation in the x and y axes, but importantly a better performance of the particle filtering method in the two dimensions, for instance, from 0.7 to 0.9 in the y-axis and from 0.5 to 0.9 in the x-axis. The results prove a natural tracking of the control signal obtained by the particle filtering method, i.e., the predicted data properly scale and shrink. In contrast, the CoM Marker or EWMA predicted data are inadequate for tracking the different phases of the gait and hence inappropriate to detect small changes, an issue which results crucial for identifying certain pathologies [262].
In despite of the so far better performance of the particle filtering method, provided

that changes are in any case very small, it is very difficult to follow differences through the time and the correlation coefficient is a global measurement. In particular, it is really difficult to figure out the real gain of any method at any time, since results are always contaminated by the particular noise in measurements at any of the two dimensions and differences, as said before, are really small. In consequence, the quality of the prediction was weighted by the noise, using a logarithmic scale and measuring the difference between the expected (control data) and the predicted values. This SNR-like or quality performance measure reads as

$$q((x_c, x_m), (y_c, y_m))^{dB} = 10 \log \left[\frac{1}{(x_c - x_m)^2 + (y_c - y_m)^2} \right] \tag{4-17}$$

where (x_c, y_c) are the ground truth coordinates at the time t and the (x_m, y_m) are the signal coordinates at the same time. The great advantage with this measure is that it allows a temporal gain follow-up of the tracking. Overall, the most important component is the vertical (y-axis) since most pathologies alter mainly this vertical pattern [110, 262]. Therefore, this quality measurement was modified by weighting the horizontal and vertical components with a coefficient α, whereby we could tune the importance of the vertical direction. Then, this quality measure can be written as:

$$q((x_c, x_m), (y_c, y_m))^{dB} = 10 \log \left[\frac{1}{((1-\alpha)(x_c - x_m)^2 + (\alpha)(y_c - y_m)^2)} \right] \tag{4-18}$$

This evaluation was performed using the cycle for which the averaged difference between the control and each of the temporal series was the smaller, when comparing with the entire sequence of differences among gait cycles. Quality was then measured using α values of 0.9, 0.8 and 0.7 and results were plotted in figure 4.1.3. The four graphs show in the y-axis the gain in decibels for each of the methods and in the x-axis the cycle with smaller averaged difference. For comparing so, each cycle was expressed as percentage since overall each cycle spans a different time interval. The four panels are distributed as follows: Upper left panel is the first quality measure, that is to say, there is no α, upper right panel corresponds to the SNR-like measure with $\alpha = 0.7$ and left and right bottom panels correspond to this measure with α set to 0.8 and 0.9, respectively.

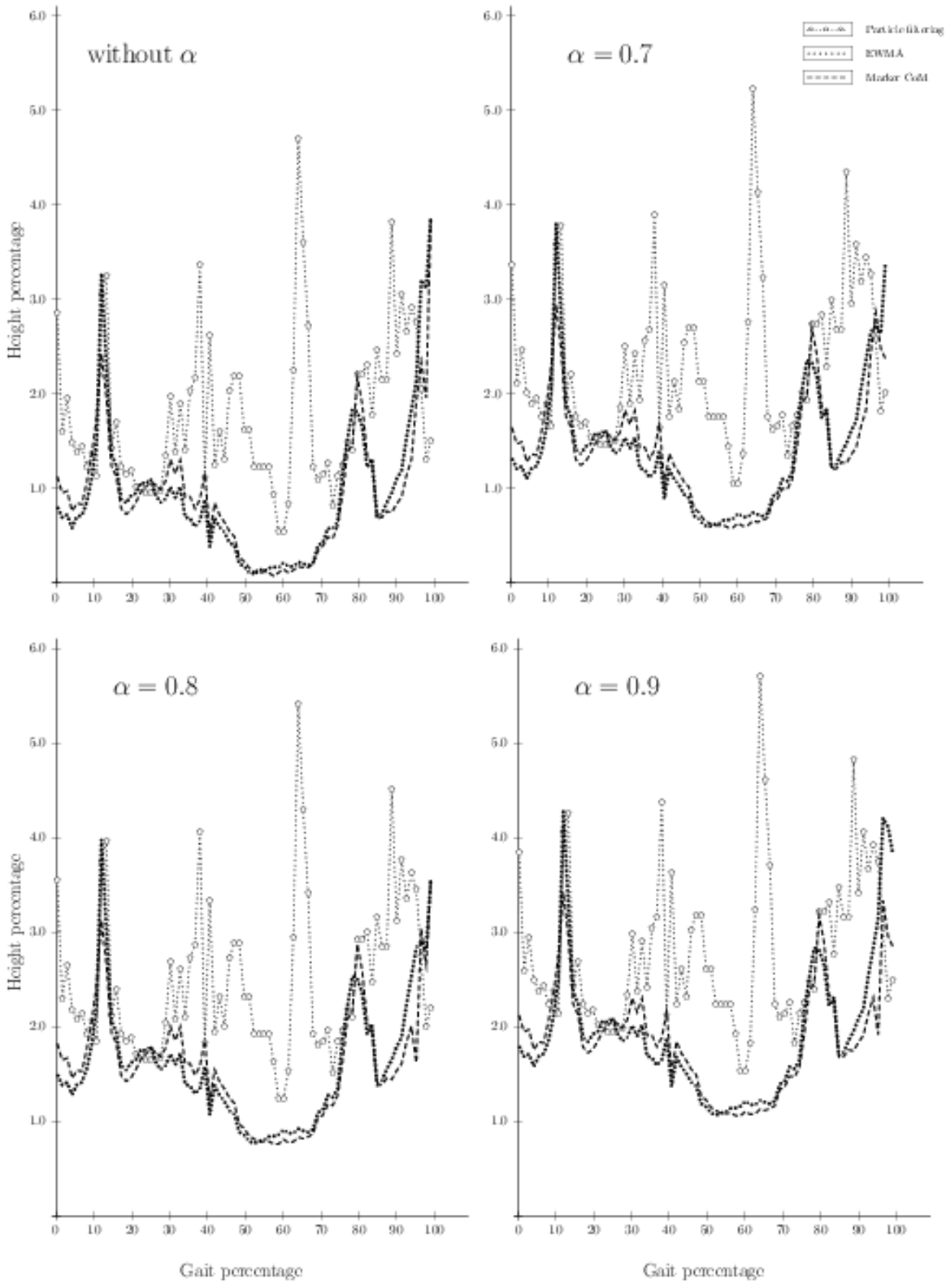


Figure **4-4**: The SNR-like measurements assess the fidelity to the control data along the best cycle for each of the methods. The graph shows in four panels different

In the image processing community, the PSNR measure is a term for the ratio between the maximum possible power of signal and the power of corrupting noise that affects the fidelity to the original data. Typical values for the PSNR in lossy image and video compression are between 30 and 50 dB, where higher is better. In the case of our evaluation a decibel entails a large difference and this accounts as a quality measurement of the fidelity to the control data. Overall, plots in the four panels show the same pattern, that is to say, the particle filtering always outperforms the other two methods. The pattern of the four panels results to be very alike at the beginning, but rapidly after the gait has reached a 30 % of the entire cycle, the particle filtering strategy shows a larger gain, between one and three decibels. The larger difference appears when the cycle has spanned about a 60 %, as expected since at this time the gait pattern is much more non linear because of the heelstrike and the particle filtering is better suited to following this kind of discontinuities. Likewise, the introduction of a particular preference for the vertical direction in the SNR-like measurement has no influence on the results. Basically, the difference between plots with α set for instance to 0.7 (upper right panel) and the SNR-like measurement with no α (upper left panel) is about 0.5 decibels, and most punctual differences are slightly amplified, indicating that the x-axis or temporal evaluation contribute very little to the whole variability of the SNR-like measurement. This result agrees, with what is known in the literature, in the sense that most important evaluation must be performed upon the vertical direction [110, 262]. Finally, the analysis performed setting α to 0.8 and 0.9, shows the same trend, i.e., the vertical direction weights more and differences are more important, demonstrating that the particle filtering strategy is able to closely follow these non linear patterns, and to more accurately predict this vertical movement.
The SNR-like assessment was also pormed on the twenty cycles of the three patients and the mean with its standard deviation were calculated from this set of data. erf

Tracking Method	without α	$\alpha = 0.9$	$\alpha = 0.8$	$\alpha = 0.7$
Marker CoM	1.35 ± 0.35	2.34 ± 0.39	2.04 ± 0.39	1.86± 0.36
EWMA	1.33 ± 0.46	2.32 ± 0.46	2.03 ± 0.46	1.85± 0.47
Tracking Bayes	2.17 ± 0.5	3.17 ± 0.5	2.86 ± 0.5	2.7± 0.5

Table **4-3**: The mean and standard deviation of the SNR-like mesurement for a total of sixty cycles available in the present investigation

Table 4.1.3 shows the mean and standard deviation of sixty cycles from three different patients. In general then, among the three methods, the higher quality gains are observed for the particle filtering, as was illustrated before in Figure 4.1.3 for the best cycle. These differences are larger when evaluation has no α value, as expected since

the particle filtering outperformed the other two methods in the two directions. In this particular case, the trajectory generated by our model had an approximated quality of 2.17, compared to the averaged 1.34, obtained by the conventional methods.
When the vertical preference is introduced into the SNR-like measurement, differences decrease but the Bayesian tracking conserves the better performance at any of the set α values. The results with this modification show a considerable gain (mean of 3.17 for $\alpha = 0.9$), whereby the three methods track much better the vertical movement but the particle filtering presents still the best performance so that its utility in clinical application could be even better. Figures for the EWMA and Marker CoM methods look very similar, in despite the Marker CoM estimation is a very noisy.
The accuracy of the particle filtering method prediction is of course dependent on the goodness of the approximation to the *pdf*. This is essentially an issue which is function of the number of particles used to get this approximation. Accuracy was herein defined as the averaged RMSE for each of different number of particles. The Bayesian fidelity to the control data was thus assessed using a different number of particles so that we could establish the influence of this parameter. Since the whole performance picture includes also the computational time, the running time is also described in table 4.

Particle	x-axis	y-axis	Duration ms
10	0.451 ± 0.022	0.423 ± 0.039	170
100	0.281 ± 0.023	0.29 ± 0.028	810
1000	0.253 ± 0.025	0.252 ± 0.031	7100

Table **4-4**: Accuracy index or the averaged RMSE with its standard deviation for different numbers of particles from the data of vertical and horizontal directions. The running time is included for illustrating the computational cost of any of them.

Table 4.1.3 shows a high gain obtained with 100 particles when comparing with 10, but a smaller one when 1000 particles are used instead. On the contrary, the variance is comparable in the three cases, indicating a systematic error decreasing with the number of particles, as expected. This statement holds true for the two dimensions (x and y) but this error reduction results very small when using 1000 particles. Computation times evidence an important increasing at using 1000 particles, when comparing with 100. It is very likely that a time of $7\,s$ is incompatible with real time applications while the gain in error reduction is actually small, a trend which will be even bigger for larger numbers of particles since this exponential difference is a well known characteristic of this type of algorithm [16, 100].
Finally, it is worthy to point out an efficiency issue within this evaluation. Efficiency

was herein conveniently defined as the precision per unit cost. For the present evaluation, the precision of a final estimate was expressed as the reciprocal of the data variance (σ^2) and the cost as the running time t. Hence, an index of efficiency E was computed as $\frac{1}{\sigma^2 t}$

Particle	x-axis	y-axis
10	0.02673	0.15082
100	0.05367	0.044091
1000	0.0056	0.00454

Table **4-5**: Efficiency index for both dimensions and running time for a number different of particles.

Table 4.1.3 shows a larger efficiency when using 100 particles, a statement that was difficult to figure out from data in table 4 and which illustrates the needed balance between accuracy and the computational cost, an important demand in real time applications.

4.1.4 Discussion

This article has introduced a novel strategy for the CoM to be closely followed during actual body gait cycles, which is based upon some probabilistic considerations, namely, there exists an extensive domain knowledge about the gait physiology and the whole process can be modeled as a Markovian process so that the future system states are stochastic functions of the past system states. The particle filtering framework used in the present investigation nicely dealt with this highly non linear gait dynamics, based on the calculation of two complementary terms: a prior adapted from a well known mathematical gait model, whose aim is to predict the gait cycle [109], and which is constantly tuned by a function that expresses the link between observations and system states (the likelihood). This approach is simple, easy to implement when the *pdf* is approximated by a particle filter and efficient in terms of computational running time and accuracy on the obtained measurements.
As extensively discussed before, the CoM has been a quantity difficult to establish due to its highly non linear dynamics. Previous works [86, 110] have presented significant differences between optical methods and the CoM estimated from the double integration of the reaction force, measured in ordinary force plates. Many authors consider that the CoM trajectory, determined from the reaction force, is the gold standard in the CoM examination [86,110,283,284]. Biomechanical literature is highly rich in methods whose aim is to estimate the CoM position. Nevertheless, their clinical use remain lim-

ited because most commercial laboratories are provided with one or two force plates, a set up that hardly adapts to the step length variability since, at the best scenario, one entire gait cycle is assessed. Besides, repeatability of the gait analysis is poorer with these force plates and so the analysis becomes inconsistent. Overall, there have existed two main approaches: the first has been a very practical solution at associating the CoM temporal pattern to a marker located within the sacral zone, defined by an expert. Yet its accuracy is enough in many clinical studies, as it was shown in the present investigation, this estimation hardly follows the non linear CoM patterns, very likely because the marker dynamics is by itself non linear and not necessarily of the same type. This picture would be worst in those disorders in which the gait is so altered that it is really difficult to define even a cycle pattern. In practice, this approximation has consisted in setting a point where the CoM is supposed to be, namely, a 60 % of the body height, a point from which there exists considerable clinical evidence for being the candidate as the more probable CoM location [110, 291]. However, this method is particularly inaccurate in pathological situations in which the relative movements of the body segments are very distorted. On the other hand, a second strategy borrowed from optical methods consists in combining the CoM optical estimation for each of the body segments into a resultant CoM [86]. The main drawback with this method is that the CoM location for each body segment is by itself a high non linear problem and therefore very inaccurate. Moreover, the error will be propagated systematically from each of these segment estimations because of the anatomical segment variability, as well as because the variable marker placement [110]. Setting a convenient marker location may be particularly difficult in patients with high anatomical variability and unfeasible in patients using ortheses. In addition, movements of the body segments lead to relative displacements of the original marker locations, an intrinsic error which is almost impossible to avoid. Even worst, many pathologies are characterized by accentuated or attenuated movements and hence most marker placement protocols fail under these extreme conditions. This performance is improved by increasing the number of markers and so the CoM estimation, nonetheless a large number of markers alters the natural gait gesture. In despite that these optical methods are the more used in clinics because of their easy implementation and control, they are still very inaccurate in calculating with sufficient precision the CoM location [110] and therefore inadequate for many diagnosis and follow-up. The strategy herein proposed, although is also based on the marker measurements, it depends initially on the location of four markers, but above all upon the underlying gait dynamics which drives the movement. Moreover, the method introduces efficient mechanisms for tracking the gait trajectory involving non linear descriptions of the human movements. Among other advantages, this strategy requires a small number of markers (four for the present investigation)

and can even be used in patients with ortheses since the sacral region is not invaded with any device, in these cases.

The trajectory described by the CoM during a gait cycle is a global indicator that correlates with the gait efficiency but that also can be used as a dynamical variable which complements the standard gait analysis [76]. The CoM temporal path is distorted in different movement abnormalities and associated with a degree of illness [76]. The CoM is considered as an efficient indicator for assessment of pathologies such as hemiplegia, paraplegia or distonia. The optimal walking, in terms of the energy, can be defined as the movement of the CoM from a place to another with minimum energy expenditure. Therefore, a pathological gait can be analyzed in terms of energy using the CoM change as the transfer of potential to kinetic energy (recovery) ,i.e., normal gait patterns loss a 40 % of this energy in this transfer, a higher lost is pathological [76]. Indeed a proper gait analysis should be based on an accurate estimation of the CoM positions and for doing so, the analytical description of the CoM movement pattern is required. The inverted pendulum system is an appropriate mechanism to representing the CoM movement in energy terms, describing the exchange between kinetic and potential energies within the different gait cycles [40, 52, 76]. Garcia's model has succeeded about representing the whole system in these terms [109] and is able to describe normal patterns. The strategy presented here takes this model as a prior, adapts it to the observations and demonstrates its effectivity for tracking such patterns with no major concern on the anatomical variabilities, i.e. this hypothesis allows to assume a fixed leg length. Importantly, the strategy herein described could be easily used to follow pathological patterns by simply replacing this prior (see [167] for models that describe pathological patterns), or changing the weight given to the likelihood function so that even with this prior, the mapping of the observations to the states could follow the actual pattern. We claim that this is possible since a mixture of Gaussians has been already used in other problems [222] for tracking non linear dynamics with comparable accuracy rates. The prior here is needed because it gives physical meaning and more importantly, clinical meaning, to the possible altered patterns and so quantification makes sense.

An important part of the routine clinical examination is the CoM estimation per cycle, a basic descriptor of many neuromuscular and musculoskeletal disorders. So far the common point, among the different approximations, has been a heuristic detection of the CoM location. Yet these approaches have coped with many clinical needs, no endeavours have been dedicated, up to now to developing tracking methods [86, 262]. The point is that any strategy, devised to uniquely achieve detection, is inefficient because it is based on noisy position mesurements and highly dependent on their inital marker location [76, 85]. In the present work, we propose an efficient strategy for

tracking CoM locations, using a non-linear gait biomechanical model whose parameters are recursively adjusted by a Bayesian filter, implemented as a particle filtering. Implementation of any Probabilistic function distribution (*pdf*) can be achieved using several techniques, namely, extended Kalman filtering, multihypothesis tracking, grid-based or topology-based representations and particle filtering [16, 99]. The Extended Kalman Filter approaches the system dynamic by a first order Taylor series expansion. This filter is useful if the state uncertainty is not too high, i.e., measures come from accurate sensors, which is obviously not the case for CoM tracking. The multihypothesis tracking represents the belief as a mixture of Gaussians and tracks each with a Kalman filter. This technique is computationally expensive and requires complicated heuristics to determine when to add or delete Gaussians. Additionally, the heelstrike rule, herein used for the dynamic model, introduces angular discontinuities that would require many Gaussians whose number would be impossible to establish beforehand. Grid-based approaches stand for a piecewise constant representations of the belief. This approach approximates discontinuities by refining the resolution grid cells, and therefore expensive computational methods are required when discontinuities are present. This grid complexity could be approached by topological representations, corresponding to a graph where each node is related to a state and each edge to the enviroments conectivity, but the computational cost is even higher. In contrast, each particle of a particle filtering can easily follow any discontinuity, with a low number of particles and minimum computational cost. Moreover, the particle filter does not require accurate measurements because the particle weights are modified proportionally to their likelihood.

The whole strategy allows a natural tracking of the non-linear gait patterns with a high degree of noise robustness, under a non-linear minimal square criterion. Accuracy and efficiency were herein assessed under several different metrics. Firstly, under RMSE metric the particle filtering is a 56 % smaller in the x-axis and 59 % in the y-axis, when comparing with other two conventional tracking strategies. Also, correlation was assessed using the correlation coefficient and again the particle filtering outperformed the other two from 0.9 to 0.7 in the x-axis and from 0.9 to 0.6 in the y-axis. Finally, the SNR-like measurement permitted to measure the quality of the tracked signal among the different strategies under the Gaussian noise conditions, introduced from different independent sources. On average, measurements showed that the trajectory generated by our model was $0.86\,db$ higher, when comparing with conventional methods, indicating a large gain in accuracy. Importantly, the largest gain difference was observed in the heelstreak phase ($1.5\,db$), demonstrating the capacity of the proposed method to follow the non linear gait patterns. Finally, the method was computationally implemented through a particle filter because of the great advantages of this implementation

when comparing with others [100], among others: accuracy, robustness, efficiency and easy implementation. The particle filter is herein used to estimate the gait Bayesian model. The Sequential Importance Resampling implementation of the particle filter approximates the *pdf* by a weighted set of particles whose importance is constantly evolving with the dynamics we introduced and then updated after system observations are available. This implementation, as shown by our results, is efficient in terms of the computational time and enough accurate as to follow the gait non linear patterns so that on-line data analysis is possible along with the routine capture of the other variables which compose an entire gait analysis.

4.1.5 Conclusions and Perspectives

This work presented a general framework for tracking complex human movements. The whole strategy consists in simulating the dynamics of the system, using some priori information of the particular problem. Simulation requires a discrete system model and an observation process. The model approximates the truth hidden system states, while the observation process adapts to non-linear dynamics using the Bayes rule, implemented as a particle filter. The method was successfully assessed. The presented procedure could be extended to other type of medical imaging problems, under the restriction that there exists a proper knowledge of the problem so that analytical or parametric expressions may be found.

4.2 Simulation of normal and pathological gaits using a fusion knowledge strategy

Gait distortion is the first clinical manifestation of many pathological disorders. Traditionally, the gait laboratory has been the only available tool for supporting both diagnosis and prognosis, but under the limitation that any clinical interpretation depends completely on the physician expertise. This work presents a novel human gait model which fusions two important gait information sources: an estimated Center of Gravity (CoG) trajectory and learned heel paths, by that means allowing to reproduce kinematic normal and pathological patterns. The CoG trajectory is approximated with a physical compass pendulum representation that has been extended by introducing energy accumulator elements between the pendulum ends, thereby emulating the role of the leg joints and obtaining a complete global gait description. Likewise, learned heel paths captured from actual data are learned to improve the performance of the physical model, while the most relevant joint trajectories are estimated using a classical inverse kinematic rule. The model is compared with standard gait patterns, obtaining

a correlation coefficient of 0.96. Additionally,themodel simulates neuromuscular diseases like Parkinson (phase 2, 3 and 4) and clinical signs like the Crouch gait, case in which the averaged correlation coefficient is 0.92. *The complete content of this section has been published as a research article in the journal of* ***NeuroEngineering and Rehabilitation*** *(see [91]).*

4.2.1 Background

Quantification of complex movements such as human locomotion is a fundamental step towards an objective characterization of particular patterns associated to a certain degree of a disease [209, 223, 225]. The gait is the result of complex interactions between several sub-systems: neuromuscular, musculo-tendinous and osteo-articular, which work together to generate the body dynamics that underlies the bipedal displacement [295,296]. In despite of the intensive research in biomechanics [103], robotics [59, 89], medicine [102] and computer animation [72, 73], the biological complexity has hindered a proper understanding of the locomotor system. This problem has been partially overcome in the clinical routine by a gait estimation inferred from the gait laboratory [102, 220, 293]. Usually, a physician or rehabilitation expert determines whether there exist pathological gait patterns using exclusively her/his expertise [106, 168, 296]. Overall, diagnosis is supported using statistical tests carried out on the acquired gait laboratory data [14, 22, 60, 128, 200], with an inherent high degree of variability. In consequence, development of gait models that provide a quantitative gait description has become important in the process of supporting physician decisions [102, 106, 252, 296]

The main contribution of the present work is a human gait model that accurately describes a set of kinematic gait patterns, normal or pathological. The model fuses two important gait information sources: an estimated Center of Gravity (CoG) trajectory and heel paths learned from actual gaits. The global motion is governed by the CoG trajectory of a compass physical pendulum representation, coupled to a spring that emulates the muscle function. This trajectory is regulated by learned heel paths, while the remaining joint patterns are estimated using a classical inverse kinematic method. The models benefit is demonstrated by accurately simulating two different sorts of neuromuscular gaits: Parkinson and Crouch patterns. Finally, a human-like leg structure is animated with the obtained trajectories, allowing the clinician to interact with the model and facilitating the interpretation of an observational analysis.

Many models have been previously proposed for simulating the human gait, with different complexity levels, depending on the application area. A first group includes bipedal descriptions that exclusively use structural information so that they are able only to determine global relationships between muscles and joint angles. These models exploit

the conceptual simplicity of mechanical systems such as the inverted pendulum or mass-springs [101, 111, 176, 203, 282]. Basically, these approximations provide a locomotion description from an energy standpoint, simulating the change from the kinematic to potential energy during the gait cycle. These models are devised to coarsely classify normal and pathological patterns [200, 292]. However, a main drawback of these approximations is that about a 20 % of the gait cycle, corresponding to the double stance phase, is completely eliminated. These physical models are useful in areas like robotics since they eliminate the dependence on a robust control mechanism. Nevertheless, they are very limited for medical applications because of their strong simplifications, missing relevant gait aspects such as the non-linearities introduced by the heel strike.

A second group of human gait models are capable of simulating muscles and tendons during the gait. These models have obtained better gait representations, introducing muscular information that is required from a clinical standpoint in terms of interpretability, i.e., specific activity of certain muscle groups in musculoskeletal disorders like hemiplegic movements. These models have introduced new elements to simulate the control and energy storage of the locomotion process. Specifically, some gait approximations include the Hill model as the base of the muscle representation [168, 295, 296], but with no relation between the muscle and the locomotor structure and hence without any clinical meaning [142]. In these approaches, each model accelerates a specific body segment, obtaining a simplistic simulation of pathological movements. Likewise, these models are not accurate enough to describe the complex interaction among different groups of muscles. In addition, they require a certain number of parameters that need to be tuned, with the consequent dependence on an expert knowledge. Scott Delp [71, 72] introduced a computational strategy that combines the Hill muscle model and structural information, accomplishing realistic normal and pathological simulations, but again, with a high degree of subjectivity at tuning the model parameters. Currently, several approaches have used some control-based strategies, requiring relatively few data to simulate simple human structures and predicting new motions [264]. These approaches include a large number of degrees-of-freedom while joint force profiles remain subjected to a large number of constraints [14, 73, 102, 137, 292]. These methods approximate human control systems and simulate some neurological pathologies [168], but these strategies require specific information about each particular motion to be simulated and therefore they demand a high degree of interaction and prior knowledge [292]. Moreover, these methods necessitate a large group of experimental data to generate natural motions so that their clinical usefulness still remains very limited.

4.2.2 Materials and Methods

The present work simulates normal and pathological kinematic patterns by fusioning two important sources of information: a prior model of the $CoG_{x,y}(t)$ and real data trajectories. The proposed model is summarized in Figure 4.2.2. Firstly, the prior knowledge of the $CoG_{x,y}(t)$ is introduced using a physical gait model, a compass pendulum with springs coupled to both ends, representing the role of the knee and smooth tissues (see Figure 4.2.2(b) and a further description in section 4.2.2.1). The inclusion of these non linear elements allows more accurate estimations of the $CoG_{x,y}(t)$. A second information source comes from actual heel trajectories that are used to regularize the estimated $CoG_{x,y}(t)$ and serve to simulate diverse pathological and normal motion (see Figure 4.2.2(a) and section 4.2.2.4). Additionally, this fusion facilitates an accurate estimation of the remaining joint trajectories, using a classical inverse kinematic framework. Finally the set of obtained trajectories animates a human-like leg structure that provides the clinician with a interpretable tool (see Figure 4.2.2(c-d)).

Figure **4-5**: Pipeline of the proposed model. First the $GoG_{x,y}(t)$ from the proposed physical model is computed (panel (B)). Additionally heel trajectories are learned for each kind of movement (panel (A)). Then, a fusion rule to compute kinematic patterns (Panel (C)) from both trajectories allows to simulate Normal and Pathological patterns (Panels (D) and (E)).

4.2.2.1 $CoG_{x,y}(t)$ gait representation

In human movement analysis, the gait is divided in cycles, coarsely classified as double and single stance phases [18, 106, 287]. The double stance period accounts for around 20 % of gait cycle and stands for the body movement with both limbs touching the ground, while the single stance represents around 80 % of gait cycle and corresponds to the interval in which a single limb supports the whole body weight. In this work the $CoG_{x,y}(t)$ for a complete gait cycle is approached using two complementary strategies: a compass pendulum for the single stance and a spring mass system for the double stance, as follows.

4.2.2.2 The single support phase

The single support phase conserves a regular periodicity which is properly captured using a compass pendulum representation. This strategy represents the upper part of the body by a mass M which moves forwards with respect to each fixed point (with

mass m), describing a harmonic oscillating trajectory, similar to the inverted pendulum [175,203]. Likewise, the free foot swings with respect to this mass, establishing a simple pendulum pattern. Provided that these processes are coupled together, the human gait is modeled by a compass pendulum as two coupled non-linear differential equations:

$$\begin{aligned} &\beta(1-\cos\phi)(3\ddot{\theta}-\ddot{\phi})-\beta\sin\phi(\dot{\phi}^2-2\theta\dot{\phi})+(\tfrac{g\sin\theta}{l})(\beta(\sin(\theta-\phi)-1))=0 \\ &\ddot{\theta}(\beta(1-\cos\phi))-\beta\ddot{\phi}+\beta\dot{\theta}^2\sin\phi+(\tfrac{\beta g}{l})\sin(\theta-\phi)=0 \end{aligned} \tag{4-19}$$

where $\beta = m/M$, θ is the angle of the stance leg at the particular time t with respect to the slope and ϕ is the angle between the stance leg, and $l_0 = l_r = l_l$. This model also allows to simulate the swing foot when it hits the ground at the heelstrike, a time in the cycle that corresponds to $\phi(t) - 2\theta(t) = 0$ [175], when the double stance starts.

4.2.2.3 Double stance phase

Classical gait models often ignore the double support stance since they have been devised to simplify the gait rather than to accurately follow gait patterns. These simplifications have ended up by considering the leg structures as rigid segments, a hypothesis that easily leads to conclude for instance that the percentage of gait recovery is inefficient in energy terms, a reason why this phase has been eliminated in most of these strategies [102, 108, 252, 296]. Additionally, important elastic contributions which produce relevant changes in the $CoG_{x,y}(t)$, during the double stance, are often neglected. These strong simplifications reduce an appropriate gait understanding and may lead to wrong interpretations when these models are used as supporting tools of clinical decisions.

A more accurate $CoG_{x,y}(t)$ description of the double stance phase was herein achieved by coupling a planar spring-mass system [33] to the compass pendulum, previously introduced. This change of the leg length l during the gait stance phase, allows to estimate the reaction force during the whole gait cycle, as illustrated in Figure 4.2.2(B). Notice that each leg reaction forces points out towards opposite sides, separated by a distance d (the distance between the heelstrike and the other toe-off phase). The coupling is obtained as:

$$\begin{aligned} M\ddot{x} &= l_l x - l_r(d-x) \\ M\ddot{y} &= l_l y + l_r y - gM \end{aligned} \tag{4-20}$$

where g is the gravity, l_l and l_r are the left and right legs, respectively and their length changes as:

$$\begin{aligned} l_l &= k(\frac{l_0}{\sqrt{x^2+y^2}} - 1) \\ l_r &= k(\frac{l_0}{\sqrt{(d-x)^2+y^2}} - 1) \end{aligned} \tag{4-21}$$

These equations simulate the periodic vertical ground forces, with a period defined by $T = 2\pi\sqrt{\frac{m}{k}}$. This independent formulation of each reaction force allows an independent analysis of each link, whereby gait abnormalities that asymmetrically affect each leg, such as the diplegia, can be simulated. Finally , the $CoG_{x,y}(t)$ is simulated by the integration of the two gait phases described as follows:

$$CoG_{x,y}(t) = \begin{cases} l_0 \left[\sin\theta(t), \cos\theta(t)\right] & \text{if} \quad \phi(t) - 2\theta(t) < 0; \\ \left[\frac{l_l \frac{x^3}{6} - l_r(d - \frac{x^3}{6})}{M}, \frac{l_l \frac{y^3}{6} + l_r \frac{y^3}{6} - gM}{M}\right] & \text{elsewhere.} \end{cases} \tag{4-22}$$

4.2.2.4 The fusion information strategy

Although the $CoG_{x,y}(t)$ is a fundamental clinical descriptor [200], a useful identification of a particular disorder also requires a proper gait analysis of other anatomical joint trajectories. Accordingly, a more complete gait description was herein achieved by the fusion of two important sources of information: the physical gait strategy previously described and the learned heel trajectories.
The learned heel trajectories were modeled as a set of normal distributions with mean μ_i and variance σ_i^2 from three different groups of patients captured in a gait laboratory as:

$$\psi_{x,y}(t) = \sum_{i=1}^{I} w_i N(t\ \mu_i, \sigma_i^2)$$

where I represents the total number of learned gait movements (normal, Crouch and Parkinsonian gaits). Each gait distribution was computed from 30 gait cycles belonging to 10 patients (7 men and 3 women). From this multi-gaussian distribution model, we can select a heel trajectory i to regularize the $CoG_{x,y}(t)$ associated to a particular gait movement. Likewise, the normal motion distribution allows a large variety of gait patterns of the same pathology. New relationships are inferred from these two trajectories by assuming the knee joint position as $r_{x,y} = \frac{l_{0x,y}}{2}$ [75]. Afterward, a classical inverse kinematic method is adapted to obtain two main kinematic patterns: the flexion-extension patterns of the hip $\omega(t)$ and knee $\gamma(t)$. For doing so, at each time

t of the gait cycle, a CCD method performs an iterative rigid transformation over each couple of joints. The two patterns are defined as:

$$\gamma(t) = acos\left(\frac{CoG_{x,y}(t)^2 - r_{x,y}^2 - r_{x,y}^2}{2}\right) \tag{4-23}$$

$$\omega(t) = atan2(\psi_{x,y}) + atan2(r_{x,y}\sin\gamma, r_{x,y} + r_{x,y}\cos\gamma) \tag{4-24}$$

where r is the distance between the $CoG_{x,y}(t)$ and $\psi_{x,y}(t)$. Unlike other approaches, this model estimates kinematic patterns with medical meaning, but the model can also obtain energy and ground force patterns for normal and pathological cases, obtained from the $CoG_{x,y}(t)$.

4.2.2.5 Building up a human leg structure

Finally, a human-like leg structure is animated using the set of kinematic patterns described above. This structure may be used as a clinical interpretability tool. For doing so, we define a human structure composed by a set of 12 rigid elements, connected together, as shown in Figure 4.2.3. The lower limbs follow a dynamics established by the proposed model, while the upper limbs are normal trajectories computed from real data of the gait laboratory.

Figure **4-6**: The figure shows the human-like leg structure used to simulate the set of kinematic patterns. It can be used as software tool that allows interpretation and interaction.

4.2.2.6 Modeling pathological movements

The proposed model is also capable of simulating pathologic patterns such as the spastic diplegia (typically represented by a Crouch Gait) and Parkinson, an advantage with respect to other classical models.
Firstly, the model is used to simulate a Crouch gait. This motion is produced by a neuro-muscular disorder known as the spastic diplegia that is characterized by the presence of muscle rigidity and loss of muscle force, affecting predominantly legs, arms and face. The clinical signs include gait pattern distortions of the sagittal plane, like bent-rigid knees, flexed hips and certain anatomical changes like lumbar lordosis. Such signs were herein modeled by setting the spring constant to values close to the estimated

leg springiness. The resultant kinematic patterns are thus related with an increase in the energy consumption, showing a flexion rigidity of the hip and knee. The model is also used to simulate typical Parkinson gait patterns. In this case, the gait patterns are produced by a degenerative disorder of the central nervous system and are characterized by rigidity and slowness of the human movements: this gait is characterized by short steps. These Parkinsonian gait features were captured by fixing the d and k parameters, associated to the step length and the the knee flexion-extension, respectively. This representation results directly related to the energy consumption since in this case a particular displacement demands more energy than that required during a normal gait. Likewise, the kinematic gait patterns are characterized by a higher frequency than the observed for a normal gait. This kind of patterns can be modeled and simulated to approximate different phases of the disease, allowing thereby to objectively characterize the pathology.

Simulation of these pathologies requires to set the l_0, k and d parameters, using actual patient data. For the Crouch gait simulation, the spring constant was fixed within a range of $k = 350$ to 400, the $d = 0.65$ and the l_0 length was reduced to 5 % of the initial 1 m, according to well known biomechanical parameters [128, 209, 223, 225]. For the Parkinsonian gait simulation the d parameter was obtained from actual data and set to $0.58\,m$, the k was set to 500 and $l_0 = 1$ meter. The heel paths fitted a normal distribution and were learned from actual patient data, as previously explained. The simulated trajectory precisely follows the different components of the abnormal gait pattern, in particular the flexed knees, the $CoG_{x,y}(t)$ attenuation and the step length, as shown in Figures 4.2.3 and 4.2.3 and reported in next section.

4.2.3 Evaluation and Results

Evaluation was carried out by comparing the estimated gait kinematic patterns with ground truth trajectories, of normal and pathological patterns, as reported in [106, 223, 252]. A quantitative evaluation was performed by calculating the correlation coefficient and the Fréchet distance between both trajectories, which are composed of temporal x and y paths, belonging to a single gait cycle.

A first part of the evaluation consisted in determining the $CoG_{x,y}(t)$ relation of two models, the physical model herein proposed and a classical compass pendulum model w.r.t the ground truth [173]. Figure 4.2.3 shows the decomposed motion for both models: the x axis representing the percentage of the gait cycle and the y axis the vertical displacement with respect to the body height, also in percentage. Both models follow a $CoG_{x,y}(t)$ periodic sequence, but the classical compass pendulum model systematically misses the discontinuity introduced by the heel strike, much more important in the x

axis, while the proposed model accurately predicts this part of the cycle. Notice that the heel strike of the contralateral foot (toe off) actually occurs at about a 60% of the gait cycle.

Figure **4-7**: Simulation of Crouch gait patterns.In the left panel it is illustrated a sequence of human poses obtained from the proposed model which better tracks actual patient poses. In the right panel it is presented the joint angle trajectory (red starred line) obtained with the proposed fusion strategy and compared with ground truth patterns (shadowed gray zone with mean trajectory represented by green line).

Figure **4-8**: Examples of the proposed method performance when estimating the dynamics of the CoG movement in different stages of the Parkinson disease: Second stage (A), third stage (B), fourth stage (C).

A second part of the evaluation compared the hip $\gamma(t)$ and knee $\omega(t)$ joint-angle patterns of simulated normal gaits and ground truth patterns. For this assessment, the fusion strategy used two different $CoG_{x,y}(t)$ estimations: the proposed physical model and the classical compass pendulum previously described. The hip $\gamma(t)$ and knee $\omega(t)$ patterns were expressed as joint-angle variations at the y-axis and plotted against the gait percentages at the x-axis, previously weighted by the entire duration of a cycle. Figure 4 shows the ground truth and the predicted gait patterns for a sagittal view (right and left) of a complete cycle, using the two $CoG_{x,y}(t)$ estimations. The joint-angle trajectories computed from the fusion strategy show a very close Correlation Coefficient (CF) w.r.t the ground truth patterns (CF = 0.8 using the compass pendulum, CF = 0.9 using the herein proposed physical approach). During the single stance phase, the angle trajectories computed from both CoG paths have a high correlation, nevertheless the conventional pendulum misses about a 40 % of the angle variation because of the nonlinearity introduced by the heel strike and therefore the curve correlation also falls down. In contrast, the joint-angle trajectories obtained from the fusion strategy with the proposed physical model follows the actual gait paths and its correlation coefficient remains larger than 0.8. Significant differences were then reported with the conventional compass pendulum, specifically for the part of the cycle dominated by the heel strike.

Pattern	R.Hip	R.Knee	L.Hip	L.Knee	CoG
Garcia's Model	0,17 ±0.02	0,38 ±0.05	-0,02 ±0.01	0,17 ±0.05	0,52 ±0,01
Our Model	0,89 ±0.02	0,99 ±0.01	0,95 ±0.003	0,99 ±0.02	0,84 ±0,01

Table **4-6**: Correlation factor computed for a normal Gait using two different physical models

Figure **4-9**: $CoG_{x,y}(t)$ trajectory obtained with two different approaches. The shadowed gray represent the normal distribution pattern for the $CoG_{x,y}(t)$ trajectory, with mean μ represented by the dark green line. The red starred line represent the $CoG_{x,y}(t)$ trajectory obtained with our physical model while the blue crossed line represent a trajectory computed from a classical pendulum model.

Table 4.2.3 shows the correlation coefficient obtained with the temporal differences between both joint angle estimates and the ground truth. This measure was applied only to those gait segments associated with the heel strike since it was previously confirmed that performance in the other gait phases are comparable because both models are based in the pendulum principle to represent the single stance phase.

Differences were found to be significant during the heel strike phase (student t-test, $p < 0.05$) for the joint angle paths computed with the conventional pendulum $CoG_{x,y}(t)$, while the joint angle trajectories estimated with our physical model, obtained correlations of about 96 %. In contrast, joint angle paths computed with the conventional pendulum $CoG_{x,y}(t)$ obtained barely correlations of about 46 %, evidencing the weakness of this type of models during this gait phase.
On the other hand, since the whole problem consists in following temporal series which are highly non-linear and whose dynamics is therefore very difficult to establish, evaluation should also include a type of measure that determines a level of agreement between two trajectories. We have then measured this level using the Fréchet distance between two temporal series, i.e., the ground truth and any of the joint-angle trajectories obtained from the fusion strategy using the two models. Briefly, the Fréchet measure estimates how close two trajectories are during the temporal capture, that is to say it estimates how similar these two curves are. Two trajectories are then similar if this distance is close to zero, the smaller this distance the closer the curves are. The

Pattern	R.Hip	R.Knee	L.Hip	L.Knee	CoG
GM	0,31 ± 0,025	0,3 ± 0,023	0,37 ± 0,017	0,35 ± 0,025	0,019 ± 0,0024
PA	0,18 ± 0,012	0,14 ± 0,018	0,26 ± 0,009	0,22 ± 0,014	0,013 ± 0,0041

Table **4-7**: Fréchet distance for a normal gait using two different physical models:Garcia's Model (GM) and proposed approach (PA)

Fréchet distance between two curves is the length of the shortest path between two points that are simultaneously moving through the two curves. The Fréchet metric uses a particular direction of the two curves because the pairs of points whose distance contributes to the Fréchet distance sweep continuously along their respective curves. This makes the Fréchet distance a better measure of similarity for curves than alternatives, such as the Hausdorff distance, for arbitrary point sets. It is possible for two curves to have a small Hausdorff distance but large Fréchet distance.
Table 4.2.3 shows the Fréchet measure obtained from temporal differences between estimates and ground truth. Again, the $CoG_{x,y}(t)$ was assessed as well as the hip $\gamma(t)$ and knee $\omega(t)$ joint-angles, showing smaller differences with our model. Interestingly, the close curve similarity between the joint angle trajectories estimated from our $CoG_{x,y}(t)$ and the ground truth, achieved a gain of 20% with respect to the classical compass pendulum model, that is to say, joint angle trajectories computed from our physical model were about 20 % more accurate.

Finally, a third part of the evaluation focused on challenging the fusion strategy to simulate pathological patterns. A Chrouch gait was tracked by changing the value of the k constant. Figure 5 shows a typical cycle obtained with the proposed model when tracking this pathological movement. It is observed in this illustration a close relationship between the trajectory described by the proposed model and the pathological pattern. A useful clinical evaluation requires a precise track of the consecutive ups and downs described by this pattern, rather than the magnitude changes.
The Crouch gait simulation was also compared with the two similarity metrics previously introduced, i.e., the correlation factor and the Fréchet distance, as illustrated in Table 4.2.3. The correlation factor is larger for upper joints, as expected since movement is much smaller, but yet correlation is high with joints such as the knees.

The proposed approach was also used to simulate the Parkinsonian gait in different stages of the disease. For each Parkinson stage it was computed the most probable k

Metric	R.Hip	R.Knee	L.Hip	L.Knee	CoG
CF	0,96 ± 0,01	0,88 ± 0,04	0,96 ± 0,01	0,93 ± 0,03	0.92 ± 0,02
FD	0,19 ± 0,006	0,26 ± 0,001	0,13 ± 0,004	0,21 ± 0,002	0,29 ± 0,003

Table **4-8**: Correlation factor (CF) and Fréchet distance (FD) when simulating the Crouch gait with the proposed fusion strategy

Model Parameter	Stage 2	Stage 3	Stage 4
k average	52.3 ± 6.22	67.42 ± 8.25	96.8 ± 2.54
d average	0.76105 ± 0.087	0.5576 ± 0.0474	0.4356 ± 0.0294

Table **4-9**: Model parameters learned from 10 patients in different stages of Parkinson disease

and d parameters (See table **4-9**). Then, using these parameters it was generated a $CoG_{x,y}(t)$ with the proposed physical model (Figure 4.2.3).

Figure **4-10**: Simulations for different gait cycles. The shadowed gray zone corresponds to the normal distribution pattern for the joint angle trajectories. The dark green line correspond to the mean joint angle patterns, i.e., the ground truth. The vertical shadowed green zone is the heel strike phase. Notice that the proposed model (red starred line) tracks better the ground truth, above all on the zones defined by the heel strike which are the most important when assessing pathological patterns. The blue crossed line corresponds to the trajectories computed from the fusion strategy but using the $CoG_{x,y}(t)$ of a classical compass pendulum.

The rule of fusion, with the computed $CoG_{x,y}(t)$ and the learned heel paths from groups of patients, was used to compute the hip $\gamma(t)$ and knee $\omega(t)$ joint-angle trajectories. The set of these trajectories allows to animate an articulated model as a virtual representation that shows characteristic Parkinsonian signs such as the short step and the slight flexion of the knee and the hip, as illustrated in figure 4.2.3. Each Parkinson disease stage was also compared with the two similarity metrics, previously introduced, i.e., the correlation factor and the Fréchet distance, and results are summarized in

2nd	Metric	R.Hip	R.Knee	L.Hip	L.Knee	CoG
	CF	0.92± 0.008	0.91± 0.01	0.95± 0.008	0.93± 0.009	0.83± 0.004
	FD	0.09 ± 0.001	0.11 ± 0.002	0.09 ± 0.001	0.12 ± 0.002	0.12 ± 0.001
3rd	Metric	R.Hip	R.Knee	L.Hip	L.Knee	CoG
	CF	0.95± 0.012	0.89± 0.009	0.93± 0.012	0.96± 0.01	0.88 ± 0.003
	FD	0.09 ± 0.001	0.13 ± 0.002	0.09 ± 0.001	0.08 ± 0.002	0.13 ± 0.016
4th	Metric	R.Hip	R.Knee	L.Hip	L.Knee	CoG
	CF	0.93± 0.008	0.90± 0.008	0.96± 0.008	0.91± 0.008	0.86 ± 0.005
	FD	0.09 ± 0.002	0.12 ± 0.001	0.17 ± 0.002	0.13 ± 0.001	0.04 ± 0.001

Table **4-10**: Results obtained simulating the Parkinsonian gait for in each stage of the disease considered in this work using Correlation factor (CF) and Fréchet distance (FD)

Table 4.2.3. These results demonstrate that our fusion strategy simulates accurately different Parkinsonian gaits stages. The angular joint variations are very similar w.r.t to real patterns, obtaining a CF higher than 0.90 while the CoG achieves an accuracy larger than 0.80. The performance for every Parkinson level was similar, showing a compact representation for the gait with our method.

Figure **4-11**: Simulation of Parkinsonian gait patterns. In the left panel it is illustrated a sequence of human poses obtained with the structural model, corresponding with actual patient poses. In the right panel it is presented the joint angle trajectory (red starred line) obtain with the fusion strategy and compared with a ground truth patterns (shadowed gray zone with mean trajectory represented by the green line).

4.2.4 Discussion

The gait can be thought of as a sequence of complex combinations of several subsystems that help the body to keep the balance while it gains support and propulsion [285].

The gait analysis aims to interpret the complex combination of several motion patterns generated by the interaction of different systems. In the clinical routine, the gait examination is considered as the most important tool for identifying motion disorders, and it is also used as a biomarker of some neuromuscular illnesses like the cerebral palsy or the Parkinson disease, supporting thereby diagnosis and follow up. Interestingly, during this analysis it is possible to evaluate the effectiveness of a specific treatment and the particular response of the multiple gait subsystems. However, these analyses are not actually carried out in the clinical practice, among others because this requires a complete correlation of all pattern recorded, which is very time consuming and examiner dependent. Moreover, measures are contaminated by noise during the capture or, by invasive devices like markers, which inevitably alter the natural motion gestures. Overall, most clinical examinations are devoted to capture a general gait picture which makes that treatments are globally addressed. The actual utility of devising gait models is that they may improve understanding of some subsystem patterns from the captured data, even if they are noisy, and hence they allow to plan more specific treatments.

This work has presented a new fusion scheme that simulates a large set of gait patterns, including pathological conditions. The model allows to identify the role of certain subsystems during the simulation, an important step towards planning oriented treatments. This scheme uses two important information sources, i.e., kinetic and kinematic components, maintaining the possibility of an energy consumption analysis. The fusion method assumes that the gait kinematic patterns must follow basic physical principles. The underlying trajectories are generated by an adapted version of the compass pendulum representation which has been extended with elements that store energy, describing a larger number of abnormalities than it has been possible so far. These trajectories are then regularized by actual learned paths, completing thereby the rule of fusion by means of an inverse kinematic approximation. A broad range of gait pathologies can thus be emulated with the proposed approach, mainly those characterized by flexion-extension restrictions or leg stiffness, as for instance patients with diabetes mellitus (DM) and peripheral neuropathy (PN) who commonly present very short stride lengths, slower walking velocities and unstable upright postures [40,41]. Likewise, these simulations may be extended to gait pathologies that compromise knee dorsiflexors and extensors like the steppage gait.

The models reported in the literature are very limited when describing particular variations of a pathological motion. Gait models can be coarsely divided in two large groups: physical based models and musculoskeletal representations. Physical models use an inverted pendulum or a spring mass system that allow an energy consumption analysis and a global dynamic description. These models nevertheless fail to mimic pathological patterns because of the strong simplifications and restrictions, i.e., pendular models

represent only the single stance while the double stance is completely omitted. Musculoskeletal representations, typically more complex than physical-based models, are able to simulate muscle patterns at each gait phase by adding some non linear muscle-tendon interactions, traditionally modeled by the Hill's model or prior information coming from data obtained from actual gait laboratories. Currently, musculoskeletal models associate each leg segment to a Hill's model. However, a main drawback of this approach is that there is no interaction between the different segments and hence simulations are quite far from experimental data, therefore, missing any anatomical meaning [71, 72]. Best performances are obtained by combining a prior model with observations coming from actual data [72, 168, 295, 296], but these models completely neglect important kinetic relationships and require a high level of expertise to properly tuning the model parameters. In contrast, physics based models are simple and usually tuned with a small number of variables. This simplicity, nonetheless, leads to most physics based models to miss important phases of the gait cycle. The fusion model proposed in this work represents the single stance by an articulated double pendulum system, while the double stance is properly simulated by a spring mass component that stands for the important knee motion contribution. The proposed model in addition emulates a whole skeleton structure by animating this architecture from the CoG trajectory. The whole strategy allows a natural simulation of non-linear gait patterns, representing several kinds of movements. Evaluation demonstrated the fusion model advantages, by comparing several kinematic patterns like the CoG trajectory and the hip and knee flexion-extension movements, considered as the most representative gait patterns for determining whether a motion is pathological or not [209, 223, 225].

The first evaluation compared the adapted physical model with a classical compass pendulum. An appropriate extraction of the CoG is essential since this biomarker is an efficient indicator of the normal/abnormal gait pattern, it constitutes one of the most important markers in pathologies such as hemiplegia, paraplegia or dystonia. A pathological gait can be analysed in terms of energy using the CoG, which tracks the transfer of potential to kinetic energy (recovery), i.e. normal gait patterns loss 40% of their energy in this transfer, a higher loss is pathological [76]. The CoG trajectory described by our physical model achieved a CF of = 0.84 while classical pendulum model only achieved a CF of = 0.52, as illustrated in Figure 3. In a second evaluation, hip and knee joint-angle paths were compared during the linear part of the cycle. Results showed a close correlation of the fusion strategy with respect to gold standard patterns, a CF of = 0.9 for our strategy and a of CF = 0.8 for trajectories computed from the classical compass pendulum. The same test was repeated exclusively for the nonlinear part of the cycle: a CF of = 0.96 was estimated with the fusion model while a CF of = 0.35 was computed with the classical compass pendulum. Likewise, while

the complete proposed strategy obtained a 0.1 Fréchet distance, the joint trajectories estimated with the classical pendulum yielded a 0.35 Fréchet distance for the normal gait. Overall, the largest error was obtained during the nonlinear parts of the gait cycle like the heelstrike and the hip moments. This fact illustrates the relevance of obtaining a complete CoG global description even to estimate the remain joint angle trajectories.

When simulations were performed for pathological patterns, the fusion method maintained in average a CF = 0.90. Two neuromuscular disorders were simulated: the cerebral palsy and the Parkinson disease in three different stages. For both abnormal movements, actual patient motion data allowed to adjust parameters and to obtain closer CoG trajectories.

During the Crouch gait simulation, the model parameters were set to $k = 400$ and $d = 0.65$, according to the data learning process. Simulations achieved a CF of = 0.9, a very close representation of actual patterns (Figure 5). The Fréchet distance maintained a comparable performance when tracking the ground truth. The Parkinson simulation, in three different disease stages, also shows a very alike representation w.r.t actual patterns. The model parameters were fixed according to Table 4 for each stage disease. The fusion model achieved in average a CF = 0.92 and a Fréchet distance average of 0.1, demonstrating the close likeness of the patterns obtained and actual data. These results demonstrated the model ability to accurately follow a different sort of gait patterns, either normal or abnormal. The model may be used as training tool for physician and also to predict the performance of a particular gait treatment.

Finally, it is worth to strengthen out the simplicity of the presented model and its ability to represent pathological movements by only tuning two parameters and with a relatively simple fusion strategy. The model may be used as a complementary tool of the classical gait analysis to determine an illness degree of any subsystem, either by correlating the compromise of any leg segment with the clinical data, or by perturbing the model and therefore the resultant trajectory.

4.2.5 Conclusions and Perspectives

This work has presented a fusion model to simulate normal and pathological kinematic gait patterns. Two main contributions are introduced in this work, a fusion strategy of two important information sources which allows the accurate estimation hip and knee joint angle trajectories. Additionally, a physical that describes the COG trajectory using a pendular motion for the single stance and a spring mass system for the double stance. The model is complemented by an animated structure that allows to visualize and quantify different gait patterns,i.e., the hip and knee flexion-extension. The pro-

posed approach can be easily extended to simulate other pathologies or even to find more dynamic gait relationships that describe a particular movement. Finally the proposed model opens up an actual possibility towards understanding more complex gait phenomena, crucial in many applications of the prostheses design such as the alignment or the relationship between those prostheses and the different muscle subsystems.

4.3 Segmentation of pelvic structures for planning CT using a geometrical shape model tuned by a multiscale edge detector

Accurate segmentation of the prostate and organs at risk in computed tomography (CT) images is a crucial step for radiotherapy planning. Manual segmentation, as performed nowadays, is a time consuming process and prone to errors due to the a high intra- and inter-expert variability. This paper introduces a new automatic method for prostate, rectum and bladder segmentation in planning CT using a geometrical shape model under a Bayesian framework. A set of prior organ shapes are first built by applying principal component analysis to a population of manually delineated CT images. Then, for a given individual, the most similar shape is obtained by mapping a set of multi-scale edge observations to the space of organs with a customized likelihood function. Finally, the selected shape is locally deformed to adjust the edges of each organ. Experiments were performed with real data from a population of 116 patients treated for prostate cancer. The data set was split in training and test groups, with 30 and 86 patients, respectively. Results show that the method produces competitive segmentations w.r.t standard methods (averaged dice = 0.91 for prostate, 0.94 for bladder, 0.89 for rectum) and outperforms the majority-vote multi-atlas approaches (using rigid registration, free-form deformation and the demons algorithm). *The complete content of this section has been published as a research article in the journal of* ***physics in medicina and biology*** *(see [202]).*

4.3.1 Introduction

Prostate cancer is one of the most commonly diagnosed male cancer worldwide, with 190.000 new cases diagnosed in USA in 2010 (American Cancer Society) and 71.000 new cases in France in 2011 (INCa 2011). External beam radiation therapy (EBRT) is a commonly prescribed treatment for prostate cancer which has proven to be efficient for tumor control [69]. EBRT uses high energy beams from multiple directions in order to deliver the dose within the patient tumor region to destroy the cancer cells.

Modern treatment techniques offer nowadays improved treatment accuracy through a better planning, delivery, visualization and the correction of patient setup errors. Segmentation plays a major role during the whole therapy. During a standard protocol for EBRT planning, CT images are manually delineated to define not only the clinical target volume (CTV), prostate and seminal vesicles, but also the neighbouring organs at risk (OARs), for instance the bladder and rectum. The CTV is then expanded to constitute the Planning Target Volume (PTV) for treatment. These spatial margins between the organs and the PTV as depicted in Figure **4-12** allow for uncertainties in delineation, patient setup, motion and organ deformations [26,120]. The segmentation allows the definition of dose constraints according to a certain recommendations. The International Commission on Radiation Units and Measurements (ICRU) 50 and 62 reports, for example, define and describe several target and critical structure volumes that aid in the treatment planning process and may provide a basis for comparison of treatment outcomes. Thus the directions, strengths, and shapes of the treatment beams are computed in a planning platform to comply with them, following a particular dose prescription. The delineated structures may also be used in further stages and with different purposes during the treatment such as the computation of cumulated dose when image guided radiotherapy (IGRT) is used or for toxicity population studies [2, 133, 256].

In current radiotherapy planning, the CT is still used to perform the segmentation since the dose computation relies on the CT electron density information. However, the CT offers a poor soft tissue contrast and therefore segmenting pelvic organs is highly time consuming (between 20-40 minutes to delineate each), and prone to errors, especially in the apical and basis regions [58]. Previous studies have reported variations in prostate delineation of about 20.60 % of the organ volume [119]. These uncertainties lead to large intra- and inter-observer variabilities and may impact the treatment planning [97] and dosimetry [19, 27]. Hence, reliable, efficient and reproducible (semi)-automatic segmentation techniques are are required in treatment planning.

Several difficulties hamper automatic segmentation methods in this context: (i) the poor border contrast between adjacent organs, (ii) the high intra- and inter-individual shape variability, (iii) the inhomogeneities in the amount of bladder and rectum filling [178] (iv) the often presence of fiducial markers used to guide dose delivery. Common automatic methods for organ segmentation of pelvic structures include deformable models [63, 66], atlas based methods [1] or statistical shape models [139, 221]. Atlas based approaches use prior learning not only for obtaining a final contour but also to provide initial organ positions for further segmentation algorithms. In atlas-based methods a pre-computed segmentation or prior information (i.e. probability map) in a template space is propagated to the image to be segmented via registra-

tion [1, 207, 241, 242]. Although atlas-based approaches may provide prior structural information, a high inter-individual variability and registration errors may lead this method to fail. Multi-atlas approaches can partly overcome some of these difficulties by selecting the most similar organ of interest among a large database, but the definition of a proper similarity metrics between the available atlases and the individual query, is still an open issue [1,1,289]. Statistical shape models may help to overcome atlas-based segmentation problems by introducing prior shapes, which in a further step may be locally adjusted [139, 221]. In general, these approaches define a cost function encoding similarities that matches a prior shape/appearance knowledge with a particular individual (test image) [96, 134]. These approaches are relatively stable and robust to noise and voxel variations, however their accuracy may be limited by the set of features selected to represent the image test and the particular definition of the match function. Examples of prostate segmentation following this approach appear in [245] where the segmentation problem was addressed without shape constraints by allowing flexibility to individual local variations, with some issues, however, in regions with poor contrast. More recently, Chen et al [56] proposed a statistical method that matches the prior shape information under a Bayesian hypothesis, incorporating anatomical constraints but without exploiting spatial information. Other approaches include learning strategies to segment the prostate, but within the context of intra-individual segmentations where previous organ delineations are required [182].

In this paper, we propose a novel method for automatically segmenting pelvic structures from CT scans based on a statistical shape model within a Bayesian framework and with local spatial variations. The proposed approach is appropiate for different types of organ, shapes, as well as stable to variations in contrast and bladder and rectum filling conditions. A method for removing CT artifacts was implemented which increases contours reliability using an adapted non-local filter, which was able to overcome the issues with fiducial markers. Since the prior is built from different available shapes, the approach is robust to the large inter-individual variability. The likelihood matching includes a geometrical and topological analysis which allows to set an optimal correspondence between a prior 3D shape model and the considered CT. Such prior is built from an actual population while a multi-scale edge descriptor allows for the mapping between the prior and the CT observations. The yielded 3D organ segmentation is spatially coherent and voxels are connected together. In the remainder of this paper we first describe the method, we then validate the accuracy of our approach on real CT data, and compared with baseline Atlas-Based segmentation methods.

(a)(b)(c)(d)

Figure **4-12**: Example of planning dose on manually segmented CT. a) original CT scan, b) prostate and rectum delineations, c) PTV is obtained by adding margins to the prostate and c) Resulting 3D planning dose. The presence of fiducial markers may hamper segmentation.

4.3.2 Methodology

Figure **4-13**: Proposed method for 3D segmentation. First, a shape space organs is built (PCA). The template is rigidly registered to the CT to be segmented followed by an automatic extraction of RoIs for preprocessing and multi-scale edge detection. A likelihood function matches the most similar PCA shape with the detected edges to finally being locally adjusted.

The proposed method consists of several stages as depicted in Figure **4-13**. Let $\widehat{S}_o$, the estimated organ shape (prostate, bladder or rectum) that optimizes a maximum a posteriori (MAP), under a Bayesian framework

$$\widehat{S}_o = \max_{\widehat{S}_o} \arg[P(\widehat{S}_o \; S^{pca}_{o_1}, S^{pca}_{o_2}, \ldots, S^{pca}_{o_N})] \tag{4-25}$$

where $S^{pca}_{o_1}, S^{pca}_{o_2}, \ldots, S^{pca}_{o_N}$ is a collection of shapes (the shape space) of the organ o. The reference template serves as the coordinate framework of the PCA geometrical shape model and corresponds simply to a random seletion of a single CT-image. The likelihood function (section 4.3.2.4), aimed to match the most similar shape with the observations extracts information from a multi-scale analysis in a RoI around the organs (section 4.3.2.3), after removing CT artifacts. Eventually, the most likely shape $\widehat{S}_o$ is locally deformed, driven by the multiscale edge descriptor (section 4.3.2.5). These steps are described in the following sections.

4.3.2.1 Learning an organ shape model: the prior

A dimensionality reduction of the whole training data was firstly carried out by applying PCA [255] to the population of manually delineated organs [188]. Hence, a collection of shapes $S^{pca}_{o_1}, S^{pca}_{o_2}, \ldots, S^{pca}_{o_N}$, for each considered organ, was computed. A

previous rigid registration scales different volumes so that the number of slices is always the same for any organ. Then, a set of "360 landmarks" was set in the polar space (a landmark per grade), using the centroid of every organ as reference. Each shape contour is the parametric curve defined by a set of landmarks, among the training data lying on the contour. In consequence, correspondences are always one-to-one and the number of points of any contour is always the same. Thus, the first two contour moments are then computed, namely the mean shape vector $\bar{s}$ R^{3M} and the covariance matrix ρ $R^{3M\times 3M}$, are computed as $\bar{s} = \frac{1}{N}\sum_{i=1}^{n} s_i$ and $\rho = \frac{1}{N-1}\sum_{i=1}^{n}(s_i - \bar{s})(s_i - \bar{s})^T$, where the vector $(s_i - \bar{s})$ characterizes the organ deviation with respect to the mean shape. A conventional spectral analysis allows diagonalization of this covariance matrix, thereby finding the directional gains or *eigenmodes*. Each eigenmode defines a 3D vector field of the correlated organ inter patient-variability. Thus, the organ statistical samples were generated by deforming the mean shape with a weighted linear combination of the L dominating eigenmodes:

$$S_{o_l}^{pca} = \bar{s} + \sum_{l=1}^{L} c_l q_l \tag{4-26}$$

where the coefficients c_l follow a Gaussian distribution and the q_l are the eigenvectors or directions with a variance defined in the interval q_l $+3\ \overline{\lambda}, \ldots, -3\ \overline{\lambda}$, accounting for the 96% of the shape variability. This procedure was independently used for each organ, obtaining a family of shape models that were aligned with the previously chosen CT template. Examples of the different shapes obtained for each organ are shown in Figure **4-13**.

4.3.2.2 RoI extraction and pre-processing

During the segmentation of a given CT image, the previously computed model is rigidly registered towards the CT template, from the training database, using a classical "block matching" method [217]. Thus, a set of RoIs of size $\bar{S}_o \pm \xi$, being $\bar{S}_o$ the average organ size and ξ a tolerance value, allows to define a particular partition for the test CT image. The computation of the organ boundaries is confined to this RoI, assuming that only two intensity classes exist, foreground (organs) and background. However, other classes may appear in the neighbourhood of the considered organ, mainly becasue of some acquisition artifacts, namely intensity inhomogeneities, noise and presence or not of fiducial markers. Then, we assume a two class $RoI_o(\mathbf{x})$: the organ of interest and some neighboring tissues, which are randomly distributed. Every RoI is then modelled as a mixture of Gaussians (GMMs), aiming to approximate the organ of interest and

the other contaminant tissues as a non uniform sum of Gaussians, each approaching a particular distribution of voxels. Such modelling is formulated as:

$$\psi(i) = \sum_{k=1,2,\ldots,n} w_k N(i\ \mu_k, \sigma_k^2) \tag{4-27}$$

where the two main distributions stand for the foreground tissues and any other structure (bones for instance), respectively. Such an estimation is consistent because the different types of noises were herein modeled as additive, usually approached by a mix of Gaussians. A classical Expectation Maximization algorithm is then applied to estimate the GMMs parameters. Let $RoI_o(\mathbf{x})$ the pixel tissue distributions, with probability density function (pdf) $p(\mathbf{x}\ \theta)$, where θ is an unknown vector of parameters (μ_i, σ_i) and $\mathbf{x}$ is every pixel of the RoI. Given an observation pixel $\mathbf{x}$, we aim at maximizing the likelihood function $L(\theta) = p(\mathbf{x}\ \theta)$ w.r.t. a given search space Θ. This problem has been classically approximated by numerical routines, as the well known Expectation-Maximization algorithm [32]
Once the Mixture of Gaussians is determined, the two main distributions $(\max_{k=1,2}(2\sigma_k))$ are chosen since they represent the two searched classes. Any other distribution is eliminated by setting its voxels to the closest main distribution, using for doing so an adapted non local mean filter, as follows: a voxel $\mathbf{x}$, within the RoI, may represent any noise $\mathbf{x} < \min_{\mathbf{k}}(\mu_{\mathbf{k}} - \mathbf{2}\sigma_{\mathbf{k}})$ $\mathbf{x} > \max_{\mathbf{k}}(\mu_{\mathbf{k}} + \mathbf{2}\sigma_{\mathbf{k}})$ and therefore is replaced by a weighted average of a neighbourhood with foreground/background information, satisfying a "non local property": weights depend on the pixel similarity in the image space as

$$\varrho(\mathbf{x}) = \mathbf{e}^{\frac{-\mathbf{d}(\mathbf{x})}{\mathbf{h}^2}}, \tag{4-28}$$

where $d(\mathbf{x}) = \sum_{\mathbf{i}\ \ \phi(\mathbf{o})} \left\| \mathbf{RoI_o}(\mathbf{x}) - \mathbf{N_{k1,k2}}(\mu, \sigma^{\mathbf{2}}) \right\|$, $\phi(o)$ is the neighborhood and h is a decay parameter. The underlying idea behind this filter is to replace a pixel belonging to a probable artifact by its nearest "foreground/background" model. Examples of the RoI pre-processing are depicted in Figure 3.

Figure **4-14**: In (a) Bladder RoI preprocessing before. Lower bladder region had different appareance because filling bladder. In (b) is shown the processed RoI. The proposed filter reduce the appareance difference allowing capture robust observations

4.3.2.3 Multi-scale CT edge detector: the observations

The multiscale image analysis was herein implemented using the first partial derivatives at each scale obtained by convolving each $RoI_o(x, y)$ with a variable gaussian kernel. Thus, the multi-scale edge estimation was obtained by a simple linear combination of the different gradient magnitudes at each scale, as follows:

$$S_o^{edge}(x, y; \sigma) = \sum_i RoI_o(x, y) * \frac{\partial G_{\sigma_i}}{\partial_x \partial_y} \tag{4-29}$$

where G_{σ_i} is the 2D Gaussian function with standard deviation σ_i. In this context, the Gaussian function is the unique kernel with an equivalent scale-space representation (linearity and shift-invariance in both frequency and space). A multiscale image decomposition allows to build a more stable set of observed edges [186] since true edges are coherently present at different scales. In this context, the detection of the boundaries is further pursued by applying a classical multiscale non-maximum suppression strategy, consisting in detecting the maximum gradient magnitude for each particular direction as cited in [186]. This multi-scale edge detection follows the universal principle of scale invariance and allows a robust edge description which is usually preserved through multiple scales [260].

4.3.2.4 Computing the geometrical likelihood

Within a Bayesian framework, the devised likelihood function $P(\widehat{S}_{o_j}^{pca}\ S_o^{edge})$ was defined to find the best geometrical matching between the sample shapes $\widehat{S}_{o_j}^{pca}$, obtained from the learned model, and each multi scale edge descriptor S_o^{edge} under a log-Euclidean metric [15]. For doing so, every shape, from the organ space, and the edges of each RoI, are characterized by the first seven Hu moments [298] as

$$P(S_{o_i}^{PCA}\ S_{o_i}^{edge}) = [\min_{S_{o_j}^{pca}} \left\{ \sum_{h_u (i=1...7)} m_i^{S_o^{edge}} - m_i^{S_{o_j}^{pca}} \right\} \tag{4-30}$$

where $m_i^{S_o^{edge}} = \log\ h_i^{S_o^{edge}}$, and $m_i^{S_{o_j}^{pca}} = \log\ h_i^{S_{o_j}^{pca}}$ are the computed shape features for the edges and the PCA learned shapes, respectively, and $h_i^{S^{edge}}$, $h_i^{S_j^{pca}}$ are their Hu moments. Under a log-Euclidean metric, null values are excluded, and the metrics allows to globally determine the most similar shape. This choice is also justified since the space of organs has a Lie group structure, that is to say, a similarity space in which

continuity applies and algebraic operations are smooth mappings. This type of log-Euclidean metrics is invariant to orthogonal transformations, change of coordinates and scaling, and sets to an infinity distance those matrices with null or negative eigenvalues. The likelihood function reaches a maximal probability when a learned shape closely match the observations after the multi-edge-scale descriptor.

4.3.2.5 Local Shape Deformation

In a final step, a 3D local deformation function was also introduced to eventually improve the local correspondence of the chosen organ, selected by the PCA model. Such deformation depends on two different measurements: (1) the radial distance of a particular shape point to the observed multi-edge-scale in the same slice and (2) the distance of the same particular organ point to the corresponding organ point in the upper and lower slices. These distances drive out the warping deformation of the organ either to the observed multi-edge-scale or to the upper and lower organ neighboring points. This deformation is controlled by a λ term, which represents a trade-off between the observed edge and the prior organ shape.

$$So_i(x,y) = \lambda \widehat{S}^{pca}_{o_i}(x,y) + (1-\lambda)(\min_{S_o}(\ \widehat{S}^{pca}_{o_i} - \ S^{edge}_{o_i}, \widehat{S}^{pca}_{o_{i\pm1}}\ \ ))$$

This local deformation tends to preserve a coherence between the prior shape and the observed edges, a compact representation of the shape given by the λ term and the nearest edge criterion. In this work, the best performance was obtained from a training database with $\lambda = 0.6$

4.3.2.6 Experimental Setup

A quantitative comparison was performed between the manual organ delineations (prostate, bladder, rectum) and the computed segmentation, using two different measures: a dice score (DSC) and the Hausdorff distance. The DSC is an overlapping similarity measure, defined as $DSC(A,B) = \frac{2\ A\ \ B}{A+B}$, where $\ \cdot\ $ indicates the number of voxels of the considered A (gold standard) and B volumes. On the other hand, the Hausdorff measure $H(A,B)$ computes the maximum distance between two set of points as $\max(h(A,B), h(B,A))$ and $h(A,B) = \max_{a\ \ A} \min_{b\ \ B}\ \ a-b\ \frac{2}{2}$. In this case, each set of points represents the organ surface. This measure allows to indirectly assess the segmentation compactness.

As mentioned before, the dataset was randomly split into: training (30 patients) and test (86 patients). The training set was used to build the organ shape space and the best representation was selected within a leave-one-out cross validation scheme. The

performance of the proposed method during the training step was compared with three multi-atlas vote methods[1], using a leave one out scheme. In a first multi-atlas method, the atlases were rigidly registered to the test volume. The other two methods followed two steps: (1) the atlases were rigidly registered to the whole set of volumes, using a "block matching" strategy, and the transformed volumes were ranked according to the normalized cross-correlation (NCC) [217]. (2) The delineations associated to each volume were propagated to a new test organ volume by non-rigidly propagating the whole set of organs. This non-rigid registration was carried out by either a standard free-form deformation (FFD) [246] or by using the demons algorithm [261]. Eventually, the majority vote decision rule was applied to obtain a single segmentation for each considered organ. A second evaluation of the presented method was carried out by segmenting the test data (86 patients) that were not included in the training phase, using the best prior model constructed during the training step.

4.3.3 Data

Data used in this paper include a total of 116 patients, who underwent Intensity Modulated Radiotherapy (IMRT) for localized prostate cancer between July 2006 and June 2007 in the same institution. The whole treatment (patient positioning, CT acquisition, and volume delineations) and dose constraints complied with GETUG 06 recommendations, as previously reported [28]. The size of the planning CT images in the axial plane was 512x512 pixels, with 1mm image resolution, and 2mm slice thickness. The used treatment planning system was Pinnacle V7.4 (Philips Medical System, Madison, WI). Each treatment plan used five field beams, in a step-and-shoot delivery configuration with gantry angles of 260°, 324°, 36°, 100°and 180°. The delivery was guided by means of an IGRT protocol, with cone beam CT images or two orthogonal images (kV or MV imaging devices), using gold fiducial markers in 57% of patients. In this study, only the CT and the delineated prostate, bladder and rectum, were considered. For evaluation purposes, the dataset was split into two parts : training (30 patients) and test (86 patients) datasets.

4.3.4 Evaluation and Results

Figure 2 illustrates an example of segmentation obtained with the proposed approach (red), overlaid on an expert segmentation (in green). A major advantage of the proposed method is the adaptability of the prior shapes to the patient-specific organ, with local variations. As mentioned above, the approach performance was evaluated

[1]The used database corresponds to the 30 patients mentioned before

as follows:

Figure **4-15**: Axial segmentation illustration of the pelvic structures ((a) rectum, (b) bladder and (c) prostate). The delineation obtained by the presented method (red) and the expert reference (green).

4.3.4.1 Evaluation of training data

In a first experiment, by using the 30 patient randomly selected from the 116 available cases a DSC leave-one-out cross validation exploratory analysis was performed. Since DSC estimates the percentage of overlapping area between the evaluated segmentations, the obtained score is strongly dependent on the volume of the organs to be evaluated. Figure 3 shows the results obtained when comparing the presented approach with the three atlas based methods. The graph depicts the mean DSC and standard deviation for the four different automatic methods. The average score observed for the proposed method was of 0.91 ± 0.033 for the prostate, 0.94 ± 0.028 for the bladder, and 0.89 ± 0.022 for the Rectum. Overall, in terms of the DSC, the presented approach outperforms the atlas-based method (using the Demons) for the the prostate and rectum, by 9 %, and 9.2 % ($p < 0.001$), respectively, while for the bladder a slight gain of 1.2 % was observed. Likewise, it should be noted that in all cases the standard deviation of the proposed approach is much smaller than the atlas based approaches.

Figure **4-16**: Dice scores comparison for vote vs the proposed approach

An additional comparison was performed by computing the Hausdorff Distance, which allows for the compactness of the segmentation to be compared since this measure penalizes the isolated voxels that are far from the ground truth. Table **4-11** summarizes the obtained performance for the three target organs, with smaller values for the proposed method in any case, indicating that the method coherently finds shapes compatible with the knowledge stored in the bank of shape organs. These results in addition illustrate that the presented approach systematically obtains shape segmentations with average distances of 5.98 for the prostate, 19.09 for the bladder and 7.52 for the rectum, while with the best atlas approach, the average distances were 9.33 for the prostate, 79.42 for the bladder and 61.44 for the rectum. It should be strengthen out that the Hausdorff distance evaluates not only the superposition of two shapes but also the quantity of scattered voxels that are outside of the segmented prostate: the larger the measure the higher the number of outlier voxels.

Segmentation method	**Prostate**	**Bladder**	**Rectum**
Vote(Rigid)	16.61±5.6	102.02±26	66.87±10.3
Vote(FFD)	14.27±4.2	78.63±20.1	65.22±6.1
Vote(Demons)	9.33±3.2	79.42±18.2	61.44±5.8
proposed approach	5.98±2.2	19.09±3.1	7.52±2.3

Table **4-11**: The Hausdorff distances obtained with the multi-atlas majority-vote method using rigid, FFD or a demons registration and with the proposed approach

The results, obtained with this metric, point put the importance of using structural priors that define a particular topology. The bladder segmentation, calculated with the three atlas-based approaches and compared with the gold standard under the DSC metric, seemed to be appropriate, but when using the Hausdorff distance, larger differences appeared and produced a very large error of 79.42 mm in the present study.

4.3.4.2 Evaluation of test data

Once a particular organ model was set, its generalization capacity was also evaluated. The best prior, determined by the PCA in the previous experiment with 30 patients, was used in a second test group with 86 patients, a larger data set with different shape organs, presence of artifacts or contrast changes in CT. Quantitative results, with the previously introduced metrics, are reported in table **4-12**. In general terms, the obtained results show a segmentation, with an averaged DSC of 0.86 and an averaged Hausdorff Distance of 16.19. The befitting obtained segmentations are consistent with what was observed in the training data, i.e. with a best score for the bladder. This fact can also be attributed to the overlapping measure since results depend on the organ volume. Although the obtained score errors are slightly larger for the three organ segmentations, the proposed approach properly captures the shape variability. The Hausdorff distance, on the other hand, shows that the proposed approach produces more stable segmentations, meaning that voxels belong mostly to a single shape since the measurement penalizes the isolated or outlier voxels. Unlike the atlas based approaches, this method reaches a tolerable margins of error, an important element when planning the radiation therapy. Finally, in both quantitative metrics, the obtained standard deviation was lower, evidencing coherency in the obtained segmentations.

	Prostate	**Bladder**	**Rectum**
Hausdorff Distance (mm)	9.98±3.4	25.07±4.6	13.52 ± 5.1
Dice score (DSC)	0.87 ± 0.071	0.89 ± 0.083	0.82 ± 0.062

Table **4-12**: The Dise score and the Hausdorff distances obtained the proposed approach in using the test data

4.3.5 Discussion

An automatic method for segmentation of pelvic organs in CT images was presented. Under a Bayesian framework, a geometrical likelihood function mapped a set of global observations, built from a structural multi-scale analysis [186], to a prior shape space that stored the shape organ knowledge. This prior captured the principal shape variation modes, constituting an organ space with samples that represent the 96 % of the shape variability [62, 134]. Observations were automatically captured from RoIs around each specific organ, namely, the prostate, the bladder and the rectum. Each of these RoIs is pre-processed by an adapted non-local filter which regularizes the region by considering only two principal pixel distributions: the organ and the background. This filter allows for artifact reduction coming from gas in the rectum, several filling bladder states or bone parts randomly captured within the RoI. Afterward, a set of structural observations, the Hu moments, are calculated from a multi-edge-scale feature. The obtained Hu-based descriptor is a global shape representation that is used by a geometrical likelihood function to find the most likely organ shape given the set of observations. The likelihood function rules out the null Hu moments and redistributes the organ space using a log-Euclidean metrics and thereby it matches the most similar.

In general, state-of-the-art methods attempt to obtain more accurate segmentations by integrating a prior to the CT image information, for instance the CT or MRI manual expert delineations or CT salient morphological points. The Bayesian approach herein presented may be included in statistical shape models family since the organ shape prior is mapped to the CT through a geometrical likelihood. Several statistical models, for segmenting CT registered structures, have been proposed. Among them, a semi-automatic Bayesian method that matches a set of organ templates and finds the most probable template transformation using an iterative increasing and bending algorithm may be found in [156]. Each of these templates is characterized by a medial axis relationship at different resolution scales. However, the media axis result in many cases very simple shape representation that can lead to wrong segmentations since the growing iterative algorithm can easily overflow the pelvic organ boundary. Likewise, segmentation of the prostate and rectum has been also approximated using a histogram

learning representation under a Bayesian framework [56], achieving an average volume overlapping of 0.89 and 0.71, for the prostate and rectum, respectively. These low figures can be attributed to the histogram representation since (1) it is not robust to contamination with any complex noise, (2) some bladder states and bone structures show practically the same gray level, (3) there exists a high inter patient variability regarding the bladder state and (4) it completely misses the spatial relationships.

Essentially, the proposed approach combines both a multi-scale (global) and derivative estimations (local) to obtain the real organ edges. Since organ boundaries in the CT images are usually very blurred and even experts can miss actual boundaries, the proposed approach sets a proper edge estimation using a likelihood function that uses global metrics (The Hu moment characterization) of the searched shape and then corrects for the possible edge mismatches. Another advantage of the proposed approach is the use of prior shapes by the likelihood function so that voxels belonging to any organ shape are always connected. This investigation has extended the usual prostate segmentation to the whole set of organs at risk, with very little change in the required parameters. The method has demonstrated to be robust to many types of noise and adaptable to different organs, with comparable accuracy values when segmenting the prostate, the bladder or the rectum. The approach was compared to 3 different atlas based methods which also included geometrical prior information, using the DSC and Hausdorff distance calculated between the automatic organ segmentation and a manual delineation carried out by a radiologist expert. Patients were grouped in training and test groups, and in both scenarios the proposed approach outperformed the other methods.

Finally, several pre-processing strategies have been proposed to overcome the natural high intra and inter patient organ appareance variability, for instance, Davis et al [70] cope with the bowel gas problem by defining a binary mask containing the regions with gas. Then, they apply a "deflation" transformation, based on the flow computed or induced by the binary image mask, and doing that this region tend to converge to the center of these regions. These approximation allows to reduce the bowel gas effect but without taking into account that very often this phenomenon appears as scattered in small regions. In the proposed approach, a new non-local filter which replaces the RoI artifacts was introduced. Taking advantage of the redundancy and assuming that a segmented RoI is composed of two tissues: organ and background. This facilitates the organ segmentation, above all because some organs like the bladder or rectum are composed of different objects that result in different distributions, this filter allows to homogenize the organ texture and to isolate the concept. After this filter, an organ and background pixel distribution are may be facilitated and used thereafter. Last but not least, the obtained results are highly encouraging and demonstrate the method

usefulness in clinical scenarios as a support of the final delineation that dramatically decreases the expert burden in the daily radiotherapy planning.

4.3.6 Conclusions and Perspectives

In this work we presented a new methodology to segment pelvic structures in CT scans. The Bayesian method combines a deformable prostate model, learned from examples, and a geometrical likelihood strategy that maps actual observation into the space of model organs and selects the best shape to superimpose it in the original CT image. The presented results offer that this segmentation technique may be suitable to use as a oncologist's support. This approach may be easily extended to other structures in CT images. Future works includes a more robust CT characterization by the association of edge observation and pixel distributions in order to find more likely shapes in the organ space and also develop a more reliable local shape adjustment.

5 Conclusions and Perspectives

This thesis has formulated, developed and largely validated a computational framework that characterizes complex dynamic patterns using different level of analysis. By emulating main motion perception mechanisms, the framework integrated three independent and complementary modules that process temporal information, capturing motion patterns and spatio-temporal relationships. Each module was extensively evaluated in diverse artificial vision tasks, such as classification, simulation, recognition and segmentation.

The first module temporally correlates visual information between consecutive frames, recovering shapes in motion and performing a global spatio-temporal analysis that was successfully used to classify normal and pathological gaits, to segment and measure polyps in video-endoscopy sequences and to segment the right ventricle in cine-MRI sequences.

The second module characterizes dynamic patterns from dense optical flow representations. In such analysis it was proposed a set of motion descriptors based on the statistical quantification of velocity flow orientations. These descriptors were successfully applied in video-surveillance applications to classify and recognize human action activities in uncontrolled scenarios. Optical flow representation were also used to compute kinematic patterns of the hummingbird during flight from a markerless standpoint and over cardiac sequences to classify normal and cardiac movements.

In the third module, prior models and simplified representations of the real world were explored. Integration of prior gait models with captured video observations was used to effectively track the center of mass during walking, achieving a proper correlation with ground truth trajectories. In addition, it was proposed a fusion model that simulates normal and pathological gait patterns from learned patient trajectories and physical gait models. This dynamical analysis was also extended to delineate the prostate in radiotherapy planning by warping a shape into a deformation model according to a set of observations.

The set of computational tools of this computational framework has opened new ways to motion analysis. Regarding human and animal motion analysis, the proposed strategies calculated salient dynamic patterns without altering the natural gestures of motion. This kind of analysis also characterizes particular patterns under non-controlled sce-

narios in surveillance applications. From biomedical analysis standpoint, the proposed strategies correlated physiological heart conditions with captured motion patterns, and identify and measure neoplastic polyps in typical endoscopy routine. These analysis may be useful as a support to diagnosis and can be easily integrated with common routine analysis. Finally, the dynamical analysis herein proposed was extended to segmentation task, classical treated from an appearance analysis perspective, obtaining remarkable results.

5.1 Perspectives

This thesis was developed attempting to construct a computational framework for quantitative and qualitative motion analysis, emulating main perception mechanisms. The analogy with visual systems has resulted in an efficient characterization of natural dynamic patterns, even in noisy scenarios. Yet understanding of these complex mechanisms has recently dramatically increased, many visual mechanisms are still unknown. Therefore, new motion perception areas would allow to complement the proposed computational framework and characterize dynamic patterns in terms of efficiency and robust representation. Overall, during this work several artificial problems were attacked proposing new directions. Some of them might be:

1. **Discovering hidden dynamic patterns**: Classical motion analysis attempts to capture specific motion trajectories from noisy measures. However it is well accepted that particular motion patterns come from hidden complex relationships among different subsystems and their interaction with the environment. In this work, motion patterns were captured from a markerless standpoint, achieving a better understanding of particular dynamics. Additional researches aimed to the discovery of new spatio-temporal relationships of objects in motion will permits the suitable analysis of pathological and abnormal trajectories correlated with physiological conditions or singular behaviours.

2. **Efficient motion representations**:

   A remarkable characteristic of the visual systems is their aptness to efficiently represent the world with a minimum number of primitives. A series of poorly understood mechanisms highlights the salient information, reconstructing information in very noisy and occluded scenarios. An important line of research in the future should then be directly related with emulating visual models and formulating representations that allow a robust capture of dynamic patterns and prediction of complex patterns of objects in motion.

3. **Fusion to build robust priors**:

   In real scenarios, any task is usually accomplished using two sources of information, actual observations and a sort of world models. The fusion of multimodal sources of information can lead to a better characterization of complex motions. Such multimodal analysis can be extended to diverse scenarios in which the object is followed from different point of views. For instance, the combination of kinematic, kinetic and consumption energy patterns may facilitate the diagnosis, treatment and following of pathological gait patterns. Also the heart function can be better understoog by combining different medical sequence modalities such as tagged images, cine-MRI sequences and echo-cardiograms.

4. **Development of tools to quantify and analyze motion in real scenarios**

   Overall, the framework proposed in this work have solved real problems in diverse scenarios. In other bio-medical applications, the proposed analysis can support the diagnosis and following up of several pathologies of the motor system. Such strategies may be used in an on-line fashion to quantitatively capture regions in motion. In other contexts, motion analysis strategies can be directly applied to remotely monitoring human activities useful in surveillance and telemedicine/healthcare applications.

5. **Performance validation in other scenarios and large datasets**

   Motion analysis is a very relevant topic in many areas since the dynamic pattern characterization results fundamental in problems such as tracking, recognition or classification, among others. Although a wide spectrum of applications was herein covered, many other scenarios would take advantage of the experience herein developed to complement classical approaches or to solve problems from a dynamical analysis perspective. Since the herein approached problems present a large variability, a extended validation with other datasets also may be useful to reach significant statistical conclusions and to capture additional patterns.

Bibliography

[1] Acosta, et a. O.: Evaluation of multi-atlas-based segmentation of ct scans in prostate cancer radiotherapy. In: *IEEE ISBI.* (2011), S. 1966 –1969

[2] Acosta, Oscar ; Drean, Gael ; Ospina, Juan D. ; Simon, Antoine ; Haigron, Pascal ; Lafond, Caroline ; Crevoisier, Renaud de: Voxel-based population analysis for correlating local dose and rectal toxicity in prostate cancer radiotherapy. In: *Physics in Medicine and Biology.* 58 (2013), Nr. 8, S. 2581

[3] Adrian, Ronald J.: Particle-Imaging Techniques for Experimental Fluid Mechanics. In: *Annual review of fluid mechanics* 23(1) (1991), S. 261–304

[4] Aggarwal, J.K. ; Ryoo, M.S.: Human activity analysis: A review. In: *ACM Computing Surveys* 43 (2011), April, S. 16:1–16:43

[5] Aggarwal, J.K. ; Ryoo, M.S.: Human Activity Analysis: A Review. In: *ACM Comput. Surv.* 43 (2011), April, Nr. 3, 16:1–16:43. `http://dx.doi.org/10.1145/1922649.1922653`. – DOI 10.1145/1922649.1922653. – ISSN 0360–0300

[6] Ahmad, Mohiuddin ; Parvin, Irine ; Lee, Seong-Whan: Silhouette History and Energy Image Information for Human Movement Recognition. In: *Journal of Multimedia* 5 (2010), S. 12–20

[7] AL., Frazier ; GA., Colditz ; CS., Fuchs ; KM., Kuntz: Cost-effectiveness of screening for colorectal cancer in the general population. In: *The Journal of the American Medical Association* 284(15) (2000), S. 1954–1961

[8] Albright, T D. ; Stoner, G R.: Visual motion perception. In: *Proceedings of the National Academy of Sciences* 92 (1995), Nr. 7, 2433-2440. `http://dx.doi.org/10.1073/pnas.92.7.2433`. – DOI 10.1073/pnas.92.7.2433

[9] Alonso Larraga, Juan ; Cossio, Sergio S.o ; Guerrero, Angelica ; Cordova Pluma, Victor ; Casillas, Jesus: The Borrmann classification. Interobserver and intraobserver agreement of endoscopists in an oncological hospital. In: *Clinical and Translational Oncology* 5 (2003), S. 345–350. – ISSN 1699–048X

[10] Altshuler, D. ; Dudley., R.: The ecological and evolutionary interface of hummingbird flight physiology. In: *J. Exp. Biol.* 205 (2005), S. 2325–2336

[11] Altshuler, Douglas L. ; Dudley, Robert: The ecological and evolutionary interface of hummingbird flight physiology. In: *Journal of Experimental Biology.* 205(16) (2002), S. 2325–2336

[12] Altshuler, Douglas L. ; Dudley, Robert ; Ellington, Charles P.: Aerodynamic forces of revolving hummingbird wings and wing models. In: *Journal of zoology* 264(4) (2004), S. 327–332

[13] Altshuler, Douglas L. ; Welch, Kenneth C. ; Cho, Brian H. ; Welch, Danny B. ; Lin, Amy F. ; Dickson, William B. ; Dickinson, Michael H.: Neuromuscular control of wingbeat kinematics in Anna's hummingbirds (Calypte anna). In: *The Journal of Experimental Biology* 213 (2010), S. 2507–2514

[14] Alvarez-Alvarez, Alberto ; Trivino, Gracian ; Cordon, Oscar: Human Gait Modeling Using a Genetic Fuzzy Finite State Machine. In: *IEEE Trans Fuzzy Systems* 20 (2012), S. 205 – 223

[15] Arsigny, Vincent ; Fillard, Pierre ; Pennec, Xavier ; Ayache, Nicholas: Fast and Simple Calculus on Tensors in the Log-Euclidean Framework. In: *Medical Image Computing and Computer-Assisted Intervention MICCAI*, 2005, S. 115–122

[16] Arulampalam, Sanjeev ; Maskell, Simon ; Gordon, Neil: A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. In: *IEEE Transactions on Signal Processing* 50 (2002), S. 174–188

[17] Atehortúa, Angélica ; Martínez, Fabio ; Romero, Eduardo: A Novel Right Ventricle Segmentation Approach from Local Spatio-temporal MRI Information. In: Ruiz-Shulcloper, José (Ed.) ; Baja, G. Sanniti d. (Ed.): *Progress in Pattern Recognition, Image Analysis, Computer Vision and Applications* Bd. 8259. Springer Berlin Heidelberg, 2013. – ISBN 978–3–642–41826–6, S. 206–213

[18] B, Pietraszewski ; S, Winiarski ; S., Jaroszczuk: Three-dimensional human gait pattern - reference data for normal men. In: *Acta Bioeng Biomech.* 14(3) (2012), S. 9–16

[19] B. Al-Qaisieh, et a.: Impact of prostate volume evaluation by different observers on CT-based post-implant dosimetry. In: *Radiother Oncol.* 62 (2002), Nr. 3, S. 267–273

[20] Baccouche, Moez ; Mamalet, Franck ; Wolf, Christian ; Garcia, Christophe ; Baskurt, Atilla: Sequential Deep Learning for Human Action Recognition. In: A.A. Salah, B. L. (Ed.): *2nd International Workshop on Human Behavior Understanding (HBU)*, Springer, 2011 (Lecture Notes in Computer Science), S. 29–39

[21] Bai, Wenjia ; Shi, Wenzhe ; O'Regan, D.P. ; Tong, Tong ; Wang, Haiyan ; Jamil-Copley, S. ; Peters, N.S. ; Rueckert, D.: A Probabilistic Patch-Based Label Fusion Model for Multi-Atlas Segmentation With Registration Refinement: Application to Cardiac MR Images. In: *Medical Imaging, IEEE Transactions on* 32 (2013), July, Nr. 7, S. 1302–1315. http://dx.doi.org/10.1109/TMI.2013.2256922. – DOI 10.1109/TMI.2013.2256922. – ISSN 0278–0062

[22] Baker, Richard: Gait Analysis Methods in Rehabilitation. In: *J Neuroeng Rehabil* 3 (2006), S. 4–14

[23] Ballan, L. ; Bertini, M. ; Del Bimbo, A. ; Seidenari, L. ; Serra, G.: Effective Codebooks for Human Action Categorization. In: *Proc of ICCV. International Workshop on VOEC*, 2009

[24] Baur, L.H.B.: Magnetic resonance imaging: the preferred imaging method for evaluation of the right ventricle. In: *The International Journal of Cardiovascular Imaging* 24 (2008), 699-700. http://dx.doi.org/10.1007/s10554-008-9321-5. – DOI 10.1007/s10554–008–9321–5. – ISSN 1569–5794

[25] Baur, L.H.B.: Magnetic resonance imaging: the preferred imaging method for evaluation of the right ventricle. In: *The International Journal of Cardiovascular Imaging* 24 (2008), Nr. 7, 699-700. http://dx.doi.org/10.1007/s10554-008-9321-5. – DOI 10.1007/s10554–008–9321–5. – ISSN 1569–5794

[26] Bayley A, Craig T Bristow R Chung P Gospodarowicz M Ménard C Milosevic M Warde P Catton C. Rosewall T T. Rosewall T: Clinical application of high-dose, image-guided intensity-modulated radiotherapy in high-risk prostate cancer. In: *Int J Radiat Oncol Biol Phys.* 77 (2010), Nr. 2, S. 477–83

[27] Beckendorf, V. ; Bachaud, J. M. ; Bey, P. ; Bourdin, S. ; Carrie, C. ; Chapet, O. ; Cowen, D. ; GuÃ©rif, S. ; Hay, H. M. ; Lagrange, J. L. ; Maingon, P. ; PrisÃ©, E. L. ; Pommier, P. ; Simon, J. M.: Target-volume and critical-organ delineation for conformal radiotherapy of prostate cancer: experience of French dose-escalation trials. In: *Cancer Radiother.* 6 (2002), Nr. 1, S. 78–92

[28] Beckendorf, VÃ©ronique ; Guerif, StÃ¨phane ; PrisÃ©, Elisabeth L. ; Cosset, Jean-Marc ; Bougnoux, Agnes ; Chauvet, Bruno ; Salem, Naji ; Chapet, Olivier ; Bourdain, Sylvain ; Bachaud, Jean-Marc ; Maingon, Philippe ; Hannoun-Levi, Jean-Michel ; Malissard, Luc ; Simon, Jean-Marc ; Pommier, Pascal ; Hay, Men ; Dubray, Bernard ; Lagrange, Jean-Leon ; Luporsi, Elisabeth ; Bey, Pierre: 70 Gy versus 80 Gy in localized prostate cancer: 5-year results of GETUG 06 randomized trial. In: *Int J Radiat Oncol Biol Phys.* 80 (2011), Nr. 4, S. 1056–1063

[29] Bengio, Yoshua: Learning Deep Architectures for AI. In: *Found. Trends Mach. Learn.* 2 (2009), Januar, Nr. 1, 1–127. http://dx.doi.org/10.1561/2200000006. – DOI 10.1561/2200000006. – ISSN 1935–8237

[30] Bentley, Jon L.: Multidimensional binary search trees used for associative searching. In: *Commun. ACM* 18(9) (1975), S. 509–517

[31] Bernal, J. ; Sï¿½nchez, J. ; Vilariï¿½o, F.: Towards automatic polyp detection with a polyp appearance model. In: *Pattern Recognition* 45 (2012), Nr. 9, S. 3166 – 3182. – ISSN 0031–3203. – Best Papers of Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA'2011)

[32] Bilmes, Jeff: A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models. In: *Tech. Rep.* (1998)

[33] Blickhan, R.: The Spring-mass model for running and hopping. In: *Journal of Biomechanics* 22 (1989), S. 1217 – 1227

[34] Bobick, A. ; Davis, J.: Real-time recognition of activity using temporal templates. In: *Applications of Computer Vision, 1996. WACV '96., Proceedings 3rd IEEE Workshop on* (1996), dec., S. 39 –42. http://dx.doi.org/10.1109/ACV.1996.571995. – DOI 10.1109/ACV.1996.571995

[35] Bobick, Aaron F. ; Davis, James W.: The Recognition of Human Movement Using Temporal Templates. In: *IEEE Trans. Pattern Anal. Mach. Intell.* 23 (2001), März, Nr. 3, 257–267. http://dx.doi.org/10.1109/34.910878. – DOI 10.1109/34.910878. – ISSN 0162–8828

[36] Bomphrey, Richard J.: Advances in Animal Flight Aerodynamics Through Flow Measurement. In: *Evolutionary Biology.* 39(1) (2011), S. 1–11

[37] Borges, P.V.K. ; Conci, N. ; Cavallaro, A: Video-Based Human Behavior Understanding: A Survey. In: *Circuits and Systems for Video Technology, IEEE Transactions on* 23 (2013), Nov, Nr. 11, S. 1993–2008. http://dx.doi.org/10.1109/TCSVT.2013.2270402. – DOI 10.1109/TCSVT.2013.2270402. – ISSN 1051–8215

[38] Bos, F. M. ; Lentink, D. ; Van Oudheusden, B. W. ; Bijl, H.: Influence of wing kinematics on aerodynamic performance in hovering insect flight. In: *Journal of fluid mechanics* 594(1) (2008), S. 341–368

[39] Boyd, Jeffrey E.: Synchronization of Oscillations for Machine Perception of Gaits. In: *Comput. Vis. Image Underst.* 96 (2004), Oktober, Nr. 1, 35–59. http://dx.doi.org/10.1016/j.cviu.2004.04.004. – DOI 10.1016/j.cviu.2004.04.004. – ISSN 1077–3142

[40] Buczek, Frank L. ; Cooney, Kevin M. ; Walker, Matthew R. ; Rainbow, Michael J. ; Concha, M. C. ; Sanders, James O.: Performance of an inverted pendulum model directly applied to normal human gait. In: *Clinical Biomechanics* 21 (2006), Nr. 3, S. 288 – 296. http://dx.doi.org/DOI:10.1016/j.clinbiomech.2005.10.007. – DOI DOI: 10.1016/j.clinbiomech.2005.10.007. – ISSN 0268–0033

[41] Burges, Christopher J. C.: A tutorial on support vector machines for pattern recognition. In: *Data Mining and Knowledge Discovery* 2 (1998), S. 121–167

[42] Bussey, T J. ; Saksida, L M.: Memory, perception, and the ventral visual-perirhinal-hippocampal stream: thinking outside of the boxes. In: *Hippocampus* 17 (2007), Nr. 9, S. 898–908

[43] Canny, John: A Computational Approach to Edge Detection. In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* PAMI-8 (1986), Nr. 6, S. 679–698. http://dx.doi.org/10.1109/TPAMI.1986.4767851. – DOI 10.1109/TPAMI.1986.4767851. – ISSN 0162–8828

[44] Canny, John: A Computational Approach to Edge Detection. In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* PAMI-8 (1986), Nov, Nr. 6, S. 679–698. http://dx.doi.org/10.1109/TPAMI.1986.4767851. – DOI 10.1109/TPAMI.1986.4767851. – ISSN 0162–8828

[45] Cao, T. ; Wu, X. ; Guo, J. ; Yu, S. ; Xu, Y.: Abnormal crowd motion analysis. In: *Int. Conf. on Robotics and Biomimetics*, 2009, S. 1709–1714

[46] Cao, Xiaochun ; Zhang, Hua ; Deng, Chao ; Liu, Qiguang ; Liu, Hanyu: Action Recognition Using 3D DAISY Descriptor. In: *Mach. Vision Appl.* 25 (2014), Januar, Nr. 1, 159–171. http://dx.doi.org/10.1007/s00138-013-0545-6. – DOI 10.1007/s00138–013–0545–6. – ISSN 0932–8092

[47] Carson, M.C. ; Harrington, M.E. ; Thompson, N. ; O'Connor, J.J. ; Theologis, T.N.: Kinematic analysis of a multi-segment foot model for research and clinical applications: a repeatability analysis. In: *Journal of Biomechanics* 34 (2001), S. 1299Ã¯Â¿Â½1307

[48] Caudron, J. ; Fares, J. ; Lefebvre, V. ; Vivier, V. ; Petitjean, C. ; Dacher, J.: Cardiac MRI assessment of right ventricular function in acquired heart disease: factors of variability. In: *Acad Radiol* 19 (2012), S. 991–1002

[49] Caudron, J. ; Fares, J. ; Vivier, P. ; Lefebvre, V. ; Petitjean, C. ; Dacher, J.: Diagnostic accuracy and variability of three semi-quantitative methods for assessing right ventricular systolic function from cardiac MRI in patients with acquired heart disease. In: *Eur Radiol* (2011), S. 2111–20

[50] Caudron, J. ; Fares, J. ; Vivier, P. ; Lefebvre, V. ; Petitjean, C. ; Dacher, J.: Diagnostic accuracy and variability of three semi-quantitative methods for assessing right ventricular systolic function from cardiac MRI in patients with acquired heart disease. In: *Eur Radiol* (2011), S. 2111–20

[51] Caudron, Jï¿½rï¿½me ; Fares, Jeannette ; Lefebvre, Valentin ; Vivier, Pierre-Hugues ; Petitjean, Caroline ; Dacher, Jean-Nicolas: Cardiac MRI Assessment of Right Ventricular Function in Acquired Heart Disease: Factors of Variability. In: *Academic Radiology* 19 (2012), Nr. 8, 991 - 1002. http://dx.doi.org/http://dx.doi.org/10.1016/j.acra.2012.03.022. – DOI http://dx.doi.org/10.1016/j.acra.2012.03.022. – ISSN 1076–6332

[52] Cavagna, G. A. ; Kaneko, M.: Mechanical work and efficiency in level walking and running. In: *Physiology* 268 (1977), S. 467–481

[53] Chadebecq, F. ; Tilmant, C. ; Bartoli, A.: Using the Infocus-Breakpoint to estimate the scale of neoplasia in colonoscopy. In: *Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on*, 2013. – ISSN 1945–7928, S. 354–357

[54] Chang, C. C. ; Lin, C. J.: LIBSVM: A library for support vector machines. In: *ACM Trans. on Intelligent Systems and Technology* 2 (2011), S. 21–27

[55] Chaudhry, Rizwan ; Ravich, Avinash ; Hager, Gregory ; Vidal, RenÃ©: Histograms of oriented optical flow and binet-cauchy kernels on nonlinear dynamical systems for the recognition of human actions. In: *Int. Conf. on Computer Vision & Pattern Recognition (CVPR)*, 2009

[56] Chen, Siqi ; Lovelock, D. M. ; Radke, Richard J.: Segmenting the prostate and rectum in CT imagery using anatomical constraints. In: *Medical Image Analysis* 15 (2011), S. 1–11

[57] Cocosco, C. ; Niessen, W. ; Netsch, T. ; Vonken, E. ; Lund, G. ; Stork, A. ; Viergever, M.: Automatic image-driven segmentation of the ventricles in cardiac cine MRI. In: *J Magn Reson Imaging* 28 (2008), S. 366–374

[58] Collier, Burnett S. Amin M. Bilton S. et a. D.: Assessment of consistency in contouring of normal-tissue anatomic structures. In: *Journal of Applied Clinical Medical Physics.* 4 (2003)

[59] Collins, Steven ; Ruina, Andy: A Bipedal Walking Robot with Efficient and Human-Like Gait. In: *Proceedings of the 2005 IEEE International Conference on Robotics and Automation* (2005), S. 1983 – 1988. – University of Michigan

[60] Collins, Steven ; Wisse, Martijn ; Ruina, Andy: A Three Dimensional Passive-Dynamic Walking Robot with Two Legs and Knees. In: *Int J Rob Res* 17 (2001), S. 607–615

[61] Condessa, Filipe ; Bioucas-Dias, Josï¿½: Segmentation and Detection of Colorectal Polyps Using Local Polynomial Approximation. In: Campilho, Aurï¿½lio (Ed.) ; Kamel, Mohamed (Ed.): *Image Analysis and Recognition* Bd. 7325. Springer Berlin Heidelberg, 2012. – ISBN 978–3–642–31297–7, S. 188–197

[62] Cootes, T. ; Taylor., C.: Statistical models of appearance for medical image analysis and computer vision. In: *Proc. SPIE Medical Imaging.*, 2001, S. 236–248

[63] Costa, MarÃa J. ; Delingette, HervÃ© ; Novellas, SÃ©bastien ; Ayache, Nicholas: Automatic Segmentation of Bladder and Prostate Using Coupled 3D Deformable Models. In: *MICCAI*, 2007, S. 252–260

[64] Cristani, Farenzena M. Bloisi D. M. ; Murino, V.: Background subtraction for automated multisensor surveillance: A comprehensive review. In: *EURASIP Journal on Advances in Signal Processing* (2010), S. 24:17

[65] Crowder, Stephen V.: A Simple Method for Studying Run-Length Distributions of Exponentially Weighted Moving Average Charts. In: *Technometrics* 29 (1987), Nr. 4, S. 401–407. – ISSN 00401706

[66] D. Bystrov, et a.: Simultaneous fully automatic segmentation of male pelvic risk structures. In: *Estro. European society for Radioherapy and oncology.*, 2012

[67] Dachman, Abraham H. ; Yoshida, Hiro: Virtual colonoscopy: past, present,and future. In: *Radiologic Clinics of North America* (2003), S. 337–393

[68] Dalal, N. ; Triggs, B.: Histograms of Oriented Gradients for Human Detection. In: *Int. Conf. on Computer Vision & Pattern Recognition* Bd. 2, 2005, S. 886–893

[69] D'Amico, et a.: Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer. In: *AMA* (1998)

[70] Davis BC, Rosenman J Goyal L Chang S Joshi S. Foskey M M. Foskey M: Automatic segmentation of intra-treatment CT images for adaptive radiation therapy of the prostate. In: *Med Image Comput Comput Assist Interv.* 8 (2005), S. 442–50

[71] Delp, Scott: An Interactive Graphics Based Model of the Lower Extremity to Study Orthopaedic Surgical Procedures. In: *IEEE Trans Biomed Eng* 37 (1990), S. 757

[72] Delp, Scott ; Loan, J: A graphics Based Software system to develop and Analyze models of musculoskeletal structure. In: *Comput Biol Medical* 25 (1995), S. 22–34

[73] Delp, Scott L. ; Anderson, Frank C. ; Arnold, Allison S. ; Loan, Peter ; Habib, Ayman ; John, Chand T. ; Guendelman, Eran ; Thelen, Darryl G.: OpenSim Open Source Software to Create and Analyze Dynamic Simulations of Movement. In: *IEEE Trans Biomed Eng* 54 (2007), S. 1940 – 1951

[74] Deluzio, Kevin J. ; Urs P. Wyss, Benny Z. ; Costigan, Patrick A. ; Serbie, Charles: Principal component models of knee kinematics and kinetics: Normal vs. pathological gait patterns. In: *Human Movement Science* 16 (1997), S. 201–217

[75] Dempster, Wilfrid ; Gaughran, George: Properties of body segments based on size and weight. In: *American Journal of Anatomy* 120 (1967), S. 33 – 54

[76] Detrembleur, C. ; Hecke, A. van d. ; Dierick, F.: Motion of the body centre of gravity as a summary indicator of the mechanics of human pathological gait. In: *Gait Posture* 12 (2000), S. 243–250

[77] Dice, L. R.: Measures of the Amount of Ecologic Association Between Species. In: *Ecology* 26 (1945), Juli, Nr. 3, S. 297–302

[78] Dice, Lee R.: Measures of the amount of ecologic association between species. In: *Ecology* 26 (1945), Nr. 3, S. 297–302

[79] Dice, Lee R.: Measures of the Amount of Ecologic Association Between Species. In: *Ecology* 26 (1945), July, Nr. 3, 297–302. `http://www.jstor.org/pss/1932409`

[80] Dickinson, Michael H. ; Lehmann, Fritz-Olaf ; Sane, Sanjay P.: Wing Rotation and the Aerodynamic Basis of Insect Flight. In: *Science* 284(5422) (1999), S. 1954–1960

[81] Dijkers, JJ ; Wijk, C van ; Vos, FM ; Florie, J ; Nio, YC ; Venema, HW ; Truyen, R ; Vliet, LJ. van: Segmentation and size measurement of polyps in CT colonography. In: *Medical Image Computing and Computer-Assisted Intervention - MICCAI* (2005), S. 712–719

[82] Dollar, P. ; Rabaud, V. ; Cottrell, G. ; Belongie, S.: Behavior Recognition via Sparse Spatio-temporal Features. In: *Proceedings of the 14th International Conference on Computer Communications and Networks*. Washington, DC, USA : IEEE Computer Society, 2005 (ICCCN '05). – ISBN 0–7803–9424–0, S. 65–72

[83] Dong, Huan ; Pelah, Adar ; Cameron, Jonathan ; Lasenby, Joan: The perceptual influences on gait transition of step parameters and optic flow in virtual environment locomotion simulators. (2008), S. 143–146. `http://dx.doi.org/http://doi.acm.org/10.1145/1394281.1394308`. – DOI http://doi.acm.org/10.1145/1394281.1394308. ISBN 978–1–59593–981–4

[84] Download MRI Data from RV segmentation Challenge MICCAI'12: *http://www.litislab.eu/rvsc/rv-segmentation-challenge-in-cardiac-mri/download*. 2012

[85] Duff-Raffaele ; Kerrigan ; Corcoran ; Saini: The proportional work of lifting the center of mass during walking. In: *Am J Phys Med Rehabilit* 75 (1996), S. 375–379

[86] Eames, M.H.A. ; Cosgrove, A. ; Baker, R.: Comparing methods of estimating the total body centre of mass in three-dimensions in normal and pathological gaits. In: *Human Movement Science* 18 (1999), S. 637–646

[87] Efros, A. ; Berg, A. C. ; Mori, G. ; Malik, J.: Recognizing Action at a Distance. In: *Int. Conf. on Computer Vision*. Washington, DC, USA, 2003

[88] Elgammal, Ahmed ; Harwood, David ; Davis, Larry: Non-parametric model for background subtraction. (2000), S. 751–767

[89] Endo, Ken ; Herr, Hugh: Human Walking Model Predicts Joints Mechanics, Electromyografy and Mechanical Economy. In: *The 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2009

[90] Evaluation Framework Right Ventricle Segmentation: *http://www.litislab.eu/rvsc*. 2012

[91] Fabio Martínez, Christian C. ; Romero, Eduardo: Simulation of normal and pathological gaits using a fusion knowledge strategy. In: *Journal of NeuroEngineering and Rehabilitation* 10 (2013), S. 73

[92] Fabio Martínez, Francisco G. ; Romero, Eduardo: A kinematic method for computing the motion of the body centre-of-mass (CoM) during walking: a Bayesian approach. In: *Computer Methods in Biomechanics and Biomedical Engineering* 14 (2011), S. 561–572

[93] Fabio Martínez, Juan Carlos L. ; Romero, Eduardo: Pathology Classification of Gait Human Gestures. In: *The International Conference on Computer Vision Theory and Applications - Vissap*, 2011

[94] Fabio Martíneznez, Eduardo R. Juan León L. Juan León: Classification of Pathological Gait Markerless Patterns. In: *VI Seminario Internacional de Procesamiento y Analisis de Imagenes MÃ©dicas (SIPAIM).*, 2010

[95] Farraye, Francis A. ; Waye, Jerome D. ; Moscandrew, Maria ; Heeren, Timothy C. ; Odze, Robert D.: Variability in the diagnosis and management of adenoma-like and non-adenoma-like dysplasia-associated lesions or masses in inflammatory bowel disease: an Internet-based study. In: *Gastrointestinal Endoscopy* 66 (2007), Nr. 3, S. 519 – 529. – ISSN 0016–5107

[96] Feng, Qianjin ; Foskey, M. ; Tang, Songyuan ; Chen, Wufan ; Shen, Dinggang: Segmenting CT prostate images using population and patient-specific

statistics for radiotherapy deformable models. In: *IEEE ISBI.* (2009), S. 282–285

[97] Fiorino, C. ; Reni, M. ; Bolognesi, A. ; Cattaneo, G.M. ; Calandrino, R.: Intra- and inter-observer variability in contouring prostate and seminal vesicles: implications for conformal treatment planning. In: *Radiother Oncol.* 47 (1998), Nr. 3, S. 285–292

[98] Florack, B. ; Romeny, L. Ter H. ; Viergever, M. ; Koenderink, J.: The gaussian scale-space paradigm and the multiscale local jet. In: *International Journal of Computer Vision* 18 (1996), S. 61–75

[99] Fox, Dieter ; Hightower, Jeffrey ; Liao, Lin ; Schulz, Dirk ; Borriello, Gaetano: Bayesian Filtering for Location Estimation. In: *IEEE Pervasive Computing* 02 (2003), Nr. 3, S. 24–33

[100] Fox, V. ; Hightower, J. ; Liao, Lin ; Schulz, D. ; Borriello, G.: Bayesian filtering for location estimation. In: *Pervasive Computing, IEEE* 2 (2003), July-Sept., Nr. 3, S. 24–33. http://dx.doi.org/10.1109/MPRV.2003.1228524. – DOI 10.1109/MPRV.2003.1228524. – ISSN 1536–1268

[101] Frank, Buczek ; Kevin, Cooney ; Walker, Matthew ; Rainbow, Michael: Performance of an inverted pendulum model directly applied to normal human gait. In: *Clin Biomech* 21 (2006), S. 288 – 296

[102] Fregly, Benjamin J.: Design of Optimal Treatments for Neuromusculoskeletal Disorders using Patient-Specific Multibody Dynamic Models. In: *Int J Comput Vision and Biomech* Jul 1;2(2): (2009), S. 145–155

[103] Full, R ; Abbas, J: Musculoskeletal dynamics in rhythmic systems. In: *Biomechanics & Neural Control of Posture & Movement* Springer Vaerlag-New York, Inc (2000), S. 192–205

[104] Gace, J R. ; Press, NY: M. (Ed.): *Gait Analysis in Cerebral Palsy.* NY: MacKeith Press, 2001

[105] Gace, J R. ; Press, Mac K. (Ed.): *The treatment of gait problems in cerebral palsy.* Mac Keith Press, 2004

[106] Gage, James R. ; Bax, Martin C. (Ed.) ; Hart, Hilary M. (Ed.): *The Treatment of Gait Problems in Cerebral Palsy.* Mac Keith Press, 2004

[107] Ganz, M. ; Yang, Xiaoyun ; Slabaugh, G.: Automatic Segmentation of Polyps in Colonoscopic Narrow-Band Imaging Data. In: *Biomedical Engineering, IEEE Transactions on* 59 (2012), Aug, Nr. 8, S. 2144–2151. – ISSN 0018–9294

[108] Garcia, Mariano ; Chatterjee, Anindta ; Ruina, Andy ; Coleman, Michael: The Simplest Walking Model: Stability, Complexity, and Scaling. In: *J Biomech Eng* 120 (1998), S. 281 – 288

[109] Garcia, Mariano ; Chatterjee, Anindya ; Ruina, Andy ; Coleman, Michael: The simplest walking model: Stability, complexity, and scaling. In: *ASME Journal of Biomechanical Engineering* 120 (1998), S. 281–288

[110] Gard, Steven A. ; Miff, Steve C. ; Kuo, Arthur D.: Comparison of kinematic and kinetic methods for computing the vertical motion of the body center of mass during walking. In: *Human Movement Science* 22 (2004), Nr. 6, S. 597 – 610. http://dx.doi.org/DOI:10.1016/j.humov.2003.11.002. – DOI DOI: 10.1016/j.humov.2003.11.002. – ISSN 0167–9457

[111] Geyer, Hartmut ; Seyfarth, Andre ; Blickhan, Reinhard: Compliant leg behaviour explains basic dynamics of walking and running. In: *Proc. R. Soc. B* 273 (2006), S. 2861 – 2867

[112] Giese, Martin A. ; T., Poggio: Neural mechanisms for the recognition of biological movements and action. In: *Nature Reviews Neuroscience* 4 (2003), 01, S. 179–192. – reviewed

[113] Goetschalckx, K. ; Rademakers, F. ; Bogaert, J.: Right ventricular function by MRI. In: *Curr Opin Cardiol* (2010), S. 451–5

[114] Gopalswamy, Narasimh ; Shenoy, Vishwanath N. ; J. Markert, Umesh C. ; Peace, Nancy ; Bhutani, Manoop S. ; Barde, Christopher J.: Is in vivo measurement of size of polyps during colonoscopy accurate? In: *Gastrointestinal Endoscopy* 46 (1997), Nr. 6, S. 497 – 502

[115] Gorelick, L. ; Blank, M. ; Shechtman, E. ; Irani, M. ; Basri, R.: Actions as Space-Time Shapes. In: *IEEE Trans. on Pattern Analysis and Machine Intelligence* 29 (2007), December, Nr. 12, S. 2247–2253

[116] Goswami, Ambarish ; Thuilot, Benoit ; Espiau, Bernard: A Study of the Passive Gait of a Compass-Like Biped Robot. In: *Journal of Robotics Research* 17 (1998), S. 1282–1301

[117] Grosgeorge, Damien ; Petitjean, Caroline ; Ruan, Su ; Caudron, Jerome ; Dacher, Jean-Nicolas: Right ventricle segmentation by graph cut shape prior. In: *Workshop in medical image computing and computer assisted intervention*, 2012

[118] Grossman, P.: Depth from Defocus. In: *Journal Pattern Recognition Letters* 5(1) (1987), S. 63–69

[119] Gual-Arnau X, Lliso F RoldÃ¡n S. IbÃ¡nez Gual G. IbÃ¡nez Gual: Organ contouring for prostate cancer: Interobserver and internal organ motion variability. In: *Computerized Medical Imaging and Graphics.* 29 (2005), S. 639–647

[120] Guckenberger M, Flentje M.: Intensity-modulated radiotherapy (IMRT) of localized prostate cancer: a review and future perspectives. In: *Strahlenther Onkol.* 183 (2007), Nr. 2, S. 57 –62

[121] Guda, Nalini M. ; Partington, Susan ; Vakil, Nimish: Inter- and intra-observer variability in the measurement of length at endoscopy: implications for the measurement of Barrett's esophagus. In: *Gastrointestinal Endoscopy* 59 (2004), Nr. 6, S. 655 – 658. – ISSN 0016–5107

[122] Gutierrez, Elena M. ; Bartonek Ã¯Â¿Â½sa ; Haglund-Ã¯Â¿Â½kerlind, Yvonne ; Saraste, Helena: Centre of mass motion during gait in persons with myelomeningocele. In: *Gait & Posture* 18 (2003), Nr. 2, S. 37 – 46. – ISSN 0966–6362

[123] Guttman, M.A. ; Prince, J.L. ; McVeigh, E.R.: Tag and contour detection in tagged mr images of the left ventricle. In: *Medical Imaging, IEEE Trans on* 13 (1994), S. 74–88

[124] Haddad, F. ; Doyle, R. ; Murphy, D. ; Hunt, S.: Right ventricular function in cardiovascular disease, part II: pathophysiology, clinical importance, and management of right ventricular failure. In: *Circulation* 117 (2008), S. 1717–1731

[125] Haddad, F. ; S. Hunt, D. R. ; Murphy, D.: Right ventricular function in cardiovascular disease, part I: Anatomy, physiology, aging, and functional assessment of the right ventricle. In: *Circulation* 117 (2008), S. 1436–1448

[126] Haiyan Luo, S. C. ; al., D. W. (.: A remote markerless human gait tracking for e-healthcare based on content-aware wireless multimedia communications. In: *IEEE Wireless Communications,.* (2010)

[127] Hateren, J. H. ; Ruderman, D. L.: Independent component analysis of natural image sequences yields spatio-temporal filters similar to simple cells in primary visual cortex. In: *Proceedings of the Royal Society of London. Series B: Biological Sciences* 265 (1998), Nr. 1412, S. 2315–2320

[128] Hausdorff, Jeffrey M.: Gait variability: methods, modeling and meaning. In: *Journal of NeuroEngineering and Rehabilitation* 2 (2005), S. 19

[129] Hebart, Martin N. ; Hesselmann, Guido: What Visual Information Is Processed in the Human Dorsal Stream? In: *The Journal of Neuroscience* 32 (2012), Juni, Nr. 24, 8107–8109. `http://dx.doi.org/10.1523/jneurosci.1462-12.2012`. – DOI 10.1523/jneurosci.1462–12.2012

[130] Hedrick, T. L. ; Biewener, A. A.: Low speed maneuvering flight of the rose-breasted cockatoo (eolophus roseicapillus) i. kinematic and neuromuscular control of turning. In: *The Journal of Experimental Biology.* 210 (2007), S. 1897–1911

[131] Hedrick, Tobalske B. W. T. L. L. T. L. ; Biewener, A. A. (.: Estimates of circulation and gait change based on a three-dimensional kinematic analysis of flight in cockatiels (Nymphicus hollandicus) and ringed turtle-doves (Streptopelia risoria). In: *J. Exp. Biol* 205 (2002), S. 1389–1409

[132] Hedrick, Usherwood J. R. T. L. L. T. L. ; Biewener, A. A.: Low speed maneuvering flight of the rose-breasted cockatoo (eolophus roseicapillus). ii. inertial and aerodynamic reorientation. In: *The Journal of Experimental Biology.* 210 (2007), S. 1912–1924

[133] Heemsbergen, et a. W.D.: Urinary obstruction in prostate cancer patients from the dutch trial (68 gy vs. 78 gy): Relationships with local dose, acute effects, and baseline characteristics. In: *Int J Radiat Oncol Biol Phys* (2010)

[134] Heimann, Meinzer H. T.: Statistical shape models for 3d medical image segmentation: A review. In: *Medical Image Analysis* 13 (2009), S. 543 – 563

[135] Hennemuth, A. ; Seeger, A. ; Friman, O. ; Miller, S. ; Klumpp, B. ; Oeltze, S. ; Peitgen, H. O.: A comprehensive approach to the analysis of contrast enhanced cardiac mr images. In: *Medical Imaging, IEEE Trans on* 27 (2008), S. 1592–1610

[136] Herbert Bay, T. T. Andreas Ess E. Andreas Ess ; Gool, L. V.: Speeded-up robust features (surf). In: *Comput. Vis. Image Underst* 110 (2008), S. 346–359

[137] Hernani, R ; Romero, G ; Martinez, M.L.: A musculoskeletal human gait model using the Bond Graph technique. In: *Proceedings of the 6th World Congress of Biomechanics (WCB 2010)* 31 (2010), S. 270–273

[138] Higgins, Longuet H. C. ; Prazdny, K.: The Interpretation of a Moving Retinal Image. In: *Proceedings of the Royal Society of London. Series B, Biological Sciences* 208(1173) (1980), S. 385–397

[139] Hodge, Fenster A. Downey D.B. L. A.C.: Prostate boundary segmentation from ultrasound images using 2D active shape models: Optimisation and extension to 3D. In: *Comput. Methods Programs Biomed.* 84 (2006), Nr. (2-3), S. 99–113

[140] Hong, DongHo ; Tavanapong, W. ; Wong, J. ; Oh, JungHwan ; Groen, P.: 3D Reconstruction of Colon Segments from Colonoscopy Images. In: *Bioinformatics and BioEngineering, 2009. BIBE '09. Ninth IEEE International Conference on*, 2009, S. 53 –60

[141] Howe, N. R. ; Deschamps, A.: Better Foreground Segmentation Through Graph Cuts. In: *ArXiv Computer Science e-prints* (2004), Januar

[142] Hoy, Melissa ; Zajac, Felix ; Gordon, Michael: A musculoskeletal model of the human lower extremity: The efect of muscle, tendon, and moment arm on the moment-angle relationship of musculotendon actuators at the hip, knee, and ankle. In: *J Biomech* 23 (1990), S. 157–169

[143] Hsu, Chih-Wei ; Lin, Chih-Jen: A comparison of methods for multiclass support vector machines. In: *IEEE Transactions on Neural Networks* 13 (2002), Nr. 2, S. 415–425

[144] Hsu, Chih-Wei ; Lin, Chih-Jen: A Comparison of Methods for Multiclass Support Vector Machines. In: *Trans. Neur. Netw.* 13 (2002), März, Nr. 2, 415–425. `http://dx.doi.org/10.1109/72.991427`. – DOI 10.1109/72.991427. – ISSN 1045–9227

[145] Hu, Ming-Kuei: Visual pattern recognition by moment invariants. In: *Information Theory, IRE Transactions on* 8 (1962), February, Nr. 2, S. 179–187. `http://dx.doi.org/10.1109/TIT.1962.1057692`. – DOI 10.1109/TIT.1962.1057692. – ISSN 0096–1000

[146] Hunter, J S.: The exponentially weighted moving average. In: *J. Quality Technol.* 18 (1986), S. 203 –210

[147] Hunter, J S.: The exponentially weighted moving average. In: *J. QUALITY TECHNOL.* 18 (1986), Nr. 4, S. 203–210

[148] Hupe, J. M. ; James, A. C. ; Bullier, J.: Cortical Feedback Improves Discrimination Between Figure and Background By V1,V2 and V3 Neurons. In: *Nature* 394 (1998), S. 784

[149] Huttenlocher, D.P. ; Klanderman, G.A. ; Rucklidge, W.J.: Comparing images using the Hausdorff distance. In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 15 (1993), sep, Nr. 9, S. 850 –863. http://dx.doi.org/10.1109/34.232073. – DOI 10.1109/34.232073. – ISSN 0162–8828

[150] Huttenlocher, D.P. ; Klanderman, G.A. ; Rucklidge, W.J.: Comparing images using the Hausdorff distance. In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 15 (1993), Sep, Nr. 9, S. 850–863. http://dx.doi.org/10.1109/34.232073. – DOI 10.1109/34.232073. – ISSN 0162–8828

[151] Ikizler, N. ; Cinbis, R.G. ; Duygulu, P.: Human action recognition with line and flow histograms. In: *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, 2008. – ISSN 1051–4651, S. 1–4

[152] Ikizler, Nazli ; Cinbis, Ramazan G. ; Duygulu, Pinar: Human action recognition with line and flow histograms. In: *19th International Conference on Pattern Recognition (ICPR 2008), December 8-11, 2008, Tampa, Florida, USA*, IEEE, 2008. – ISBN 978–1–4244–2175–6, S. 1–4

[153] J., Morimoto ; G., Cheng ; C.G., Atkeson ; Zeglin, G.: A simple reinforcement learning algorithm for biped walking. In: *Proceedings of the 2004 IEEE. International Conference on Robotics & Automation* 3 (2004), S. 3030 – 3035

[154] Ji, Shuiwang ; Xu, Wei ; Yang, Ming ; Yu, Kai: 3D Convolutional Neural Networks for Human Action Recognition. In: *IEEE Trans. Pattern Anal. Mach. Intell.* 35 (2013), Januar, Nr. 1, 221–231. http://dx.doi.org/10.1109/TPAMI.2012.59. – DOI 10.1109/TPAMI.2012.59. – ISSN 0162–8828

[155] Jia, Xiaohong ; Wang, Rencheng ; Zhang, Ming ; Li, Xiaobing: Influence of Prosthetic Sagittal Alignment on Trans-Tibial Amputee Gait and Compensating Pattern: A Case Study. In: *Tsinghua Science & Technology* 13 (2008), Nr. 5, S. 581 – 586. http://dx.doi.org/DOI:10.1016/S1007-0214(08)70092-0. – DOI DOI: 10.1016/S1007–0214(08)70092–0. – ISSN 1007–0214

[156] Joshi, Sarang ; Pizer, Stephen ; Fletcher, P. T. ; Yushkevich, Paul ; Thall, Andrew ; Marron., J. S.: Multiscale deformable model segmentation and statistical shape analysis using medial descriptions. In: *Transactions on medical imaging.* 21 (2012), Nr. 5, S. 538 –550

[157] K. R. D. Dawson, J. W. Dawson-Scully ; Robertson., R.: Forewing asymmetries during auditory avoidance in flying locusts. In: *The Journal of Experimental Biology.* 200(17) (1997), S. 2323–2335

[158] Kamruzzaman, J. ; Begg, R. K.: Support vector machinesand other pattern recognition approaches to the diagnosis of cerebral palsy. In: *IEEETrans. Biomed. Eng.* 53 (2006), S. 2479–2490

[159] Kaufman, Arie ; Wang, Jianning: 3D Surface Reconstruction from Endoscopic Videos. In: Linsen, Lars (Ed.) ; Hagen, Hans (Ed.) ; Hamann, Bernd (Ed.): *Visualization in Medicine and Life Sciences*. Springer Berlin Heidelberg, 2008 (Mathematics and Visualization). – ISBN 978–3–540–72630–2, S. 61–74

[160] Kay, S. M. ; Hall, Prentice (Ed.): *Fundamentals of Statistical Signal Processing: Estimation Theory. Secaucus*. NJ, USA., 1993

[161] Ke, Y. ; Sukthankar, R. ; Hebert, M.: Efficient Visual Event Detection Using Volumetric Features. In: *Int. Conf. on Computer Vision*, 2005, S. 166–173

[162] Ke, Yan ; Sukthankar, R. ; Hebert, M.: Efficient visual event detection using volumetric features. In: *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on* Bd. 1, 2005. – ISSN 1550–5499, S. 166–173 Vol. 1

[163] Kirtley, Christopher: *Clinical Gait Analysis: Theory and Practice*. Churchill Livingstone, 2005

[164] Klein, S. ; Staring, M. ; Murphy, K. ; Viergever, M.A. ; Pluim, J. P W.: elastix: A Toolbox for Intensity-Based Medical Image Registration. In: *Medical Imaging, IEEE Transactions on* 29 (2010), Jan, Nr. 1, S. 196–205. `http://dx.doi.org/10.1109/TMI.2009.2035616`. – DOI 10.1109/TMI.2009.2035616. – ISSN 0278–0062

[165] Klein, Stefan ; Pluim, JosienP.W. ; Staring, Marius ; Viergever, MaxA.: Adaptive Stochastic Gradient Descent Optimisation for Image Registration. In: *International Journal of Computer Vision* 81 (2009), Nr. 3, 227-239. `http://dx.doi.org/10.1007/s11263-008-0168-y`. – DOI 10.1007/s11263–008–0168–y. – ISSN 0920–5691

[166] Klempous, Ryszard: Biometric Motion Identification Based on Motion Capture. 243 (2009), S. 335–348

[167] Komura, T. ; Nagano, A. ; Leung, H. ; Shinagawa, Y.: Simulating pathological gait using the enhanced linear inverted pendulum model. In: *Biomedical Engineering, IEEE Transactions on* 52 (2005), Sept., Nr. 9, S. 1502–1513. http://dx.doi.org/10.1109/TBME.2005.851530. – DOI 10.1109/TBME.2005.851530. – ISSN 0018–9294

[168] Komura, Taku ; Nagano, Akinori ; Kudoh, Shunsuke ; Shinagawa, Yoshihisa: Simulating Pathological Gait using the Enhanced Inverted Pendulum Model. In: *IEEE Transactions on Biomech Eng* 52 (2005), S. 1502 – 1513

[169] Koppula, Hema S. ; Gupta, Rudhir ; Saxena, Ashutosh: Learning Human Activities and Object Affordances from RGB-D Videos. In: *Int. J. Rob. Res.* 32 (2013), Juli, Nr. 8, 951–970. http://dx.doi.org/10.1177/0278364913478446. – DOI 10.1177/0278364913478446. – ISSN 0278–3649

[170] Kraitchman, D.L. ; Young, A.A. ; Chang, Cheng-Ning ; Axel, L.: Semiautomatic tracking of myocardial motion in mr tagged images. In: *Medical Imaging, IEEE Trans on* 14 (1995), S. 422–433

[171] Kratz, L. ; Nishino, K.: Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models. In: *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on* 0 (2009), S. 1446–1453. http://dx.doi.org/http://doi.ieeecomputersociety.org/10.1109/CVPRW.2009.5206771. – DOI http://doi.ieeecomputersociety.org/10.1109/CVPRW.2009.5206771. ISBN 978–1–4244–3992–8

[172] Kudo, Shin ei ; Tamura, Satoru ; Nakajima, Takashi ; Yamano, Hiro o ; Kusaka, Hisashi ; Watanabe, Hidenobu: Diagnosis of colorectal tumorous lesions by magnifying endoscopy. In: *Gastrointestinal Endoscopy* 44 (1996), Nr. 1, S. 8 – 14. – ISSN 0016–5107

[173] Kuo, Arthur: A Simple Model of Bipedal Walking Predicts the Preferred Speed Step Length Relationship. In: *J Biomech Eng* 123 (2001), S. 264–269

[174] Kuo, Arthur: The six determinants of gait and the inverted pendulum analogy : A dynamic walking perspective. In: *Human movement science* 26 (2007), S. 617–656

[175] Kuo, Arthur D.: Energetics of Actively Powered Locomotion Using the Simplest Walking Model. In: *J Biomech Eng* 124 (2002), S. 113–120

[176] Kuo, Arthur D. ; Donelan, J. M.: Dynamic Principles of Gait and Their Clinical Implications. In: *Physical Therapy* 90(2) (2010), S. 157–176

[177] L. Florack, M. V. B. Ter Haar Romeny R. B. Ter Haar Romeny ; Koenderink, J.: The Gaussian scale-space paradigm and the multiscale local jet. In: *International Journal of Computer Vision.* 18(1) (1996), S. 61–75

[178] Langen, K.M ; Jones, D.T.L: Organ motion and its management. In: *International Journal of Radiation Oncology-Biology-Physics.* 50 (2001), S. 265–278

[179] Lappe, Markus ; Bremmer, Frank ; Berg, A.V. van d.: Perception of self-motion from visual flow. In: *Trends in Cognitive Sciences* 3 (1999), Nr. 9, 329 - 336. http://dx.doi.org/http://dx.doi.org/10.1016/S1364-6613(99)01364-9. – DOI http://dx.doi.org/10.1016/S1364–6613(99)01364–9. – ISSN 1364–6613

[180] Laptev, Ivan: On Space-Time Interest Points. In: *Int. J. Comput. Vision* 64 (2005), September, Nr. 2-3, S. 107–123. – ISSN 0920–5691

[181] Lewis, JPl: Fast template matching. In: *Vision Interface* Bd. 95, 1995, S. 15–19

[182] Li W, Feng Q Chen W Shen D. Liao S S. Liao S: Learning image context for segmentation of the prostate in CT-guided radiotherapy. In: *Physics in Medicine and Biology* 57 (2012), S. 1283

[183] Liang, Zongxian ; Dong, Haibo: Computational Analysis of Hovering Hummingbird Flight. In: *In 48th AIAA aerospace sciences meeting including the new horizons forum and aerospace exposition*, 2010

[184] Liedlgruber, M. ; Uhl, A. ; Vecsei, A.: Statistical analysis of the impact of distortion (correction) on an automated classification of celiac disease. In: *Digital Signal Processing (DSP), 2011 17th International Conference on*, 2011. – ISSN Pending, S. 1 –6

[185] Lindeberg, T.: Feature detection with automatic scale selection. In: *Int. J. Comput. Vision* 30 (1998), S. 79–116

[186] Lindeberg, T.: Feature detection with automatic scale selection. In: *Int. J. Comput. Vision* 30 (1998), S. 79–116

[187] Liu, Jianfei ; Subramanian, KalpathiR. ; Yoo, TerryS.: A robust method to track colonoscopy videos with non-informative images. In: *International Journal of Computer Assisted Radiology and Surgery* (2013), S. 1–18. – ISSN 1861–6410

[188] Lorenz, et a. C.: Generation of Point-Based 3D Statistical Shape Models for Anatomical Objects. In: *CVIU* 77 (2000), S. 175–191

[189] Lynch, M. ; Ghita, O. ; Whelan, P.F.: Automatic segmentation of the left ventricle cavity and myocardium in mri data. In: *Computers in Biology and Medicine* 36 (2006), S. 389–407

[190] M. Joll, C. G. ; Guehring, J.: Cardiac segmentation in mr cine data using inverse consistent deformable registration. In: *ISBI*, 2010

[191] Mahapatra, D. ; Buhmann, J.M.: Automatic cardiac RV segmentation using semantic information with graph cuts. In: *Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on*, 2013. – ISSN 1945–7928, S. 1106–1109

[192] Maier, Oskar ; Jimenez-Carretero, Daniel ; Santos, Andres ; Ledesma-Carbayo, Maria: Right ventricle segmentation with 4D Region-Merging Graph Cut in MR. In: *Workshop in medical image computing and computer assisted intervention*, 2012

[193] Manal, Kurt T. ; Buchanan, Thomas S.: Biomechanics of human movement. In: *In Standard Handbook of Biomedical Engineering & Design* (2004), S. 26

[194] Manzanera, A.: Local Jet Based Similarity for NL-Means Filtering. In: *Pattern Recognition (ICPR), 2010 20th International Conference on*, 2010. – ISSN 1051–4651, S. 2668–2671

[195] Manzanera, Antoine: Local Jet Feature Space Framework for Image Processing and Representation. In: *In International Conference on Signal Image Technology and Internet Based Systems (SITIŚ11). 261-268*, 2011

[196] Manzanera, Antoine ; Richefeu, Julien C.: A new motion detection algorithm based on [Sigma]-[Delta] background estimation. In: *Pattern Recognition Letters* 28 (2007), Nr. 3, 320 - 328. `http://dx.doi.org/http://dx.doi.org/10.1016/j.patrec.2006.04.007`. – DOI http://dx.doi.org/10.1016/j.patrec.2006.04.007. – ISSN 0167–8655. – Advances in Visual information Processing Special Issue of Pattern Recognition Letters on Advances in Visual Information Processing. (ICVGIP 2004)

[197] Manzanera, Antoine ; Richefeu, Julien C.: A new motion detection algorithm based on Σ-Δ background estimation. In: *Pattern Recogn. Lett.* 28 (2007), Februar, Nr. 3, 320–328. http://dx.doi.org/10.1016/j.patrec.2006.04.007. – DOI 10.1016/j.patrec.2006.04.007. – ISSN 0167–8655

[198] Margerie, S. D. J. B. Mouret d. J. B. Mouret M. J. B. Mouret ; Meyer., J.A.: Artificial evolution of the morphology and kinematics in a flapping-wing mini uav. In: *Bioinspir. Biomim* 2 (2007), S. 65–82

[199] Martinez, F. ; Manzanera, A. ; Santa Marta, C. ; Romero, E.: Characterization of motion cardiac patterns in magnetic resonance cine. In: *Image Information Processing (ICIIP), 2011 International Conference on*, 2011, S. 1–5

[200] Martinez, Fabio ; Gomez, Francisco ; Romero, Eduardo: A kinematic Method for computing the motion of the body center-of-Mass (CoM) during walking: A bayesian Approach. In: *Comp Meth Biomech Biomed Eng* 14 (2010), S. 561 – 572

[201] MartÃnez, Fabio ; Manzanera, Antoine ; Romero, Eduardo: *Representing activities with layers of velocity statistics for multiple human action recognition in surveillance applications.* http://dx.doi.org/10.1117/12.2042588. Version: 2014

[202] MartÃnez, Fabio ; Romero, Eduardo ; DrÃ©an, GaÃ«l ; Simon, Antoine ; Haigron, Pascal ; Crevoisier, Renaud de ; Acosta, Oscar: Segmentation of pelvic structures for planning CT using a geometrical shape model tuned by a multi-scale edge detector. In: *Physics in Medicine and Biology* 59 (2014), Nr. 6, 1471. http://stacks.iop.org/0031-9155/59/i=6/a=1471

[203] McGeer, Tad: Passive Dynamic Walking. In: *J Biomech Eng* 123 (1990), S. 264 – 269

[204] McHugh, Konrad J. Saligrama V. J. ; Jodoin, P.-M.: Foreground-adaptive background subtraction. In: *Signal Processing Letters, IEEE,* 16(5) (2009), S. 390–393.

[205] Mehran, Ramin ; Oyama, Alexis ; Shah, Mubarak: Abnormal crowd behavior detection using social force model. In: *CVPR*, IEEE, 2009. – ISBN 978–1–4244–3992–8, S. 935–942

[206] Melo, R. ; Barreto, J.P. ; Falcao, G.: A new solution for camera calibration and real time image distortion correction in medical endoscopy - Initial technical

evaluation. In: *Biomedical Engineering, IEEE Transactions on* 12 (2012), march, Nr. 3, S. 634 –644. http://dx.doi.org/10.1109/TBME.2011.2177268. – DOI 10.1109/TBME.2011.2177268. – ISSN 0018–9294

[207] Micheli-Tzanakou, Evangelia ; Rohlfing, Torsten ; Brandt, Robert ; Menzel, Randolf ; Russakoff, Daniel ; Maurer, Calvin: Atlas-Based Segmentation. In: *Handbook of Biomedical Image Analysis. Springer US.* (2005), S. 435–486

[208] Mukundan, R: *Moment Functions in image analysis-Theory and applications*. World Scientific Publishing Co, 2010

[209] Murray, M ; Drought, Bernard ; Kory, Ross: Walking Patterns of Normal Men. In: *The Journal of Bone & Joint Surgery* 46 (1964), S. 335 – 360

[210] Murray, Scott O. ; Kersten, Daniel ; Olshausen, Bruno A. ; Schrater, Paul ; Woods, David L.: Shape perception reduces activity in human primary visual cortex. In: *Proceedings of the National Academy of Sciences* 99 (2002), Nr. 23, 15164-15169. http://dx.doi.org/10.1073/pnas.192579399. – DOI 10.1073/pnas.192579399

[211] Muto, T ; Bussey, HJR ; Morson, BC: The evolution of cancer of the colon and rectum. In: *Cancer* 36 (1975), Nr. 6, S. 2251–2270

[212] Orchard, J. ; Ebrahimi, M. ; Wong, A.: Efficient nonlocal-means denoising using the SVD. In: *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*, 2008. – ISSN 1522–4880, S. 1732–1735

[213] Organization, World H.: *World health report 2003*. Bd. 2. 3. Oxford-Clarendon, 2003. – 68–73 S.

[214] Organization, World H.: *Ten leading causes of death worldwide*. 2008

[215] Ortega-Jimenez VM, Dudley R. S.: Aerial shaking performance of wet Annaś hummingbirds. In: *Journal of The Royal Society Interface.* 9(70) (2012), S. 1093–1099

[216] Ostrovsky, Yuri ; Meyers, Ethan ; Ganesh, Suma ; Mathur, Umang: Visual Parsing After Recovery From Blindness. In: *Psychological Science* 20 (2009), Dezember, Nr. 12, S. 1484. http://dx.doi.org/10.1111/j.1467-9280.2009.02471.x. – DOI 10.1111/j.1467–9280.2009.02471.x. – ISSN 0956–7976

[217] Ourselin, et a. S.: Reconstructing a 3D structure from serial histological sections. In: *Image and Vision Computing* 19 (2001), S. 25 – 31

[218] Paisitkriangkrai, S. ; Shen, Chunhua ; Zhang, Jian: Fast Pedestrian Detection Using a Cascade of Boosted Covariance Features. In: *Circuits and Systems for Video Technology, IEEE Transactions on* 18 (2008), Aug, Nr. 8, S. 1140–1151. http://dx.doi.org/10.1109/TCSVT.2008.928213. – DOI 10.1109/TCSVT.2008.928213. – ISSN 1051–8215

[219] Paragios, N.: A level set approach for shape-driven segmentation and tracking of the left ventricle. In: *IEEE Trans on Medical Imaging* 22 (2003), S. 773–776

[220] Paul, Chandana ; Bellotti, Mario ; Jezernik, Saso ; Curt, Armin: Development of a human neuro-musculo-skeletal model for investigation of spinal cord injury. In: *Biological Cybernetics* 93 (2005), S. 153–170

[221] Pekar, McNutt T.R. Kaus M. V.: Automated model-based organ delineation for radiotherapy planning in prostatic regions. In: *Int. J. Radiat. Oncol. Biol. Phys.* 3 (2004), S. 973–980

[222] Pennec, Xavier ; Joshi, Sarang: Geometrical and Statistical Methods for Modelling Biological Shape Variability. In: *Proceedings of the Second International Workshop on Mathematical Foundations of Computational Anatomy* (2008), September

[223] Perry, Jacquelin ; 1556421923, ISBN: (Ed.): *Gait Analysis: Normal and Pathological Function*. SLACK Incorporated, 1992

[224] Perry, Jacquelin ; Burnfield, Judith M.: *Gait Analysis: Normal and Pathological Function*. Slack, 2010

[225] Perry, MD ; Ayyappa, Edmond ; Shan, Stewart ; Torburn, Leslie: Below knee amputee gait with dynamic elastic response prosthetic feet A pilot study. In: *J Rehabil Res Dev* 27 (1990), S. 369 – 384

[226] Petitjean, C. ; Dacher, J.: A review of segmentation methods in short axis cardiac MR images. In: *Med Image Anal* 15 (2011), S. 169–184

[227] Petitjean, Caroline ; Dacher, Jean-Nicolas: A review of segmentation methods in short axis cardiac MR images. In: *Medical Image Analysis* 15 (2011), Nr. 2, 169 - 184. http://dx.doi.org/http://dx.doi.org/10.1016/j.media.2010.12.004. – DOI http://dx.doi.org/10.1016/j.media.2010.12.004. – ISSN 1361–8415

[228] Pinzur, Michael S. ; Cox, William ; Kaiser, James ; Morris, Ted ; Patwardhan, Avinash ; Vrbos, Lori: The Effect of Prosthetic Alignment on Relative Limb Loading in Persons with Trans-tibial Amputation : A Preliminary Rep. In: *Journal of Rehabilitation Research and Development* 32 (1995), S. 373–378

[229] Poppe, R.: A survey on vision-based human action recognition. In: *Image and Vision Computing* 28 (2010), Nr. 6, S. 976 – 990. http://dx.doi.org/10.1016/j.imavis.2009.11.014. – DOI 10.1016/j.imavis.2009.11.014. – ISSN 0262–8856

[230] Punithakumar, Kumaradevan ; Ayed, Ismail B. ; Islam, Ali ; Goela, Aashish ; Ross, Ian G. ; Chong, Jaron ; Li, Shuo: Regional heart motion abnormality detection: An information theoretic approach. In: *Medical Image Analysis* 17 (2013), Nr. 3, 311 - 324. http://dx.doi.org/http://dx.doi.org/10.1016/j.media.2012.11.007. – DOI http://dx.doi.org/10.1016/j.media.2012.11.007. – ISSN 1361–8415

[231] Punithakumar, Kumaradevan ; Ayed, Ismail B. ; Islam, Ali ; Goela, Aashish ; Ross, Ian G. ; Chong, Jaron ; Li, Shuo: Regional heart motion abnormality detection: An information theoretic approach. In: *Medical Image Analysis* 17 (2013), Nr. 3, 311 - 324. http://dx.doi.org/http://dx.doi.org/10.1016/j.media.2012.11.007. – DOI http://dx.doi.org/10.1016/j.media.2012.11.007. – ISSN 1361–8415

[232] R. Fry, S. N. ; Dickinson., M. H.: The aerodynamics of free-flight maneuvers in drosophila. In: *Science* 300(5618) (2003), S. 495–498

[233] Ramsey, Scott ; Yoon, Paula ; Moonesinghe, Ramal ; Khoury, Muin: Population-based study of the prevalence of family history of cancer: Implications for cancer screening and prevention. In: *Journal of American Collague of Medical Genetics and Genomics* 8(9) (2006), S. 571–575

[234] Richefeu, Julien ; Manzanera, Antoine: A new hybrid differential filter for motion detection. In: *Computational Imaging and Vision* 32 (2006), S. 727–732

[235] Ridler, T. ; Calvard, S.: Picture Thresholding Using an Iterative Selection Method. In: *Systems, Man and Cybernetics, IEEE Transactions on* 8 (1978), Nr. 8, S. 630–632. http://dx.doi.org/10.1109/TSMC.1978.4310039. – DOI 10.1109/TSMC.1978.4310039. – ISSN 0018–9472

[236] Ridler, T ; Calvard, S: Picture Thresholding Using an Iterative Selection Method. In: *Systems, Man and Cybernetics, IEEE Transactions on* 8 (1978),

Aug, Nr. 8, S. 630–632. http://dx.doi.org/10.1109/TSMC.1978.4310039. – DOI 10.1109/TSMC.1978.4310039. – ISSN 0018–9472

[237] Riemenschneider, Hayko ; Donoser, Michael ; Bischof, Horst: Bag of Optical Flow Volumes for Image Sequence Recognition. In: *BMVC*, British Machine Vision Association, 2009

[238] Ringenberg, Jordan ; Deo, Makarand ; Devabhaktuni, Vijay ; Berenfeld, Omer ; Boyers, Pamela ; Gold, Jeffrey: Fast, accurate, and fully automatic segmentation of the right ventricle in short-axis cardiac MRI . In: *Computerized Medical Imaging and Graphics* 38 (2014), Nr. 3, 190 - 201. http://dx.doi.org/http://dx.doi.org/10.1016/j.compmedimag.2013.12.011. – DOI http://dx.doi.org/10.1016/j.compmedimag.2013.12.011. – ISSN 0895–6111

[239] Roger, V. ; Go, A. ; Lloyd-Jones, D. ; Adams, R. ; l, J. B.: Heart Disease and Stroke Statisticts - 2012 Update: a report from the American Heart Association. In: *Circulation* 125 (2012), S. 200–220. http://dx.doi.org/10.1161/CIR.0b013e31823ac046. – DOI 10.1161/CIR.0b013e31823ac046

[240] Rogers, B. ; Graham, M: Motion parallax as an independent cue for depth perception. 8 (1979), S. 125 – 134

[241] Rohlfing, et a.: Multi-classifier framework for atlas-based image segmentation. In: *Computer Vision and Pattern Recognition, IEEE Computer Society Conference*, 2004, S. 255–260

[242] Rohlfing, T ; Maurer, CR: Multi-classifier framework for atlas-based image segmentation. In: *Pattern Recognition Letters* 26 (2005), Nr. 13, S. 2070 –2079

[243] Roldan, Luis ; Hani, Albis ; Aponte, Diego ; Sabbagh, Luis ; Gil, Fabio ; Santacoloma, Mario ; Rengifo, Adriana ; Albis, Rosario ; Arango, Lazaro ; Vélez, Fausto ; Aponte, Mauricio (Ed.) ; Restrepo, Angela (Ed.): *Técnicas en Endoscopia Digestiva*. ACED - Asociación Colombiana de Endoscopia Digestiva, 2007

[244] Romeny, B.M. ; Mathematica, Springer Publishing C. (Ed.): *Front-End Vision and Multi-Scale Image Analysis: Multi-scale Computer Vision Theory and Applications*. 2009

[245] Rousson, Mikael ; Khamene, Ali ; Diallo, Mamadou ; Celi, Juan ; Sauer, Frank: Constrained Surface Evolutions for Prostate and Bladder Segmentation

in CT Images. In: *Computer Vision for Biomedical Image Applications. Springer Berlin / Heidelberg* (2005)

[246] Rueckert, et a. D.: Nonrigid registration using free-form deformations: Application to breast mr images. In: *IEEE Transactions on Medical Imaging* (1999), S. 712–721

[247] Ryoo, M. S. ; Aggarwal, J. K.: *UT-Interaction Dataset, ICPR contest on Semantic Description of Human Activities (SDHA).* http://cvrc.ece.utexas.edu/SDHA2010/Human_Interaction.html, 2010

[248] Scovanner, Paul ; Ali, Saad ; Shah, Mubarak: A 3-dimensional sift descriptor and its application to action recognition. (2007), S. 357–360

[249] Scovanner, Paul ; Ali, Saad ; Shah, Mubarak: A 3-dimensional sift descriptor and its application to action recognition. (2007), S. 357–360

[250] Shyy, W. ; Lian, Y. ; Tang, J. ; Liu, H. ; Trizila, P. ; Stanford, B. ; Bernal, L. ; Cesnik, C. ; Friedmann, P. ; Ifju, P.: Computational aerodynamics of low Reynolds number plunging, pitching and flexible wings for MAV applications. In: *Acta Mechanica Sinica* 24(4) (2008), S. 351–373

[251] Sigal, Leonid ; Black., Michael J.: Humaneva: Synchronized video and motion capture dataset for evaluation of articulated human motion. / Dept. of Computer Science, Brown University. 2006. – Forschungsbericht

[252] Simon, Sheldon: Gait Analysis, Normal and Pathological Function. In: *The Journal of Bone & Joint Surgery* 75 (1993), S. 476–b–477

[253] Sinha, Pawan ; Ostrovsky, Yuri ; Meyers, Ethan: Parsing visual scenes via dynamic cues. In: *Journal of Vision* 6 (2006), Nr. 6, 95. `http://dx.doi.org/10.1167/6.6.95`. – DOI 10.1167/6.6.95

[254] Smal, Ihor ; Niessen, Wiro ; Meijering, Erik: Particle filtering methods for motion analysis in tagged mri. In: *ISBI*, 2010

[255] Sohn, et a. M.: Modelling individual geometric variation based on dominant eigenmodes of organ deformation: implementation and evaluation. In: *Physics in Medicine and Biology* 50 (2005), S. 5893

[256] Söhn M, et a.: Dosimetric treatment course simulation based on a statistical model of deformable organ motion. In: *Phys. Med. Biol.* 57 (2012), S. 3693

[257] Straw, Andrew D. ; Branson, Kristin ; Neumann, Titus R. ; Dickinson, Michael H.: Multi-camera Realtime 3D Tracking of Multiple Flying Animals. In: *Journal of The Royal Society Interface.* 8(56) (2011), S. 395–409

[258] T. B.P. D. Warrick, D.R. ; Dickinson, M.: The aerodynamics of hummingbird flight. In: *American Institute of Aeronautics and Astronautics*, 2007

[259] Taylor, Graham W. ; Fergus, Rob ; LeCun, Yann ; Bregler, Christoph: Convolutional Learning of Spatio-temporal Features. In: *Proceedings of the 11th European Conference on Computer Vision: Part VI.* Berlin, Heidelberg : Springer-Verlag, 2010 (ECCV'10). – ISBN 3–642–15566–9, 978–3–642–15566–6, S. 140–153

[260] Ter Haar Romeny, et a.: Scale space: Its natural operators and differential invariants. In: *Colchester.: Information Processing in Medical Imaging. Springer Berlin / Heidelberg* (1991), S. 239–255

[261] Thirion, J.P.: Image matching as a diffusion process: an analogy with maxwell's demons. In: *Medical Image Analysis* 2 (1998), S. 243 – 260

[262] Thirunarayan, Mandyam A. ; Kerrigan, D. C. ; Rabuffetti, Marco ; Croce, Ugo D. ; Saini, Meera: Comparison of three methods for estimating vertical displacement of center of mass during level walking in patients. In: *Gait & Posture* 4 (1996), Nr. 4, S. 306 – 314. – ISSN 0966–6362

[263] Tobalske, Bret W.: Biomechanics of bird flight. In: *Journal of Experimental Biology.* 210(18) (2007), S. 3135–3146

[264] Trifonov, Kalin ; Hashimoto, Shuji: Active Knee-release Mechanism for Passive-dynamic Walking Machines and Walking Cycle Re search. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems* Conf Pub (2008), S. 179 – 184

[265] Tucker, Carole A. ; Ramirez, Jose ; Krebs, David E. ; Riley, Patrick O.: Centre of gravity dynamic stability in normal and vestibulopathic gait. In: *Gait and Posture* 8 (1998), S. 117–123

[266] Turaga, P. ; Chellappa, R. ; Subrahmanian, V. S. ; Udrea, O.: Machine Recognition of Human Activities: A Survey. In: *Circuits and Systems for Video Technology, IEEE Transactions on* 18 (2008), September, Nr. 11, S. 1473–1488

[267] Verri, A. ; Girosi, F. ; Torre, V.: Mathematical properties of the two-dimensional motion field: from singular points to motion parameters. In: *J. Opt. Soc. Am. A.* 6(5) (1989), S. 698–712

[268] Viola, P. ; Jones, M.: Rapid object detection using a boosted cascade of simple features. In: *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on* Bd. 1, 2001. – ISSN 1063–6919, S. I–511–I–518 vol.1

[269] Viola, P. ; Jones., M. J. .. Int. J.: Robust real-time face detection. In: *Comput. Vision* 57 (2004), S. 137–154

[270] Vishwakarma, Sarvesh ; Agrawal, Anupam: A survey on activity recognition and behavior understanding in video surveillance. In: *The Visual Computer* 29 (2013), Nr. 10, 983-1009. http://dx.doi.org/10.1007/s00371-012-0752-6. – DOI 10.1007/s00371–012–0752–6. – ISSN 0178–2789

[271] Vries, AysoH. ; Bipat, Shandra ; Dekker, Evelien ; Liedenbaum, MarjoleinH. ; Florie, Jasper ; Fockens, Paul ; Kraan, Roel ; Mathus-Vliegen, ElizabethM. ; Reitsma, JohannesB. ; Truyen, Roel ; Vos, FransM. ; Zwinderman, AeilkoH. ; Stoker, Jaap: Polyp measurement based on CT colonography and colonoscopy: variability and systematic differences. In: *European Radiology* 20 (2010), Nr. 6, S. 1404–1413

[272] Walker, Simon M. ; Thomas, Adrian L. R. ; Taylor, Graham K.: Deformable wing kinematics in free-flying hoverflies. In: *Journal of The Royal Society Interface.* 7(42) (2010), S. 131–142

[273] Wallach, Hans ; O'connell, D: The kinetic depth effect. (1953)

[274] Wan, C ; Peng, C ; Chen, H: A simple and full automatic right ventricle Segmentation method for 4-dimensional cardiac mr images. In: *Workshop in medical image computing and computer assisted intervention*, 2012

[275] Wang, Ching ; Peng, Chun ; Chen, Hsiang: A Simple and Full Automatic Right Ventricle Segmentation Method for 4-Dimensional Cardiac MR Images. In: *RV Segmentation Challenge in Cardiac MRI MICCAI'12*, 2012

[276] Wang, Heng ; Ullah, Muhammad M. ; Kliser, Alexander ; Laptev, Ivan ; Schmid, Cordelia: Evaluation of Local Spatio-temporal Features for Action Recognition. In: *BMVC*, British Machine Vision Association, 2009

[277] Wang, Liang ; Suter, D.: Recognizing Human Activities from Silhouettes: Motion Subspace and Factorial Discriminative Graphical Model. In: *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on* (2007), S. 1–8. http://dx.doi.org/10.1109/CVPR.2007.383298. – DOI 10.1109/CVPR.2007.383298

[278] Wang, Z. J. ; Birch, J. M. ; Dickinson, M. H.: Unsteady forces and flows in low Reynolds number hovering flight: two-dimensional computations vs robotic wing experiments. In: *Journal of Experimental Biology* 207(3) (2004), S. 449–460

[279] Warrick, Bundle M. D. R. R. D. R. ; P., Dial K.: Bird maneuvering flight: blurred bodies, clear heads. In: *Integrative and Comparative Biology* 42 (2002), S. 141–148

[280] Warrick, Dial K. P. D. R. R. D. R. ; Biewener, A. A.: Asymmetrical force production in the maneuvering flight of pigeons. In: *The Auk.* 115(4) (1998), S. 916–928

[281] Warrick, Tobalske B. W. D. R. R. D. R. ; Powers, D. R.: Aerodynamics of the hovering hummingbird. In: *Nature* 435(7045) (2005), S. 1094–1097

[282] Whittington, Ben ; Thelen, Darryl: A simple mass-spring model with roller feet can induce the ground reactions observed in human walking. In: *J of Biomech Eng* 131 (2009), S. 011013–1 – 011013–8

[283] Whittle, Michael: Three-dimensional motion of the center of gravity of the body during walking. In: *Human Movement Science* 16 (1998), S. 347–355

[284] Whittle, Michael W.: Clinical gait analysis: A review. In: *Human Movement Science* 15 (1996), S. 369–387

[285] Whittlesey, Saunders ; Emmerik, Richard van ; Hamill, Joseph: The Swing Phase of Human Walking Not a Passive Movement. In: *Motor control* 4 (2000), S. 273 – 292

[286] Willems, Geert ; Tuytelaars, Tinne ; Gool, Luc: An Efficient Dense and Scale-Invariant Spatio-Temporal Interest Point Detector. In: *Proceedings of the 10th European Conference on Computer Vision: Part II.* Berlin, Heidelberg : Springer-Verlag, 2008 (ECCV '08). – ISBN 978–3–540–88685–3, S. 650–663

[287] Winiarski, Sawomir ; Rutkowska-Kucharska, Alicja: Estimated ground reaction force in normal and pathological gait. In: *Acta of Bioengineering and Biomechanics* 11(1) (2009), S. 53–60

[288] Wolf, Sebastian ; Loose, Tobias ; Schablowski, Matthias ; Dï¿½derlein, Leonhard ; Rupp, Rï¿½diger ; Gerner, Hans J. ; Bretthauer, Georg ; Mikut, Ralf: Automated feature assessment in instrumented gait analysis. In: *Gait and Posture* 23 (2006), S. 331–338

[289] Wu, Minjie ; Rosano, Caterina ; Lopez-Garcia, Pilar ; Carter, Cameron S. ; Aizenstein, Howard J.: Optimum template selection for atlas-based segmentation. In: *Neuroimage.* 34 (2007), Nr. 4, S. 1612–1618

[290] Wu, P. ; Stanford, B.K. ; Sallstrom, E. ; Ukeiley, L. ; Ifju, P.G.: Computational aerodynamics of low Reynolds number plunging, pitching and flexible wings for MAV applications. In: *Bioinspiration and Biomimetics. Acta Mechanica Sinica* 24(4) (2008), S. 351–373

[291] W.Zijlstra ; A.L.Hof: Displacement of the pelvis during human walking: empirical data and model predictions. In: *Gait and Posture* 6 (1997), S. 249–262

[292] Xiang, Yujiang ; Arora, Jasbir S. ; Abdel-Malek, Karim: Physics-based modeling and simulation of human walking: a review of optimization-based and other approaches. In: *Struct Multidisc Optim* 42 (2010), S. 1–23

[293] Yamazaki, N ; Ogihara, N: Generation of human bipedal locomotion by a bio-mimetic neuro-musculo-skeletal model. In: *Bio Cybern* 84 (2001), S. 1 – 11

[294] Yao, Jianhua ; Miller, M. ; Franaszek, M. ; Summers, R.M.: Colonic polyp segmentation in CT colonography-based on fuzzy clustering and deformable models. In: *Medical Imaging, IEEE Transactions on* 23 (2004), Nov, Nr. 11, S. 1344–1352. `http://dx.doi.org/10.1109/TMI.2004.826941`. – DOI 10.1109/TMI.2004.826941. – ISSN 0278–0062

[295] Zajac, Felix ; Neptune, Richard ; Kautz, Steven: Biomechanics and muscle coordination of human walking Part I: Introduction to concepts, power transfer, dynamics and simulations. In: *Gait Posture* 16 (2002), S. 215– 232

[296] Zajac, Felix ; Neptune, Richard ; Kautz, Steven: Biomechanics and muscle coordination of human walking Part II: Lessons from dynamical simulations and clinical implications. In: *Gait Posture* 17 (2003), S. 1–17

[297] Zajac, Felix E. ; Neptune, Richard R. ; Kautz, Steven A.: Biomechanics and muscle coordination of human walking Part II: Lessons from dynamical simulations and clinical implications. In: *Gait & Posture* 17 (2003), S. 1–17

[298] Zhang, D.: Review of shape representation and description techniques. In: *Pattern Recognition* 37 (2004), Nr. 1, S. 1–19

[299] Zhuang, Xiahai ; Arridge, S. ; Hawkes, D.J. ; Ourselin, S.: A Nonrigid Registration Framework Using Spatially Encoded Mutual Information and Free-Form Deformations. In: *Medical Imaging, IEEE Transactions on* 30 (2011), Oct, Nr. 10, S. 1819–1828. http://dx.doi.org/10.1109/TMI.2011.2150240. – DOI 10.1109/TMI.2011.2150240. – ISSN 0278–0062

[300] Zuluaga, Maria ; Cardoso, Jorge ; Ourselin, Sebastien: Automatic righ ventricle segmentation using a multi-label fusion in Cardiac MRI. In: *Workshop in medical image computing and computer assisted intervention*, 2012