UNIVERSITÀ DI PADOVA  FACOLTÀ DI INGEGNERIA

Dipartimento di Ingegneria dell'Informazione

Scuola di Dottorato di Ricerca in Ingegneria dell'Informazione
Indirizzo: Bioingegneria

XXIII CICLO

# NEW FRONTIERS OF MARKERLESS MOTION CAPTURE: APPLICATION TO SWIM BIOMECHANICS AND GAIT ANALYSIS

**Direttore della Scuola**: Ch.mo Prof. Matteo Bertocco

**Supervisore**: Ch.mo Prof. Claudio Cobelli

**Dottoranda**: Elena Ceseracciu

CONTENTS

ABSTRACT ..... iii

SOMMARIO ..... v

1 INTRODUCTION ..... 1

2 MARKERLESS MOTION CAPTURE TECHNIQUE ..... 9

2.1 STATE OF THE ART ..... 9

2.2 BACKGROUND SUBTRACTION ..... 12

2.3 THREE DIMENSIONAL REPRESENTATION ..... 14

2.4 MODEL DEFINITION ..... 16

2.5 MODEL MATCHING ..... 23

3 DEVELOPMENT OF AN UNDERWATER MARKERLESS MOTION CAPTURE TECHNIQUE FOR STUDYING SWIM BIOMECHANICS ..... 29

3.1 BACKGROUND ..... 29

3.2 EXPERIMENTAL SETUP ..... 30

3.3 DATA PROCESSING ..... 39

3.4 EXPERIMENTAL RESULTS ..... 48

3.5 DISCUSSION ..... 57

4 COMPARISON OF MARKERLESS AND MARKER-BASED MOTION CAPTURE TECHNOLOGIES THROUGH SIMULTANEOUS DATA COLLECTION DURING GAIT ..... 59

4.1 BACKGROUND ..... 59

4.2 EXPERIMENTAL SETUP ..... 62

4.3 DATA PROCESSING ..... 65

4.4 EXPERIMENTAL RESULTS ..... 82

4.5 DISCUSSION ..... 88

5 CONCLUSIONS ..... 91

BIBLIOGRAPHY ..... 93

LIST OF FIGURES ..... 101

LIST OF TABLES ..... 107

# ABSTRACT

During the last decade markerless motion capture techniques have gained an increasing interest in the biomechanics community. These techniques are only based on synchronized video sequences of the subject and allow noninvasive measurements of body segments kinematics. The main advantage over commercial marker-based motion capture systems is that they do not rely on skin based markers for tracking, capturing the subject's motion in a natural and easier way. Furthermore, markerless analysis can be performed in non-controlled environments where markers cannot be employed. An exemplary case is the kinematic analysis of swimmers, which requires underwater human motion tracking, therefore it cannot rely on retro reflective marker based motion capture systems. The major drawbacks of the state of the art of swim biomechanics are the manual digitization of feature points on video sequences of the subject, which is a time consuming task, and that it has often been limited to the sagittal plane. Since the motion in front-crawl swimming occurs on different planes, for a descriptive analysis of three-dimensional motion a multi-camera setup is needed, which requires not only calibration, but also synchronization of the cameras. These procedures are hindered by the underwater experimental environment. In this context markerless systems, without requiring to draw or attach markers to the athlete skin, are suitable for underwater quantitative analysis during training or even competition. One of the aims of the present thesis is the development of a video-based, markerless system for the analysis of arm movements during front crawl swimming,

On the other hand, in the clinical field, the application of markerless techniques is still debated. This is mainly due to a limited number of papers dedicated to the comparison with the state of the art of marker based motion capture, especially in term of repeatability and accuracy in the estimation of the three dimensional joint rotations. So far in the present thesis the application of markerless techniques to

data acquired with a marker-based system has been investigated. The latter was employed to directly compare the joint kinematics estimated with the markerless technique with the one determined with traditional marker based systems, while excluding differences in calibration and synchronization of the cameras. In this context for marker placement the CAST protocol was adopted, which could be considered the most reliable in term of accuracy for joints rotation estimation. Indeed, a combined markerless and markerbased system could be considered a versatile solution for gait laboratories, which could perform different type of analyses based on the desired degree of accuracy and unencumbrance of the subject.

The present thesis is organized as follows: the first chapter contains an introduction to markerless motion capture systems and the fields in which they have been applied. The second chapter presents the state-of-the-art of markerless techniques reported in literature, and a detailed description of the actual approach that has been employed and developed. The third and fourth chapters focus respectively on the application of the markerless technique to front-crawl swimming employing underwater cameras, and to gait analysis employing video sequences acquired with a marker-based system. Some background information is given first, presenting the open issues that are addressed in this thesis. A description of the experimental setup is provided, along with the explanation of the data processing method that has been developed to address the related issues. Finally, experimental results are presented and discussed.

Conclusions and future developments are the object of the fifth chapter.

# SOMMARIO

Nell'ultimo decennio le techiche di acquisizione del movimento di tipo "markerless" hanno suscitato un interesse sempre maggiore all'interno della comunità biomeccanica. Queste tecniche si basano solo su sequenze video sincronizzate del soggetto e permettono misure non invasive della cinematica segmentale. Il vantaggio principale che presentano rispetto ai sistemi di acquisizione commerciali basati su marcatori è che non è richiesta la presenza di marcatori o altri dispositivi sulla pelle del soggetto, pertanto è possibile registrarne il movimento in maniera più naturale e pratica. Inoltre le analisi markerless possono essere effettuate anche in ambienti non controllati dove i marcatori non possano essere utilizzati. Un caso esemplare è l'analisi cinematica dei nuotatori, che richiede la ricostruzione di traiettorie in ambiente subacqueo; i sistemi che sfruttano i marcatori retro-riflettenti non possono quindi essere utilizzati. Gli svantaggi principali dei metodi allo stato dell'arte della biomeccanica del nuoto sono la digitazione manuale dei punti di interesse sulle sequenze video del soggetto, compito che richiede molto tempo, e il fatto che sono spesso limitati al piano sagittale. Dal momento che il movimento, nel nuoto a stile libero, avviene su piani diversi, per ottenere un'analisi descrittiva del movimento tridimensionale è necessario un setup che comprenda molteplici telecamere; queste devono essere non solo calibrate, ma anche sincronizzate. Le procedure per la predisposizione di tale setup sono tuttavia rese difficoltose dall'ambiente sperimentale subacqueo. In questo contesto, i sistemi di tipo markerless, che non richiedono di disegnare o fissare marcatori alla pelle dell'atleta, sono adatti per analisi quantitative subacquee durante l'allenamento o addirittura in competizione. Uno degli scopi di questa tesi è lo sviluppo di un sistema markerless, basato su video, per l'analisi dei movimenti del braccio nel nuoto a stile libero.

D'altra parte, nel campo clinico, l'applicazione di tecniche markerless è ancora dibattuta. Questo è dovuto principalmente alla scarsa presenza in letteratura di lavori dedicati al confronto tra questi sistemi e lo stato dell'arte dell'acquisizione

del movimento basata su marker, soprattutto in termini di ripetibilità e accuratezza nella stima delle rotazioni articolari tridimensionali. In questa tesi è stata studiata l'applicazione di tecniche markerless a dati acquisiti con un sistema stereofotogrammetrico sviluppato per l'utilizzo dei marcatori. Quest'ultimo è stato sfruttato per effettuare un paragone diretto tra la cinematica articolare stimata con la tecnica marrkerless e quella ricostruita tradizionalmente mediante i marcatori, escludendo eventuali differenze nella calibrazione o nella sincronizzazione dei due sistemi. In questo contesto è stato adottato, per il posizionalemento dei marcatori, un protocollo di tipo CAST, che può essere considerato il più affidabile in termini di accuratezza nella stima delle rotazioni articolari. La combinazione di un sistema markerless-marker potrebbe essere invero considerata una soluzione versatile per i laboratori di analisi del movimento, che potrebbero effettuare diversi tipi di analisi a seconda del desiderato grado di accuratezza e di disagio procurato al paziente.

La presente tesi è organizzata come segue: il primo capitolo contiene un'introduzione ai sistemi markerless e ai campi in cui sono stati applicati. Il secondo capitolo riporta lo stato dell'arte delle tecniche markerless presenti in letteratura e una descrizione dettagliata dell'approccio markerless che è stato impiegato e sviluppato. Il terzo e il quarto capitolo sono incentrati rispettivamente sull'applicazione della tecnica markerless all'analisi, mediante telecamere subacquee, del nuoto a stile libero, e all'analisi clinica del passo impiegando sequenze video acquisite con un sistema predisposto per l'utilizzo di marcatori. Sono innanzitutto fornite alcune informazioni introduttive, che presentano gli argomenti trattati in questa tesi. Vengono quindi descritti il setup sperimentale e i metodi di analisi dati che sono stati sviluppati per affrontare i relativi problemi. Infine vengono presentati e discussi i risultati sperimentali.

Il quinto capitolo è dedicato alle conclusioni e alla discussione di possibili sviluppi futuri.

# 1 INTRODUCTION

The purpose of motion capture systems is the description and measurement of the movement performed by a subject. A great variety of applications in different areas exploit these systems, ranging from the clinical to the sports to the entertainment fields. Equally wide is the range of technologies and materials employed by these systems; each one presents some advantages, but also has requirements, in terms of applicability to different environments, complexity of use, costs, measured kinematics variables and relative precision.

Generally, motion capture systems are divided into two main categories: optical and non-optical. The latter involve the placement of sensor devices on the subject. The strength of these systems is that they are cheap, highly portable and only require a simple setup, as there is no need to ensure that the sensors are in line of sight of separate acquisition devices, as in case of optical motion capture systems. The main disadvantage, however, is that the presence of the devices encumbers the subject, sometimes even restricting their movements; also, some safety issues may arise in some environments, for example for the presence of water. The main types of sensors belonging to this category are: electromagnetic, electromechanical and inertial sensors.

*Electromagnetic systems* calculate position and orientation of body segments by measuring the relative magnetic flux between sensors placed on them and a transmitter source. The transmitter generates a low-frequency electromagnetic field, either direct current (DC) or alternating current (AC). Sensors are made of three orthogonal coils, and need to be cabled to an electronic control unit that correlates their reported locations within the field, becoming therefore quite cumbersome (Figure 1-1). The electronic control units are networked with a host computer that uses a software driver to represent these positions and rotations in 3D space. Electromagnetic systems use 6–11 or more sensors per person to record body joint motion. The sensors report position and rotational information, then inverse kinematics is used to solve the angles for the various body joints, and

compensate for the fact that the sensors are offset from the actual joint's centre of rotation. These systems are very sensitive to the presence of metal or other magnetic fields in the environment: currents inducted in the metal create new electromagnetic fields (called *eddy currents*) that interfere with the originally emitted one. Finally, the sensors only work within a limited distance from the transmitter, imposing therefore constraints on the maximum acquisition volume [1].

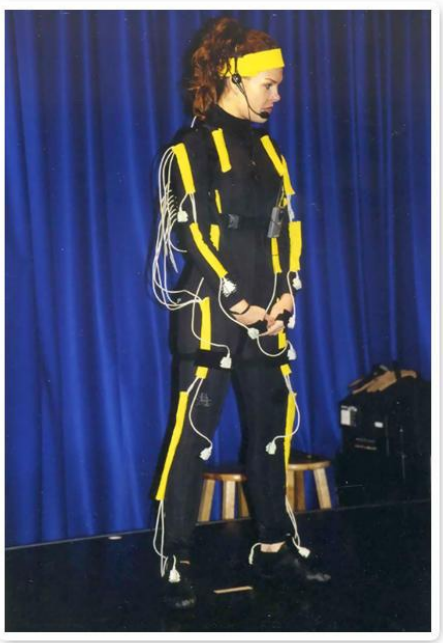


**Figure 1-1 Electromagnetic motion capture system by JZZ Technologies, Inc**

*Electromechanical systems* consist of exoskeleton-like structures attached to the subject's body, so that the relative movement of the articulated mechanical parts reflects the movement of the person. Typically, they are rigid structures of jointed, straight metal or plastic rods, linked together with potentiometers, which articulate at the joints of the body, so that every rod is coupled with a body segment (Figure 1-2). Joint angles are calculated immediately, but the position of the body in space must be determined with other types of systems. The constraints imposed at joint

level are the main disadvantage of these systems for accurate 3D kinematics estimation.

**Figure 1-2 Electromechanical system by MetaMotion (Gypsy 4)**

*Inertial systems* are based on miniature sensors, usually MEMS (micro-electro-mechanical systems), such as accelerometers and gyroscopes. Accelerometers measure linear acceleration of the device along one axis; for 3D analysis, three orthogonal accelerometers are needed. Three-axial gyroscopes measure the angular velocity of the device in the inertial reference frame. Position and orientation of the device are obtained by integration or double integration of the measured signals, employing sensor fusion techniques to combine rotational information of the device and accelerations along moving axes. The inevitable presence of noise however causes these signals to drift significantly (accumulation error), rendering them useful only when very short and quick movements are examined. The combination of accelerometric and gyroscopic sensors is called an

inertial measurement unit; each unit can be used separately, or more units can be combined in an inertial sensing suit (Figure 1-3)[2].

**Figure 1-3 Inertial motion sensors (left) and sensing suit (right) by Xsens**

Optical motion capture systems are the most commonly used, especially in clinical settings, because of the higher level of accuracy they provide. The cost of this type of systems is usually considerably greater than non-optical systems, and their application is generally limited to laboratory situations. They consist of multiple video-cameras that record from multiple views the moving subject, on whose skin particular markers have been attached (Figure 1-4).

**Figure 1-4 Subject wearing retroreflective markers**

A distinction can be made between passive and active marker systems. In the former, markers are simple plastic balls covered with a retroreflective coating, whose function is to reflect the illumination, typically at infra-red frequencies, provided by strobes mounted around each camera, which are in turn equipped with infra-red filters (Figure 1-5).

**Figure 1-5 Infrared camera and strobes (BTS S.r.l)**

In this way, markers appear on each camera image as bright dots, easily detectable, and the position of their centroids can be measured. Knowing the position and orientation of each camera in space, information that is obtained through a calibration procedure performed before the actual acquisition, it is possible to reconstruct the 3D position of the marker from its projection on each camera image (Figure 1-6). Since multiple markers are present in the scene at the same time, a great effort has been spent on algorithms that are able to perform the correct associations between 2D projections of markers, and then track their 3D trajectories throughout a sequence, even when markers are occluded in some views, or when they are close to each other. Markers' complete occlusion, markers swapping, markers going outside of the calibration volume are the main problems that affect this type of optical systems.

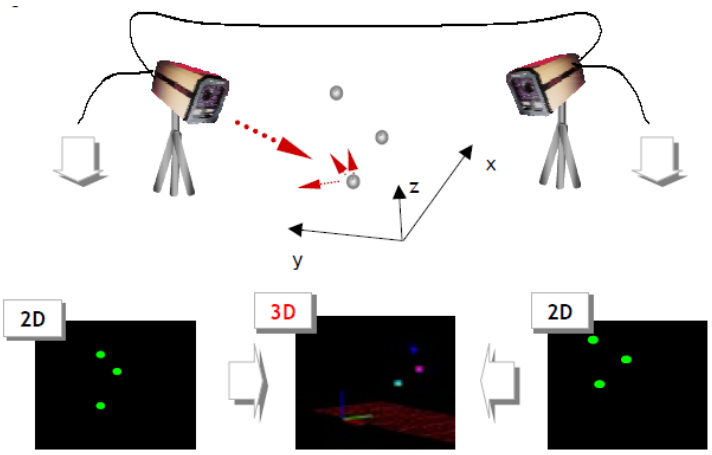



**Figure 1-6 Two-camera optoelectronic system; 3D reconstruction is performed from 2D views recorded from the cameras**

Marker identification and marker swapping issues can be addressed through the use of active marker systems. Active markers are light emitting diodes (LEDs) that are powered and cabled; their presence is therefore more obtrusive with respect to passive ones. Each LED pulses in a set sequence, so it can easily be tracked and identified; the acquisition rate is however reduced proportionally to the number of markers in the scene. In most recent years, these systems have been refined, yielding to time modulated active markers. Time modulation, instead of turning on one LED at a time, strobes them: several are turned on at the same time, but each flashes at a different frequency rate, allowing them to be individually identified. This form of unique marking radically increases the resulting capture frequency, at the expense of a far higher processing overhead.

Regardless of the type of markers employed, the output of these systems is their 3D trajectory; in order to estimate the pose of the different body segments, at least three non-collinear markers need to be attached to each segment under study, and they need to be reconstructed at all times.

A great variety of protocols has been proposed in literature for marker placement, depending on the movements and body area under examination (i.e. upper limb, lower limb, full body). Commercial acquisition software includes a set of

consolidated protocols, for which the extraction of kinematic variables is automated. Many studies have been performed in literature on comparisons and validation of these protocols ([3],[4],[5]).

In some cases, none of the abovementioned systems can be employed to perform the desired analyses, due for example to adverse environment conditions that affect the operation of the devices, such as the presence of water, or to excessive difficulties in securing sensors or markers to the subject. Traditionally, in these situations, simpler analyses are performed on video recordings of the subject that executes the movement. While a qualitative analysis from videos is very straightforward, in order to obtain quantitative measures the videos have to be processed. Employing one camera, motion on a single plane can be analyzed, by the identification and tracking of features of interest in the video; some algorithms have been developed by the computer vision community for the automatic tracking of features, but they are not always applicable to real-life situations with acceptable results. Often, the tracking process is performed manually by a human operator, who digitizes the desired points on each video image. This approach however involves long processing times, and manual intervention introduces a certain degree of subjectivity. Three-dimensional information can be obtained through the use of multiple cameras; however, they need to be not only calibrated, but also synchronized. Moreover, if manual digitization is involved, data processing becomes more time-consuming and errors are more likely to occur.

For these reasons, the development of advanced techniques, that automatically perform pose estimation of body segments from videos, has been researched in the last two decades [6],[7],[8]. A more detailed description of the state-of-the-art of this field will be given in section 2.1 of the present thesis. Some applications of these techniques have already been proposed in literature: Pansiot studied tennis actions recreating a canonical view from multiple cameras, and performing pseudo-skeletonization on the silhouette of the subject [9]; Corazza et al. analyzed the tennis serve employing the visual-hull based approach described in section 2

of this thesis to study tennis serve kinematics for the right arm [10]; Rosenhahn et al. investigated the motion of athletes interacting with sports gear, in particular cyclists and snowboarders, where the presence of the equipment results in joint restrictions [11].

In the present thesis, the application of markerless motion capture to two different branches of biomechanics was investigated. First, sports biomechanics was taken into consideration, with the analysis of arm movements during front crawl swimming. Second, clinical gait analysis was considered, and a procedure for comparison of joint angles' calculation with the two methods was developed and evaluated.

Therefore, the thesis is organized as follows: the second chapter presents the state-of-the-art of markerless techniques reported in literature, and a detailed description of the actual approach that has been employed and developed. The third and fourth chapters focus respectively on the application of the markerless technique to front-crawl swimming employing underwater cameras, and to gait analysis employing video sequences acquired with a marker-based system. Some background information is given first, presenting the open issues that are addressed in this thesis. A description of the experimental setup is provided, along with the explanation of the data processing method that has been developed to address the related issues. Finally, experimental results are presented and discussed. Conclusions and future developments are the object of the fifth chapter.

# 2 MARKERLESS MOTION CAPTURE TECHNIQUE

## *2.1 STATE OF THE ART*

The development of passive markerless motion capture systems originated from the fields of computer vision and machine learning, where the analysis of human actions by a computer has been gaining increasing interest, as demonstrated by the great number of publications and surveys [6],[7], [8],[12].

These systems vary in the number of cameras used (camera configuration), types of algorithms, representation of captured data, use of various models, and the application to specific body regions and whole body.

In this brief review we will neglect the discussion of methods for human tracking for surveillance purposes, which aim at associating humans present in a video frame with those in the previous frames, and for action recognition, whose objective is to recognize the identity of individuals and understand behaviors and activities they perform.

We will focus instead on methods for 3D human pose estimation; in particular, we are interested in multiple-view based methods. The monocular case represents a really complex and ill-posed problem, for which learnt motion models are needed as a-priori information to constrain the search over the possible different 3D poses [13].

Pose estimation algorithms are usually divided into three categories, based on how they make use of human models: model-free methods do not use *a priori* knowledge; a model can be used as a reference to constrain and guide the interpretation of measured data (indirect use of a model), or it can be a representation of the observed subject, maintained and updated by the data, that provides any desired information including the pose at any time (direct use of a model).

An example of model-free algorithm is the “probabilistic assembly of parts”, in which likely locations of body parts are first detected and then assembled to obtain the configuration which best matches the observations.

Among model-free approaches, many “example-based” methods have been proposed, which compare the observed images with a database of samples. To represent the mapping from 2D silhouette sequences (in image space) to skeletal motion in 3D pose space, hidden Markov models [14], nonlinear regression [15], direct silhouette lookups using Chamfer distance [16] have been used. Commonly, marker-based motion capture data is exploited to provide training examples with known 3D pose. A classic approach also involves a stick-figure representation of the subject obtained from the images using medial axis transformation [17] or distance transformation. Isomaps [18] and Laplacian Eigenmaps [19] are interesting instruments for transforming a 3D representation into a pose-invariant graph for extracting kinematics.

Indirect models usually include a-priori knowledge varying from height of the subject to body part shape, relative size and configuration, and dynamic information. Labeling and localization of different body parts are usually performed, on 2D images or on 3D visual hull data (for details on visual hull, see section 12.3), searching for structures similar to the model. Tracking can be achieved through the use of an extended Kalman filter [20], or through 3D-to-3D non-rigid surface matching [21].

Human models employed by *direct use* techniques are usually very detailed, and they provide the ability to handle occlusions and include kinematic constraints into the motion capture system. A typical model definition, with joints, the sticks (bones) connecting them and a surface representation, is described in section 2.3. Simple stick-figures, cylinders, super-quadrics, and CAD models are also often used. The concrete representation of the human model is a “state space” where each axis represents a degree of freedom of a joint in the model. Constraints are often used to limit the search space, either by partitioning it into legal and illegal

regions, or by introducing forces that act on an unconstrained state phase. Dynamic information, such as the assumption of a known motion pattern, can be used to reduce considerably the number of possible poses, giving up on generality.

Different approaches can be used to relate the image-data acquired to this pose-data representation; the most common one is referred to as "analysis-by-synthesis" and is based on a matching process, either on the camera planes (reprojecting the model on them and finding correspondences with cues like edges, silhouettes, blobs, texture) or directly in 3D space, for example generating a 3D representation like a visual hull.

The matching problem is generally formulated as a function optimization, solved through numeric iterative methods; gradient-descent methods or Kalman filter exploitation were common until a few years ago, but research has lately moved to more sophisticated stochastic approaches, like the simulated annealing [22] or the *particle filter* tracker [23]. Their application to the human pose problem is difficult because of the computational requirements for such a high-dimension state space. A combination of stochastic and deterministic approaches is often sought, like in the *annealed particle filter* [24], for which at each time step the particle set is refined through a series of annealing cycles with decreasing temperature to approximate the local maxima in the fitness function), or using the *stochastic meta descent* [25], which consists in a stochastic sampling of the set of model points used at each iteration of the gradient descent algorithm).

Evaluation of methods proposed in literature remains however mostly heuristic and qualitative; furthermore, it is difficult to compare performances presented by different groups because no standard error measures exist and results are reported in a variety of ways which prevent direct comparison. For these reasons, common datasets have been recently made available, such as the HumanEva [26], which aim to provide synchronized video sequences and marker-based motion capture data, along with a set of error measures and support software for manipulating the data and evaluating results.

## *2.2 BACKGROUND SUBTRACTION*

The background subtraction step is common among almost all markerless video based approaches. The simplest way to perform it in a controlled environment is by covering walls and floor of the area with panels painted in a plain colour that is not present on the subject, i.e. blue or green [27]. The portions of the video which match the preselected color are considered as background. For relatively less controlled situations, such as clinical gait laboratories where multiple instruments are used that remain in view of the cameras, a more general approach must be used: a reference background image, where no subjects are present, is taken and compared to each frame of the video sequence. When colour video cameras are employed, as commonly proposed in literature, the first step in this comparison consists in subtracting from each channel (Red, Green, Blue) of the frame under exam the corresponding channels of the background image. If the sum of the values of the three channels for one pixel in the image thus generated is below a fixed threshold (the sum is seldom zero, due to the presence of noise), that pixel will be labeled as "background". This is usually not enough for a satisfying foreground-background separation because of the presence of the soft shadows cast by the subject. To avoid considering these shadows as part of the subject itself, a second comparison is performed in the RGB space (Figure 2-1): the vectors for each pixel in the two images are computed, and if the angle between them is small enough (below a certain threshold) the pixel is still labeled as background.

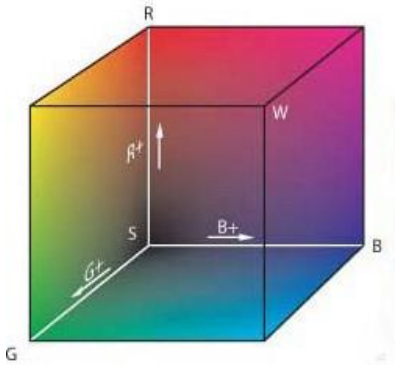



**Figure 2-1 RGB space**

In typical sports or outdoor situations, however, the background is subject to changes during the sequence, therefore the adoption of a single static reference background image may become inadequate. More sophisticated stochastic approaches, developed in the computer vision field, are needed. They generally create a model of the background, based on past history of the video sequence; an example is the creation of a background model from a running Gaussian averaging of the value of each pixel. Another common approach entails the definition of eigenbackgrounds [28], obtained from dimensional reduction of the series of frames in the video sequence through principal component analysis: the assumption is that the PCA-reduced space will represent only the static parts in the scene. Finally, one of the most common algorithms creates, then iteratively updates, a Gaussian mixture model of the RGB values assumed by each pixel in the past frames of the video sequence [29]. RGB values that are deemed to belong to the main components of the Gaussian mixture are therefore considered to belong to the background.

Regardless of the actual technique employed for background/foreground segmentation, only the information relative to the shape of the subject is retained from the images. Each frame image is binarized assigning, for example, the value 0 (black) to all the background pixels and the value 1 (white) to all the foreground

pixels. Usually morphological operations are performed, such as dilation and erosion (*binary closure*) [30], in order to get rid of spurious pixels or holes in the foreground patch. The images thus obtained are called *silhouettes* (Figure 2-2).

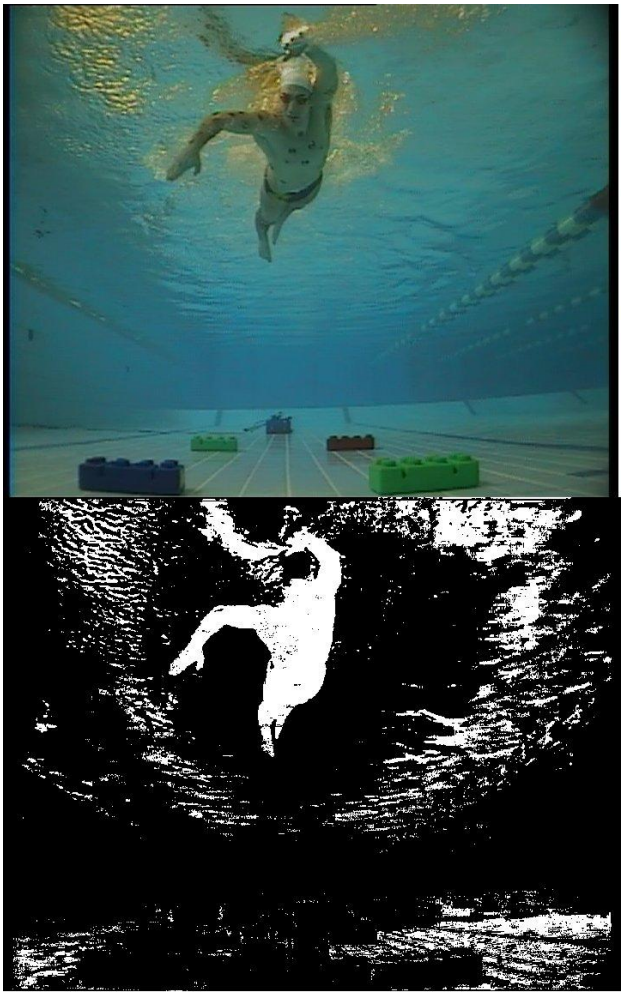


**Figure 2-2 Example of original video frame (top) and the binary silhouette image (bottom) resulting from background subtraction**

## *2.3 THREE DIMENSIONAL REPRESENTATION*

Silhouettes from all the different, synchronized views are used to compute a visual hull [31]. Since the position and orientation of each camera in space is known, a generalized cone can be back-projected into space; from the intersection

of all these cones, a 3D volume is obtained, which represents a locally convex over-approximation of the actual person's volume (Figure 2-3).

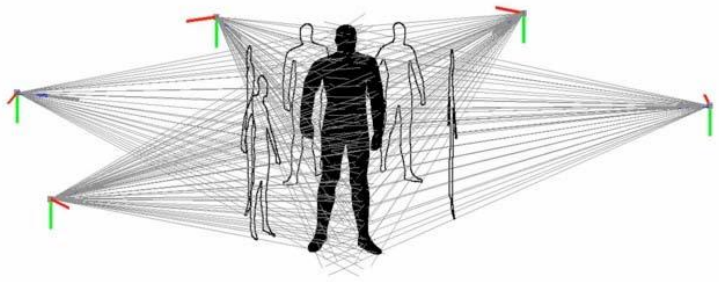


**Figure 2-3 Visual hull generation from silhouettes**

The simplest way to reconstruct the visual hull of a person is by partitioning the calibrated volume in small cubic samples, called voxels, then checking all the vertices against the input silhouettes: only the voxels whose projections on the camera planes belong to the silhouette are considered as belonging to the visual hull. The resolution of the visual hulls depends on the size of the voxels; however, this approach suffers from quantization errors, especially when the projection of a voxel on a camera plane has smaller size than a pixel. To improve the speed of the visual hull generation, a method based on octree representation of the voxels in the working volume can be employed: partitioning the space recursively subdividing it into eight octants, the only small voxels which need to be consideres are the ones near sensitive areas, such as edges. Since the registration algorithm consists in the matching of three-dimensional surfaces, visual hull volumes are filtered in order to extract 3D contours.

An alternative approach to visual hull creation represents the silhouette contour as a polygon and finds the surfaces which delimit the intersection of the cones in space by solving a series of bidimensional problems [32]. The resulting surface is called *polyhedral visual hull*.

One of the main problems from which the visual hull approach suffers is the presence of phantom volumes. These are artifacts which appear when an area of the working volume is occluded from the view of all cameras, for example because of an unfavorable position of body limbs of the subject at some moment during the acquisition (Figure 2-4). The number of cameras is a critical parameter to deal with this problem [33].

**Figure 2-4 Example of visual hull. Phantom volumes are present, i.e. between arms**

## *2.4 MODEL DEFINITION*

The model represents the a priori knowledge available about the subject's morphology (3D shape) and kinematics.

### Morphology

The 3D shape information is expressed as a triangle mesh (Figure 2-5), preferably obtained using a laser scanner. A general triangle mesh is described by an ordered list of 3D points, called vertices of the mesh, and by a list of the triangles which describe how the vertices are connected. In particular, each vertex is uniquely identified by its position in the list, and defined by its three coordinates in a

reference coordinate system; each triangle is described by the identifiers of its three vertices in counter-clockwise order (when viewed from outside).

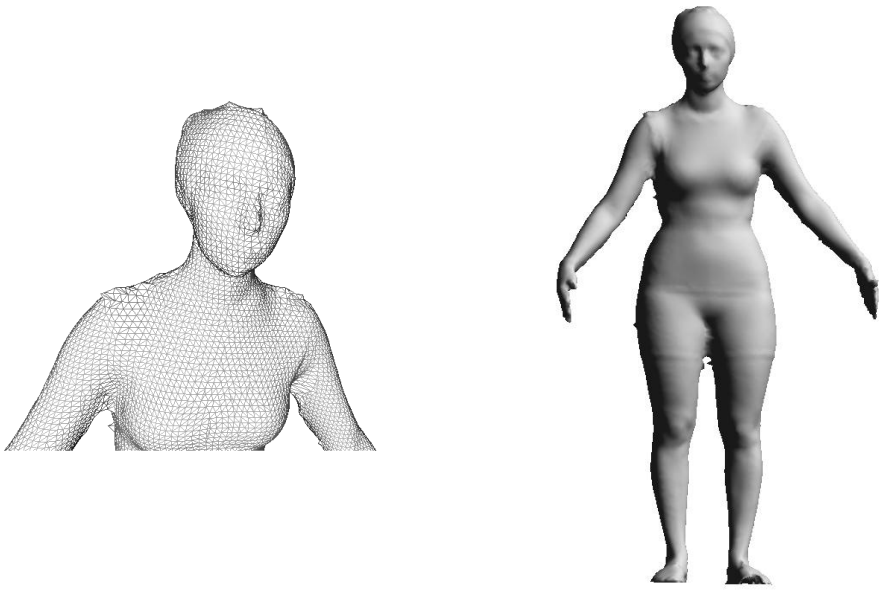


**Figure 2-5 Model mesh: detail of the triangles (left) and whole subject's mesh (right)**

## Kinematics

The kinematic information is represented by an articulated model of the human body. The model for general full-body analyses consists of 15 rigid segments, corresponding to the principal clusters of rigidly-linked bones (head, torso, pelvis, arms, forearms, hands, thighs, shanks, and feet; Figure 2-6), linked through joints with 6 degrees of freedom (3 translational and 3 rotational). However, different models can be defined, depending on the particular needs or problems at hand. Each segment has a "parent" segment (except for the root), to which it is connected through its "parent joint"; similarly, segments and can have multiple "child segments", to which they are connected through "child joints". A local technical coordinate system is embedded in each segment; its origin is coincident with the parent joint and, for segments that only have one child, the longitudinal

(Z) axis is the line connecting child and parent joint. The root segment is initialized with the same orientation as the global coordinate frame, while segments without children inherit the Z orientation from their parent (Figure 2-6). X and Y axes are defined starting from the global frame of reference so that orthogonality is respected for every segment. The pose of the model is completely determined by the position and orientation of each body segment's coordinate frame relative to its parent's: the body segments form a kinematic chain. Rotations are expressed using the exponential maps formulation [34][35] to allow an easier calculation of points' position derivatives with respect to rotational parameters, which are needed in the matching process described in section 2.5.

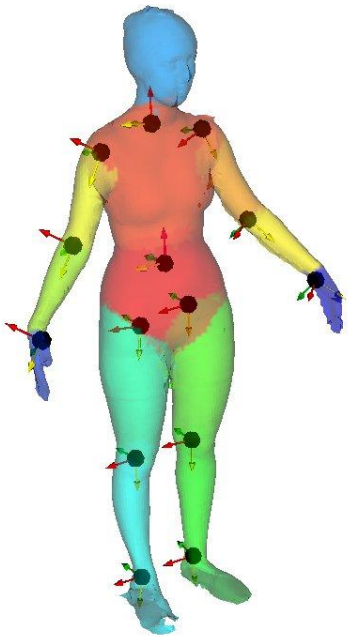


**Figure 2-6 Segmented model of a subject; each color represents a body segment. Joint centers' position (red circles) and segments' embedded frames of reference are shown.**

Joint constraints are also part of the *a priori* information that can be exploited, as they prevent the model from reaching anatomically incorrect configurations.

### Automatic model generation

A procedure has been developed to obtain subject-specific models from simple 3D surface representations of the subjects [36]: joint centers' positions are automatically obtained, as well as segmentation of the mesh into the different body parts.

This algorithm exploits the space of human shapes [37], built on a database of laser scans of people in the same reference pose. Dimensionality reduction has been performed on the meshes in the database through Principal Component Analysis (PCA). In this way, human body shapes can be expressed as free-form meshes by a reduced number of parameters, as linear combinations of principal components. Usually, the first ten principal components give a sufficient level of detail for the applications described in this thesis. Examples of body shapes' variations in function of principal components are shown in Figure 2-7.

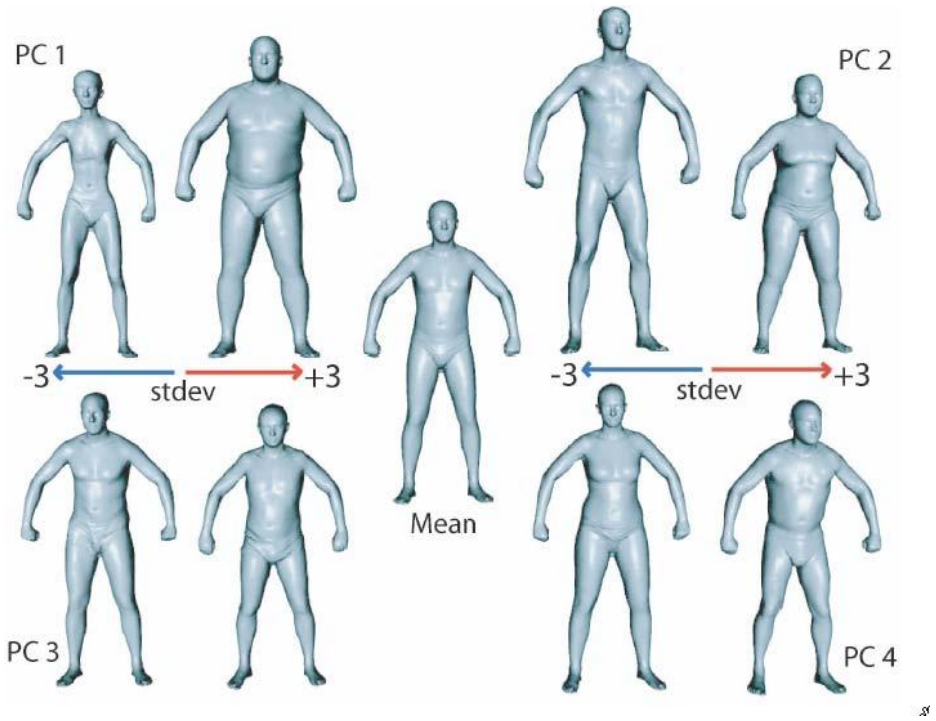



**Figure 2-7 Effects of principal components' variation on human body shape space**

A reference mesh (the center of the space of human shapes) has been segmented, assigning a body part to each of its vertices, so that registering it with any other

mesh will also provide the segmentation of the latter. The registration iterates the following steps, summarized by Figure 2-8:

- find the transformation to be applied to each segment of the reference mesh to match the pose of the subject' mesh, using a pose registration algorithm such as the one described in section 2.5
- segment the subject's mesh, assigning each vertex to the same segment as the closest vertex on the transformed mesh obtained in the previous step
- apply the inverse of the transformation found in the first step to all the segments of the subject's mesh; this way, the subject's mesh is registered in the reference pose
- morph the reference mesh into this new subject's mesh: this is achieved finding the human shape space parameters which describe the mesh most “similar” to the subject's mesh, where the similarity is mathematically defined as the sum of squared distances from the vertices of the mesh to the corresponding closest vertices on the target mesh.

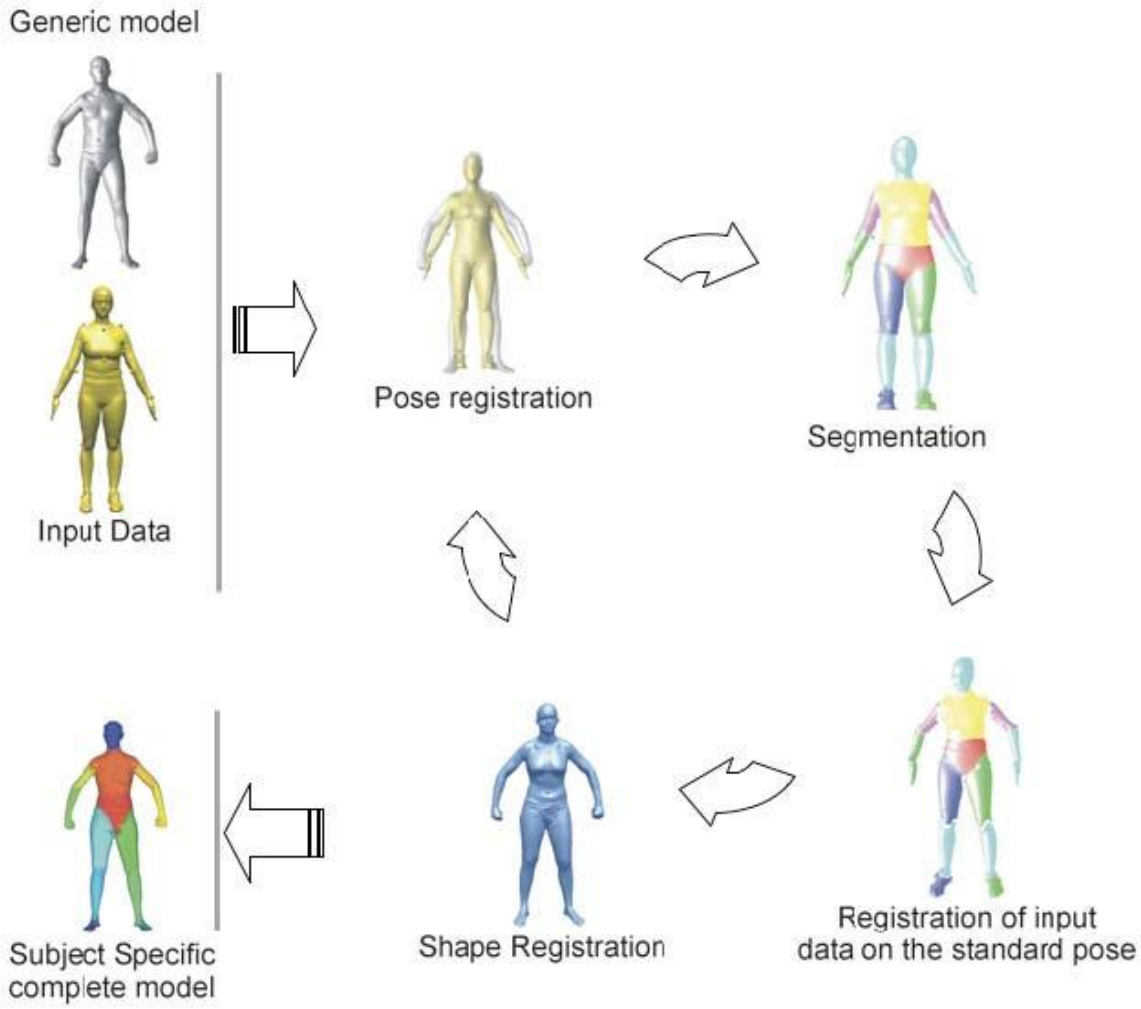



**Figure 2-8 Automatic model generation pipeline**

At every iteration, the reference mesh has a different shape; in 4 or 5 iterations, the algorithm usually converges, and the segmentation of the subject's mesh is obtained.

Each joint center position is then given as a linear combination of the positions of seven vertices on the morphed reference mesh; the number of necessary vertices, their identifiers and the regression coefficients were determined from a training set of 9 subject meshes, in which the joint centers had been identified through virtual palpation, after they had been registered in the reference pose and shape as described above [36]. For most of the joints the error has been shown to be similar

to the marker placement error due to inaccurate identification of the bony landmark on the subject's skin when marker-based motion capture is performed. For the hip joint center the error is significantly greater, because the anthropometric method, commonly employed in biomechanics, used to identify them can lead to errors on the order of several centimeters [38]. In order to overcome this problem, a functional method has been developed to bring the error in the estimation of hip joint center location within biomechanically acceptable limits [39].

It is also important to notice that whenever a subject's laser scan is not available, a visual hull of the subject standing in orthostatic posture can be used to generate a model from the human shape space. The model thus generated usually looks "fatter" than the actual subject, since the visual hull is an over-approximation of the object it represents. The model of the same subject as in Figure 2-6, generated from a visual hull instead of a laser scan, is shown in Figure 2-9.

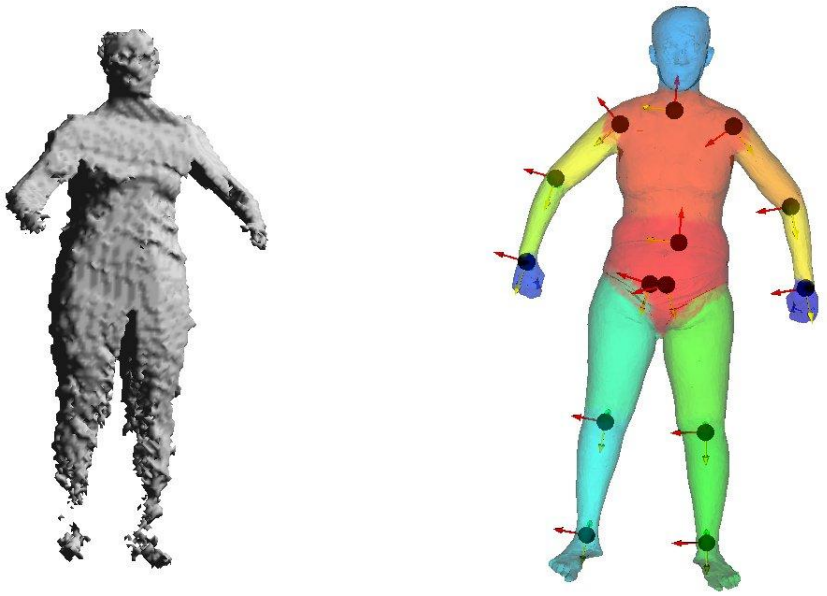


**Figure 2-9 Static visual hull of subject in Figure 2-6 (left) and model generated from it (right)**

## *2.5 MODEL MATCHING*

In order to identify the motion of the subject throughout a sequence, an algorithm is needed to recover his instantaneous pose frame by frame, exploiting all the available information. This is achieved finding the particular configuration of the model which better explains the data, represented by the visual hull surface, while being anatomically correct because it respects the bounds imposed on the kinematic chain. The configuration is found matching the points of the visual hull to the points of the model mesh, using the Articulated Iterative Closest Point (ICP) algorithm, a modification of the standard ICP [40]. The opposite matching could be performed (model to data); the two behave differently in the handling of corrupted data. In the first case, the algorithm is less robust to "phantom volumes" in the visual hull, caused by concavities or self-occlusion; in the second, missing chunks in the visual hull (especially missing limbs), due for example to errors in the background subtraction, lead to the model reaching completely wrong configurations.

ICP-like algorithms perform the point matching between two sets of points (e.g. set $\mathcal{V}$ and set ) iterating between the two following steps:

- for each point belonging to $\mathcal{V}$, find the corresponding closest one in $\mathcal{M}$ (or viceversa), according to some metric
- find and apply the relative transformation between the two sets (usually a rigid body transformation) which minimizes the total distance between corresponding points.

In our case, the metric chosen for the first step is the Euclidean distance; the search for the closest point (nearest neighbour) involves the pre-computation of a particular structure, a KD-tree [41], to store the points in . The query of this structure is significantly faster than an exhaustive search over all the points in .

As for the second step, the transformation consists in a change of the model configuration. The configuration of an articulated model with $n$ body segments is function of $6 \cdot n$ parameters, which can be represented by a vector $\beta$. Six parameters represent the degrees of freedom associated to the position and orientation of the root segment relative to the global coordinate frame:

Parameters are the components of the translation vector that defines the position of the origin of the root segment's embedded frame of reference. Parameters are the components of the rotation vector , that is a concise representation of the rotation (*special orthogonal group*) of the root segment's frame of reference. The direction of , identified by the unit vector , indicates the axis along which the rotation occurs; the magnitude of the vector, is the value of the angle that describes the magnitude of the rotation about the axis. The two formulations (rotation matrix and rotation vector) are related by an exponential map transformation

:

creating, from a normalized vector , the skew-symmetric matrix

, where $so$ denotes the *special orthogonal Lie algebra*, the following relationship holds:

2.1

The rotation vector representation can be retrieved from rotation matrix with the following calculations:

2.2

2.3

Similarly, for all other body parts, the six parameters represent their position and orientation, each relative to their own parent's coordinate frame.

The position of a vertex on the model mesh is therefore a function of a subset of these roto-translation parameters: only the segment the vertex belongs to, and all the segments which link it to the root of the kinematic chain, need to be taken into consideration for that vertex.

In mathematical formulae, this problem can be expressed as follows: given a set of $N$ points belonging to the visual hull, , a set of points belonging to the model, which are function of a vector of parameters $\beta$, and an injective function so that, for fixed is the point on the model that is closest (Euclidean distance) to the point we need to find

2.4

This is a nonlinear least-squares problem, since the position of a model point in the global frame of reference, given its position in the corresponding segment's frame of reference, and the set of frames that connects this frame of reference to the global one, is

2.5

where are constrained to be rotation matrices. The Levenberg-Marquardt approach [42] has been employed in order to solve this problem. It is an iterative procedure which interpolates between the Gauss-Newton algorithm and the method of gradient descent, depending on a viscosity (damping) factor which changes at each iteration according to the linearity of the problem. Defining as the vector between and the origin of the frame of reference of segment *d*, linearization of around the working position by first-order Taylor expansion yields:

2.6

where

2.7

The first three columns of derive from the formulation of infinitesimal rotations as exponential maps, so that first-order derivative of the rotation of vector *v* can be expressed as

2.8

At each iteration of the Levenberg-Marquardt algorithm, the following linear system must be solved in :

2.9

where $J$ is the matrix given by the vertical concatenation of all , and D is the vector of length 3N given by the vertical concatenation of corresponding distances .

Since the number of visual hull vertex can be very large ( >20000), the memory occupancy of matrix $J$ can highly affect the computational performance of the algorithm. To overcome this problem, only matrix (size ) and vector (length ) are kept in memory, and updated sequentially when each point on the visual hull is considered.

As for the choice of damping parameter , indications from Nielsen [43] were followed

2.10

Furthermore, a weighting factor multiplies all equations relative to rotational degrees of freedom, in order to distribute the damping parameter effect on translations and rotations.

Finally, two mechanisms are available to prevent the model from reaching anatomically incorrect configurations:

- a soft bound in the form of a polynomial function to be added to the cost formulation of the least squares problem, which discourages the minimization algorithm from moving to configurations close to the edges of the “anatomically correct” range
- a clipping routine, which at the end of each ICP iteration checks if the bounds are respected, and in case replaces the model configuration with the acceptable solution that best approximates it.

The articulated-ICP approach to shape matching is based on correspondence between couples of closest points on two meshes: the algorithm therefore acts "locally", not considering all the possible combinations of corresponding points on the two meshes. This reduces significantly the computational costs, at the expense of an increased sensitivity to the initialization of the pose of the model. Similarly, the choice of an iterative algorithm such as LM instead of MonteCarlo techniques, which are more robust, but extremely computationally expensive, has been motivated by the need to keep processing times at an acceptable level.

# 3 DEVELOPMENT OF AN UNDERWATER MARKERLESS MOTION CAPTURE TECHNIQUE FOR STUDYING SWIM BIOMECHANICS

## *3.1 BACKGROUND*

In order to move faster, swimmers move cyclically their trunk and upper limbs on multiple reference planes. In particular, arms' movement involves multiple changes in direction and velocity (called "sweep" phases): this way, the swimmer yearns for "still" water, which offers more resistance to the hand. The resulting movement is complex, and difficult to visualize. Analysis of swimmers' kinematics is indeed a challenging problem in the field of sports biomechanics because of the difficult experimental conditions that affect the setup of motion capture devices. The presence of water and the limited space available in swimming pools are some of the factors that hinder the use of electronic devices. For example, commercial stereophotogrammetric systems that employ reflective markers cannot be adopted. Hence, research on motion analysis of swimmers is commonly based on video recordings of the subject's motion. 2D analyses require a single camera; points of interest are digitized on each video and analysis of movement on the sagittal plane is performed [44]. Since the motion in front-crawl swimming occurs on different planes [45], for a descriptive analysis of three-dimensional (3D) motion a multi-camera setup is needed, which requires not only calibration, but also synchronization of the cameras. These procedures are hindered by the underwater experimental environment[46]. Furthermore, videos are analyzed by manual digitization of feature points on all images, which may correspond to either visual markers drawn on the subject or crucial points, such as joint centers, identified by the operator. This procedure has two main drawbacks: it is time-consuming and can easily lead to misidentification of features, especially when a large number of points are involved. Alternative approaches based on accelerometric sensing units have also been adopted [47][48]. This technology is relatively cheap and provides higher sampling rates; however processing and interpretation of measured data is not straight-forward.

Furthermore, sensing units may encumber the subject, which hinders their application for analysis of swimmers' performances during competitions.

In this context, the application of video-based, markerless approaches to motion analysis of front crawl swimming was investigated. They do not require to draw or attach markers to the athlete skin, thus are suitable for sport quantitative analysis during training or even competition.

## *3.2 EXPERIMENTAL SETUP*

Video-based markerless analysis of front crawl swimming requires subaqueous video cameras that record the subject; they need to be synchronized and spatially calibrated in order to perform a three-dimensional reconstruction of the subject. The appropriate positioning of the set of cameras is one of the most challenging aspects for the adoption of the markerless technique in the swimming pool environment. The equipment employed for these studies is described in the following paragraphs, along with the synchronization technique and the calibration procedure adopted. A great effort has been made in trying to perfect a low-cost setup, which could be widely used by coaches and researchers, since it is based on out-of-the shelf cameras and computer hardware. Finally, the setup adopted in laboratory conditions for dry acquisitions is described. It has been employed for a full body acquisition of the subject standing in a static pose; the visual hull (see section 2.3) obtained from these videos is needed for the creation of a subject-specific model, exploiting the algorithm presented in section 2.3.

### Equipment

Six underwater colour analog cameras (TS-6021PSC; Tracer Technology Co. Ltd) were employed in the experiments presented in this thesis (Figure 3-1).


**Figure 3-1 Color analog camera by Tracer Technology Co. Ltd**

Each camera was connected to a FireWire (IEEE 1394a)-equipped notebook through an Analog to Digital Video Converter (Canopus ADVC55; output DV video, PAL interlaced, 25 frames/second). A total of 4 notebooks was used; out of them, 2 were equipped with PCMCIA IEEE 1394a cards and could therefore be connected to two cameras. Although in theory two cameras could be connected to the same Firewire controller, provided that multiple input ports are available, this solution has proved experimentally unreliable. Each DV-PAL video involves a data rate of ~36 Mbps, and the converters impose a transfer speed of 100Mbps to the controllers, only the 80% of which is available for isochronous utilization. The bandwidth needed to capture from two cameras is 36*2≈72Mbps; however, allocation of isochronous bandwidth has failed in multiple occasions during tests, and a more reliable configuration has been adopted adding PCMCIA cards. All notebooks were linked to a hub through Ethernet cables. The swimming pool where cameras were placed was 25 meters long and 16 meters wide; depth varied from 1.20 to 1.70 meters, but acquisitions only regarded the deepest half. A calibration grid, sized 2.07 x 1.07 x 1.40 meters was placed at the center of the volume of interest before the actual acquisition, and used for the calibration of the extrinsic parameters of the cameras (Figure 3-2). 12 control points were in fact identified on it, and their position relative to an embedded system of reference were known.

**Figure 3-2 Grid employed for extrinsic calibration of the cameras**

A wooden panel with a checkerboard pattern drawn on it (13 x 9 black and white squares, side 42 millimeters) was employed for the calibration of the intrinsic parameters (Figure 3-3).

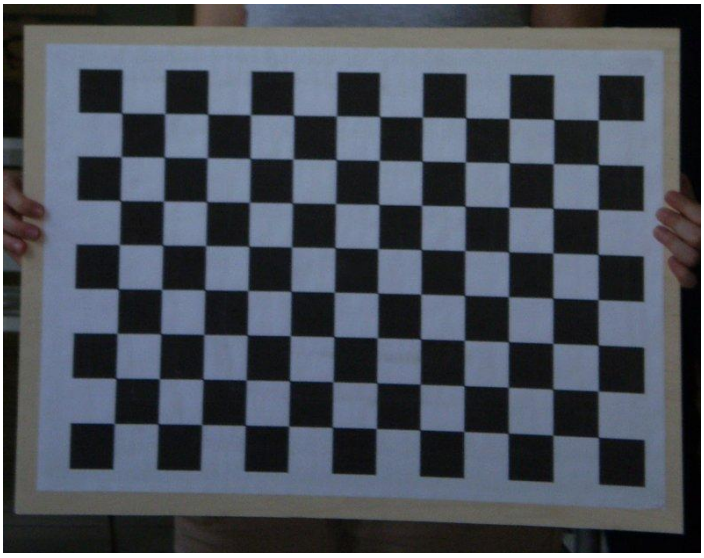


**Figure 3-3 Checkerboard panel employed for intrinsic calibration of the cameras**

Finally, for manual digitization of feature points on videos, commercial software SIMI Motion (SIMI Reality Motion Systems GmbH) was adopted. This was needed for a comparison of the novel markerless technique with traditional methods.

**Synchronization**

In order to synchronize the acquisition from all cameras, a software solution was investigated, that could allow to acquire data simultaneously from all A/D converters. Different hardware solutions were excluded for various reasons: first, the need to keep the system setup as simple and safe as possible; secondly, the desire to keep the setup as cheap and accessible as possible, and to employ the cameras and converters that were available. The custom-made software that was developed in C++ language is based on the Microsoft *DirectShow* framework, which follows the *Windows Component Object Model* (COM) interface, and on TCP (Transmission Control Protocol) and UDP (User Datagram Protocol) network protocols.

*DirectShow* is an extensible, filter-based framework that can render or record media files. In particular, software code for accessing DV video streams from FireWire inputs, and writing them to hard disk, was based on the DVapp Sample application ( *http://msdn.microsoft.com/en-us/library/ms783409(v=VS.85).aspx* ) included in the Windows Software Development Kit (SDK). Handling of multiple video devices, which is a feature needed when more than one converter are connected to the same PC through additional FireWire cards, was implemented following the design provided by the open source *videoInput* library by Theodore Watson (*http://www.muonics.net/school/spring05/videoInput/*). Synchronization of capture from all devices is then achieved setting the same "reference clock" for all **IMediaControl** DirectShow interfaces, the function of which is to control the streaming of media through the filters in the graph.

Network protocols that allow to synchronize the capture in all PCs were developed through the **Windows Sockets API** (**WSA**, or **Winsock**). A different IP (Internet Protocol) address of type 192.127.1.xxx is manually set on each PC (LAN configuration properties). One of the PC acts as "server", while the others are "clients". When the server starts running, clients can connect to it via TCP/IP protocol. Once the communication has been established, setup information is

exchanged: each client transmits the number of devices connected to it, and receives the name of the file that video data will be saved with. In this phase, it is possible to preview the video images transmitted by each camera. When the start signal is given to the server, the message is relayed to all clients as an UDP broadcast signal, which is less reliable but generally faster than TCP-transmitted signals. This is due to the fact that, for UDP, no error-correction mechanisms are in place, which would require that the client sends confirmation of message reception before program execution continues.

No direct control is however possible on the operation of A/D video converters, which are independent one from the other. Therefore, due to the temporal discreteness of input video signals, delays of up to one frame have sometimes been found on the videos, and in some case corrected through the identification of the same event in all video sequences (i.e., hand entering the water).

**Calibration**

Calibration of the subaqueous cameras is required for the three-dimensional reconstruction of the subject. The position and orientation of the cameras in space (extrinsic parameters), as well as the parameters of the optical model that explains how the 3D scene is projected to the 2D image frame (intrinsic parameters), need to be estimated. The presence of water however renders this operation more complicated, because of the distortion due to refraction of the light beams [49]. Commonly, calibration techniques based on direct linear transformation (DLT) have been proposed in literature [50], [51]. Intrinsic and extrinsic parameters are estimated at the same time from the projection image of a calibration grid of known size. Similarity equations are written for every control point, and then combined into a system that is solved by standard methods. In this study, a two-step calibration algorithm was adopted instead [52], in which intrinsic parameters are estimated first from a series of images of a checkerboard-patterned plane. This allows to perform estimation of intrinsic parameters on a greater number of points (all corners, in all images), distributed on the whole image. The optical model that

has been assumed is the “pin-hole” camera, with inclusion of radial and tangential distortion. For ideal pin-hole cameras, the following relationship holds between a 3D point expressed in the camera frame of reference, and the corresponding point’s projection on the image plane :

– 3.1

where is the focal length expressed in pixel size, respectively horizontal and vertical; is the position on the image plane of the central point; describes the skewness of the two image axes, but is often considered unnecessary and set to 0. When the distortion component is added to the model to take into account the presence of the lenses, coefficients for radial polynomial ( and and tangential “thin prism” ( and ) distortion appear in the equations. Defining , the projection equation becomes

3.2

–

Initial estimate of focal length and central point parameters is given by the closed form solution of the pin-hole camera equations; from these, an initial estimate of distortion parameters is calculated, then all parameters are optimized minimizing the global distance between feature points (corners) measured on the images, and the projection, through the complete model, of the 3D points that generate them. A non-linear least square minimization algorithm (Levenberg-Marquardt) is employed.

Once the intrinsic parameters are known, the extrinsic parameters can be easily calculated via DLT technique from the projections of a set of >3 control points; more numerous sets are usually exploited to increase the estimation robustness.

Since calibration of intrinsic parameters has been performed out of water, they need to be corrected for use in the underwater environment. The change in refractive index from air ( to water has to be taken into account. Following the work of Lavest et al. [53], the focal length has been multiplied by ; radial distortion coefficients consequently become

— —

This calibration procedure has been developed employing the Calibration Toolbox for Matlab by Jean-Yves Bouguet (*http://www.vision.caltech.edu/bouguetj/calib_doc/*). The position and orientation of the cameras as reconstructed by calibration is shown in Figure 3-4.

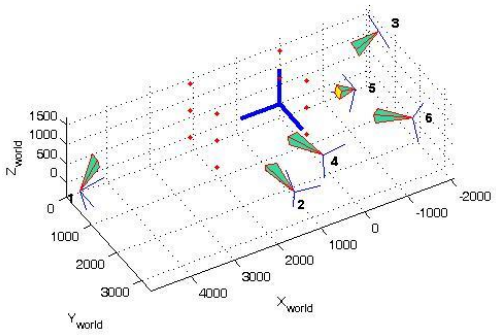



**Figure 3-4 Cameras' position as resulting from extrinsic calibration (coordinates in mm). Control points are shown in red.**

Effects of intrinsic calibration for distortion correction are shown in Figure 3-5, which is the "undistorted version" of Figure 3-2.

**Figure 3-5 Result of distortion compensation on an image. Straight lines do not longer appear as bended.**

Calibration accuracy has been evaluated both in terms of calibration error and reconstruction error for the 12 control points. Calibration error is determined for each camera as the mean squared distance between measured control points' positions and the projections of the corresponding 3D control points. Results are reported in Table 3-1. Reconstruction error is the mean squared distance between known coordinates of 3D control points, and the corresponding points reconstructed from measured 2D projections; results are reported in Table 3-2.

| camera ID | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| calibration error (pixel) | 2.7543 | 2.4167 | 2.9602 | 2.7560 | 3.2410 | 2.8070 |

**Table 3-1 Calibration error (average over the 12 calibration points in each camera)**

| coordinate | X | Y | Z | 3D |
|---|---|---|---|---|
| reconstruction error (mm) | 6.525 | 6.035 | 8.271 | 11.687 |

**Table 3-2 Average reconstruction error on the 12 calibration points**

### Dry setup

Creation of a subject-specific model has been obtained employing the algorithm described in section 2.3 on a full-body visual hull of the subject. The static acquisition of the subject in a reference pose has been performed in a laboratory, where a synchronized 3-camera system was in place (BASLER A-501 k/kc cameras; F 1.2, resolution 1024 x 754 pixels). Calibration of the system has been performed employing the same calibration grid used in the swimming pool, and the Calibration Toolbox for Matlab (Figure 3-6).

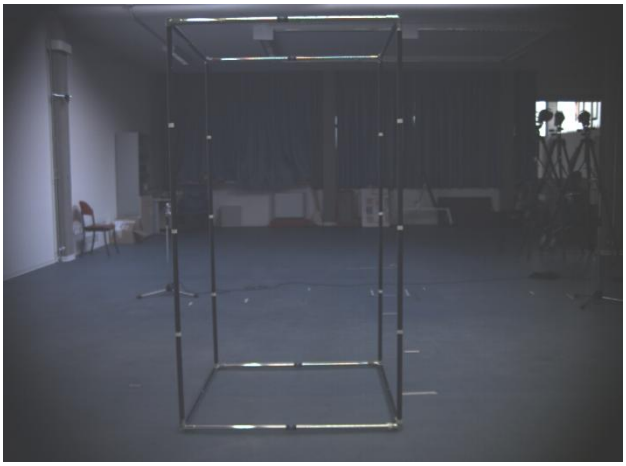


**Figure 3-6 Calibration grid (laboratory setup)**

An example of the acquired images is shown in Figure 3-7.

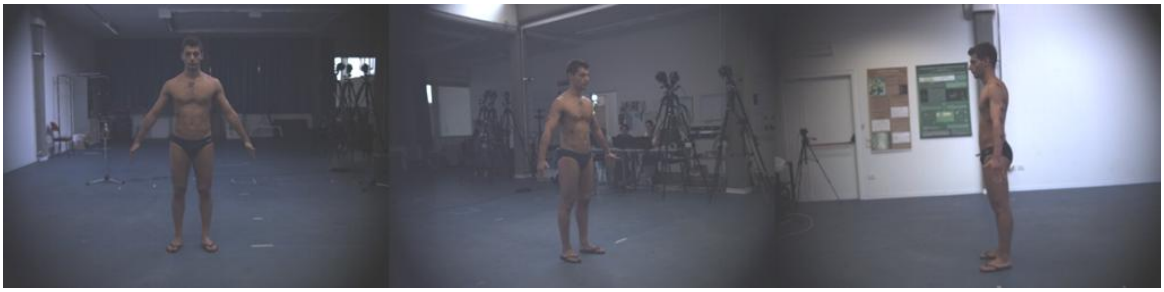


**Figure 3-7 Static acquisition of the subject: view from the three cameras.**

## *3.3 DATA PROCESSING*

In the present section, all the processing steps that have been followed to extract kinematics data from the videos acquired with the setup presented in section 3.2 are explained. An example of an image frame from each video sequence is shown in Figure 3-8.

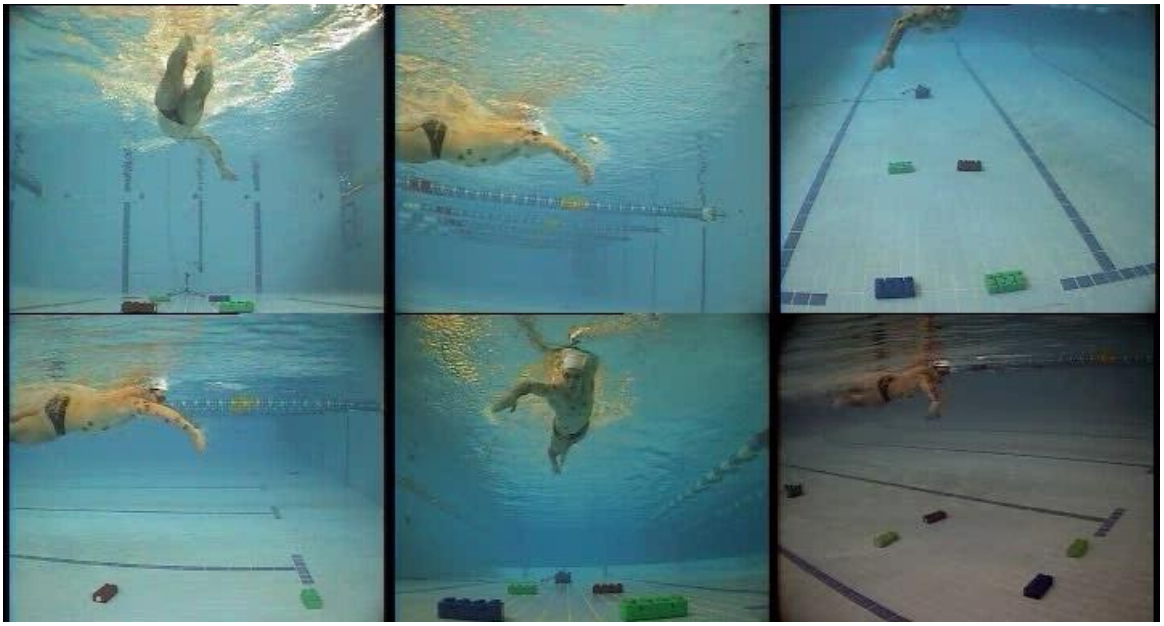


**Figure 3-8 Frame from each synchronized view of the front crawl trial acquisition**

*Deinterlacing*

DV videos obtained from A/D video converters are interlaced. This means that odd and even rows of each image are captured at different instants, where the difference corresponds to half of the delay between two consecutive frames. For PAL video, for instance, nominal acquisition rate is 25 frames per second, but a half image (called field) is captured every 50 seconds. Although this strategy is very useful for recording videos that must be displayed on CRT monitors or televisions, images thus recorded cannot be directly employed for 3D reconstruction. The process of creating a progressive video, where each frame corresponds to a different time instant, from an interlaced video is called *deinterlacing*. The most advanced deinterlacing techniques consist basically in a

spatial and temporal interpolation of pixel values in the images (bob-deinterlacing), so that the resulting video has double frame rate. The adopted implementation was the *Smart-Bob filter for VirtualDub* developed by Donald Graft (*http://neuron2.net/bob.html*). *VirtualDub* is a video processing utility for Microsoft Windows, written by Avery Lee and available as free software (*www.virtualdub.org)*. In order to apply the bob-deinterlacing filter to the videos, they need to be input to VirtualDub as double rate, half vertical resolution videos. This is achieved on-the-fly through the use of the Avisynth frameserver (*www.avisynth.org)*.

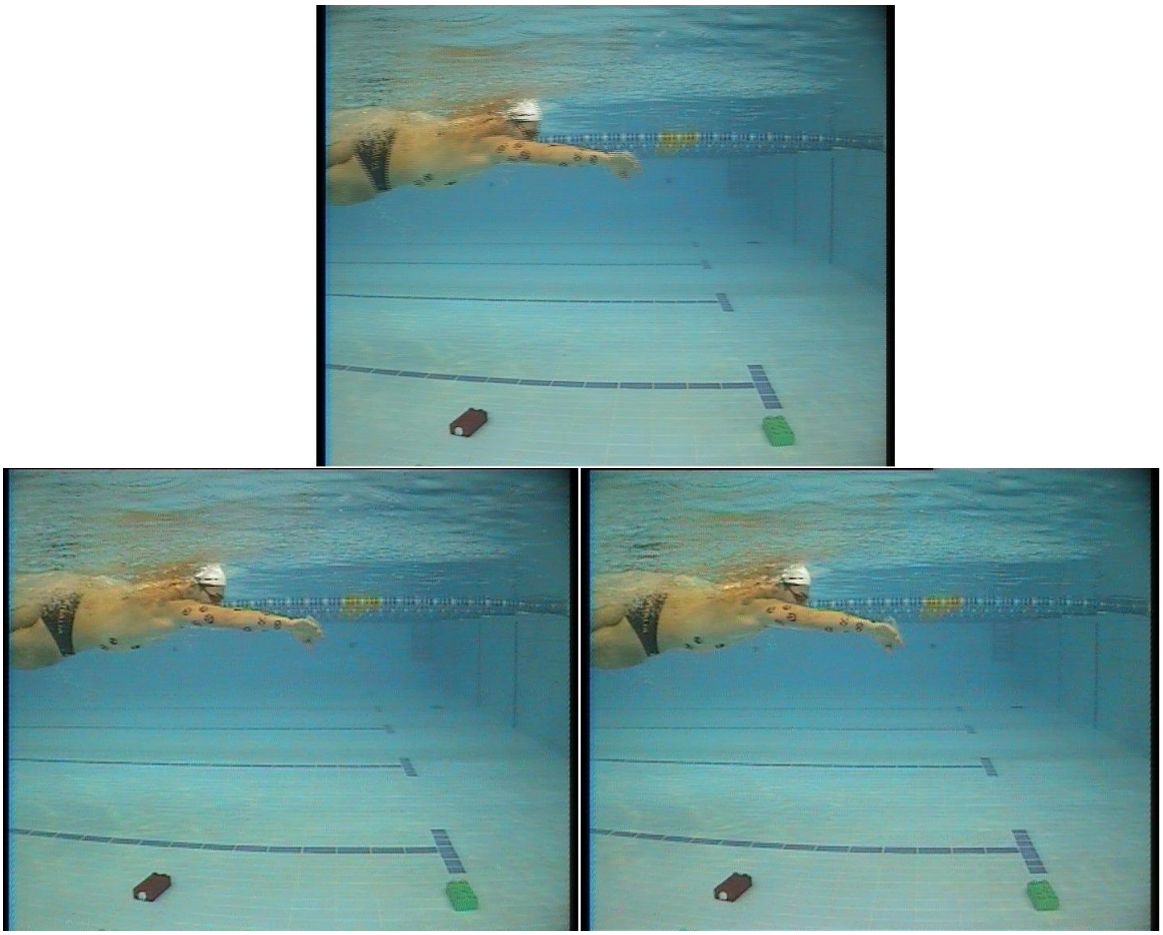


**Figure 3-9 Example of original frame from a DV video (top), and the corresponding frames obtained through interpolation by bob-deinterlacing (bottom)**

The result of this operation on one of the acquired frames is shown in Figure 3-9.

*Background subtraction*

The background of underwater videos is not static: since the air-water surface interface is in view, the ripples that move through it result in a continuous variation of the scene. For this reason, the use of a single background image as reference is inappropriate. A background model is built instead from a video sequence of the empty swimming pool, aiming to characterize the variations in pixel value intensities that define the sole background. The Gaussian-mixture-model approach, mentioned in section 2.2, has been adopted. The implementation provided by the Intel OpenCV open source C++ library (*http://opencv.willowgarage.com/wiki/*) has been taken as starting point. The algorithm has been slightly modified with the addition of a Gaussian component to the Gaussian-mixture-model that is created for each pixel from the background sequence. This is a "white Gaussian", centered on an RGB value sampled from a foamy area of an image in the sequence (for example, red=107, green=170, blue=163; values range from 0 to 255), and with the same initial variance as the other components. Its function is to account for the possible presence of foam, generated by the movement of the swimmer: when a pixel presents an RGB value that is classified as belonging to this white component, it will be considered to be background.

Reflexes however remain the main problem within underwater images, as they cannot be removed from the foreground based on color content of the pixels. Their appearance is too similar to the actual subject, therefore at the moment it is not possible to discriminate between the two.

An example of foreground extraction is given by Figure 3-10.

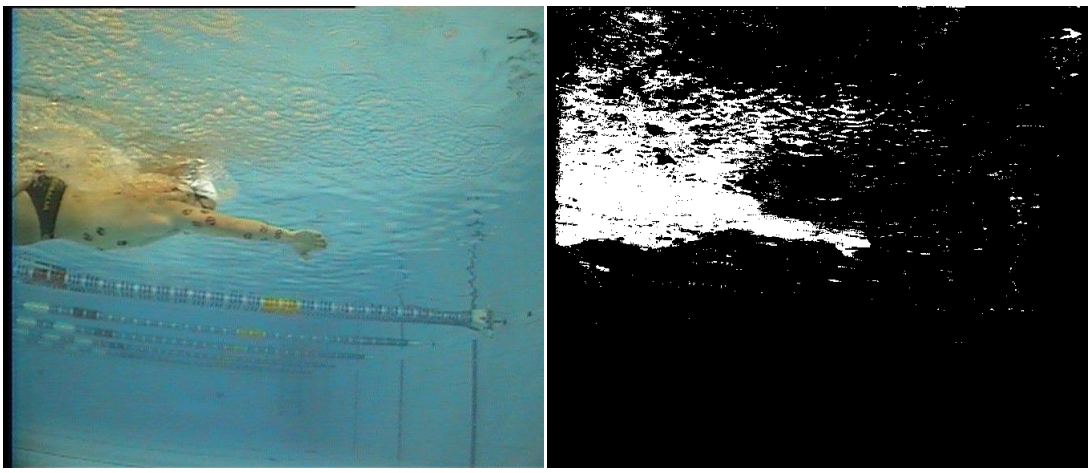


**Figure 3-10 Example of foreground extraction for a lateral-view camera**

Finally, a last step is added to the background subtraction procedure for videos taken from the bottom of the swimming pool (cameras 1 and 5 in Figure 3-4). From these views, the swimmer appears darker than the surrounding water, because external light is blocked; therefore, pixels for which the sum of red, green and blue values is greater than a fixed threshold (i.e. 400) are considered to be part of the background (Figure 3-11).

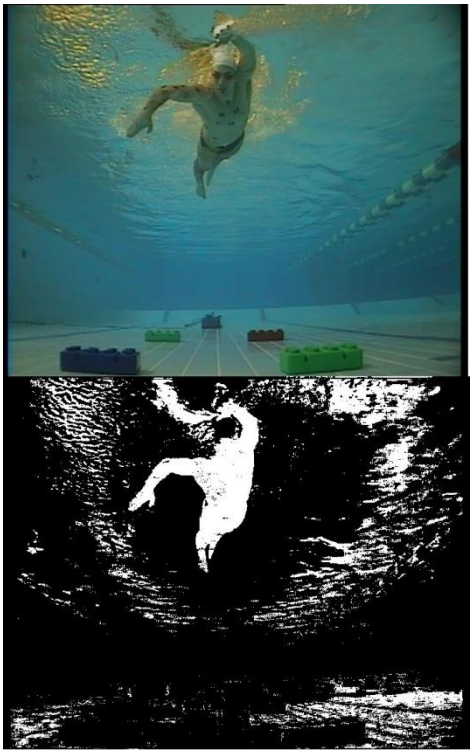


**Figure 3-11 Example of foreground extraction for a bottom-up view camera**

*Visual hull creation*

Visual hulls are created from silhouettes as indicated in section 2.3. The volume of interest is partitioned into voxels of size 0.01 m; then, each of the voxels' vertices is projected on the input silhouettes, exploiting the calibration functions of the OpenCV library, that are analogous to those provided by the Calibration Toolbox for Matlab. The Visualization Toolkit (VTK), an open source software system for 3D computer graphics, image processing and visualization by Kitware, Inc (*http://www.vtk.org*) has been exploited to obtain the 3D surface contour of all the points belonging to the visual hull. The 3D meshes thus obtained are saved in PLY format, also known as Polygon File Format or Stanford Triangle Format. Some examples are shown in Figure 3-12.

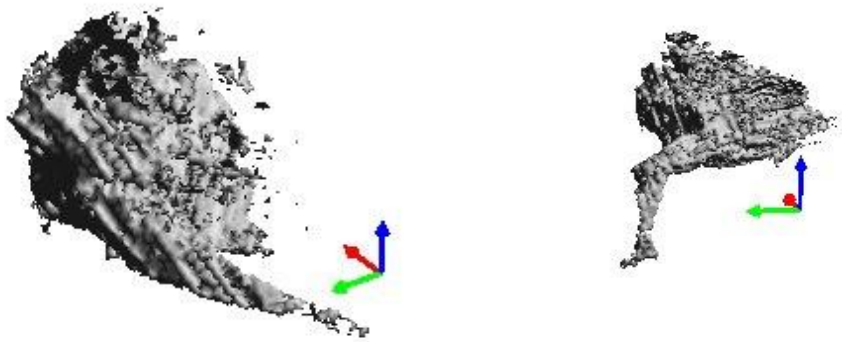


**Figure 3-12 Example of two visual hulls reconstructed from the video sequence. Orientation of global frame of reference is also shown.**

*Model definition and initialization*

A subject-specific model, generated automatically from a dry visual hull (see end of section 3.2) of the subject, employing the procedure explained in section 2.3, has been used. Although the model represents the whole body of the subject, only the body parts that belong to the right arm and the torso are taken into consideration for pose estimation. Furthermore, the kinematic chain starts at the right hand (root segment), then continues with right forearm, upper arm and torso; the latter has the head, the left arm, and the pelvis as children (Figure 3-13).

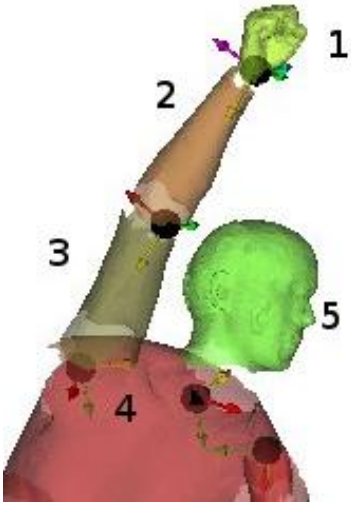



**Figure 3-13 Segmented model of the subject (particular of upper body). Numbers indicate the sequence of the kinematic chain.**

Torso, pelvis and the inferior part of the body are considered as a rigid body. The reason for these choices lies in the fact that for most part of the video sequences, only the arm and the upper part of the body are in view, therefore pose estimation on the lower part of the body can easily reach completely wrong results. This phenomenon is exacerbated by the presence of spurious volumes due to incorrect background subtraction, caused for example by foam or reflexes. The inclusion of all lower body degrees of freedom in the minimization problem would affect its solution with respect to the joints of interest. Similarly, the hand is chosen as root segment because this improves its chance to fit properly visual hull data, without being hindered by an incorrect estimation of torso pose, especially in the initial part of the stroke when the torso is not in view. This is a consequence of using the articulated-ICP algorithm for model matching: if the configuration of the model becomes too different from the truth, the algorithm could incur in local minima that lead the solution further away from the global optimum.

Model initialization is also important for the same reason. The model's orientation has been adapted to the swimming pool's global frame of reference, so that it looks toward negative Z (vertical) direction, while the head points towards

negative X (longitudinal) direction with respect to the feet (for visualization of global frame of reference, see Figure 3-4). The initial rotation of the torso's frame of reference relatively to the upper arms' frame has been set to

as to reproduce initial shoulder flexion (Figure 3-14). Initial position of the wrist joint is manually set to the value found by triangulation of joint center positions on the camera planes.

**Figure 3-14 Initial model configuration, with flexed shoulder. Orientation of global frame of reference is also shown.**

Finally, translation constraints are relaxed especially at shoulder level, because the shoulder elevation that occurs during initial *downsweep* phase deforms the surface of the body in a way that cannot be properly described by rigid body transformation of the model body parts: in the absence of a deformable shoulder model, greater translations are allowed so that incorrect pose estimation for one of the two linked segments does not affect the other. The inevitable consequence is a greater level of uncertainty for estimation of shoulder "joint center" position.

*Extraction of joint centers trajectories*

Once the sequence of visual hulls has been tracked employing the articulated-ICP algorithm described in section 2.5 with a data-to-model approach (visual example

shown in Figure 3-15), the roto-translation matrix defining the position and orientation in space of each body segment's embedded frame of reference is obtained. However, only information regarding the joint centers, which correspond to the frame origin of "children" segments of the kinematic chain, is retained at the moment. Since frames of reference are technical and not anatomical, the interpretation of rotation matrices, or of their decomposition into Cardan angles, is not straightforward. Also, estimation of internal-external rotation angles for arm segments is deemed to be unreliable because of their strong cylindrical symmetry: volume-based registration algorithms cannot therefore discriminate between different levels of internal-external rotation.

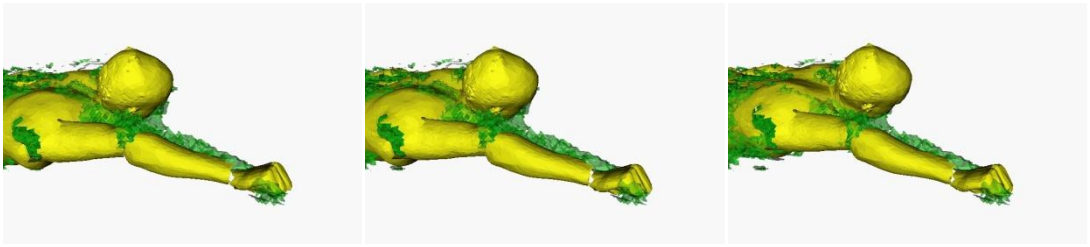


**Figure 3-15 Example of three iterations of the articulated-ICP algorithm for matching of the model (yellow) to visual hull data green)**

The following joint centers are taken into consideration, for comparison with results from traditional swimming motion analysis techniques:

- wrist, connecting hand and forearm;
- elbow, connecting forearm and upper arm;
- shoulder, connecting upper arm and torso.

*Manual digitization*

Multiple operators were asked to digitize the approximate location of wrist, elbow and shoulder on each frame of the video sequences from all cameras through the use of SIMI Motion software. Videos were undistorted first, employing the distortion parameters estimated by intrinsic calibration of the cameras, as to

minimize the difference in calibration of extrinsic parameters between the markerless and SIMI systems. Joint trajectories reconstructed this way have been used to characterize the variability introduced by human intervention in the reconstruction process (inter-operator variability). Similarly, one of the operators was asked to perform the digitization multiple times, in order to assess the intra-operator variability. It is however to be expected that operators from different laboratories may be instructed differently on how to perform the digitization, especially on identification of shoulder position; this would increase the variability of the method.

Even though distinctive features were drawn on the skin on the subject with permanent markers, they have not been used for this study, as the position of markers is different from the definition of joint centers in the markerless model.

## *3.4 EXPERIMENTAL RESULTS*

### Experiment details

One elite sprint swimmer (age 25 years, height 1.82 m, weight 91 kg) participated as subject for the experiment presented in this thesis. A front crawl trial was acquired and analyzed with markerless and manual-digitization techniques. Five operators were required to manually digitize coordinates of shoulder, elbow, and wrist joints on each frame of the video sequences from all cameras, and results were compared (inter-operator variability). In addition, one of the operators performed the manual digitization five times, as to assess intra-operator variability. Mean and standard deviation (SD) curves were calculated for each coordinate of each joint trajectory, for both multiple operators and single operator cases. Root mean square distance (RMSD) was calculated to quantify the agreement between mean curves reconstructed with the manual digitization and coordinate curves reconstructed with the markerless technique. Distance between the reconstructed trajectories was finally compared to the SD that characterizes

the variability of the manual tracking technique. Furthermore, joint velocity was calculated by first-order differentiation of wrist trajectories. No filtering was applied to the data.

## Results

Trajectories obtained with the markerless and manual tracking (multiple operator) techniques for shoulder, elbow and wrist joint centers are reported respectively in Figure 3-16, Figure 3-17 and Figure 3-18. The following phases are identified for the underwater phase of the stroke: A) gliding, B) downsweep C) insweep D) upsweep E) exit. A good agreement is demonstrated especially for wrist joint.

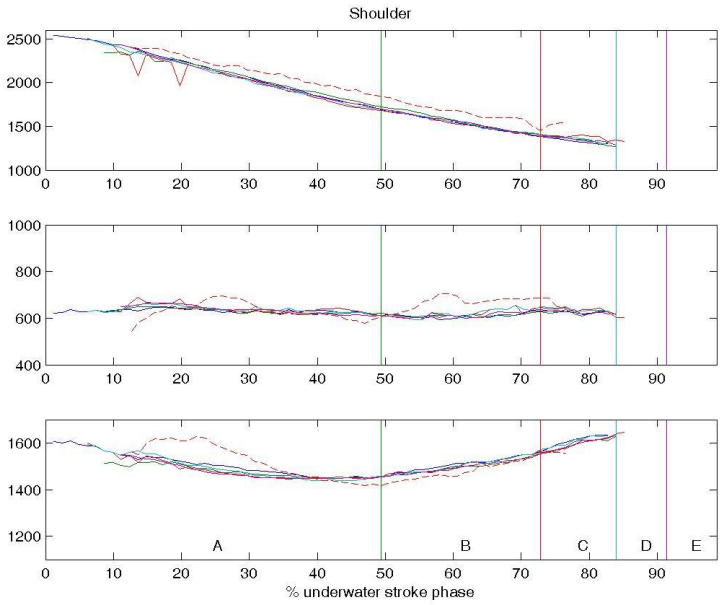



**Figure 3-16 Shoulder trajectory reconstructed through digitization by 5 operators employing a commercial software (solid lines) and with the markerless technique (red dotted line). Phases A-E are explained in the text.**

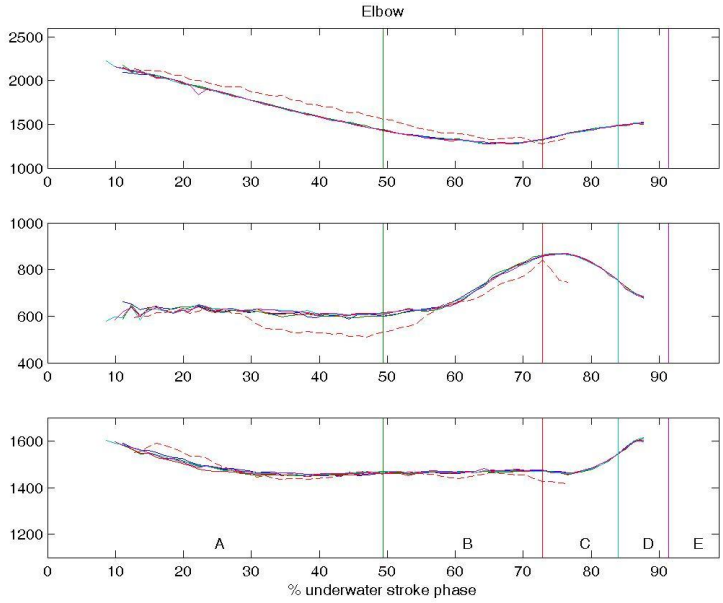



**Figure 3-17 Elbow trajectory reconstructed through digitization by 5 operators employing a commercial software (solid lines) and with the markerless technique (red dotted line). Phases A-E are explained in the text.**

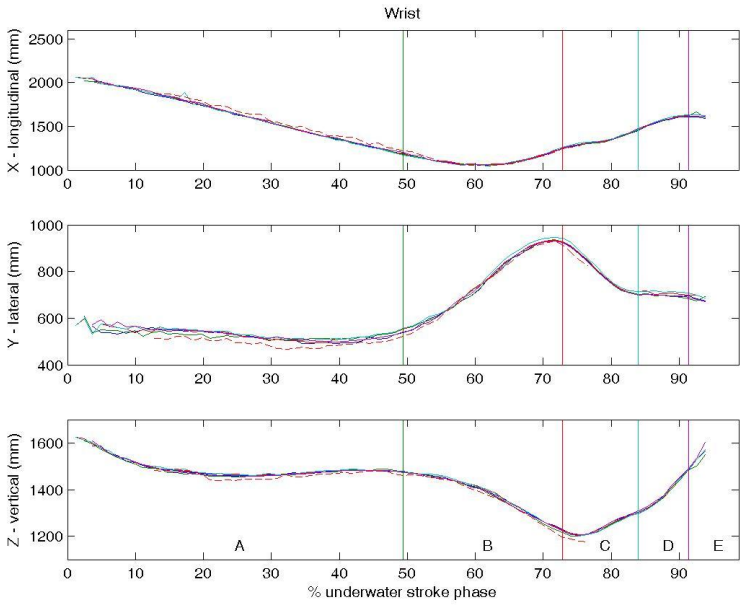



**Figure 3-18 Wrist trajectories reconstructed through digitization by 5 operators employing a commercial software (solid lines) and with the markerless technique (red dotted line). Phases A-E are explained in the text.**

Comparison with trajectories obtained from five manual digitizations by the same operator is reported in Figure 3-19. Figure 3-20 shows the distances between the markerless and average manual-tracking trajectories, compared to ±3 SD curves.

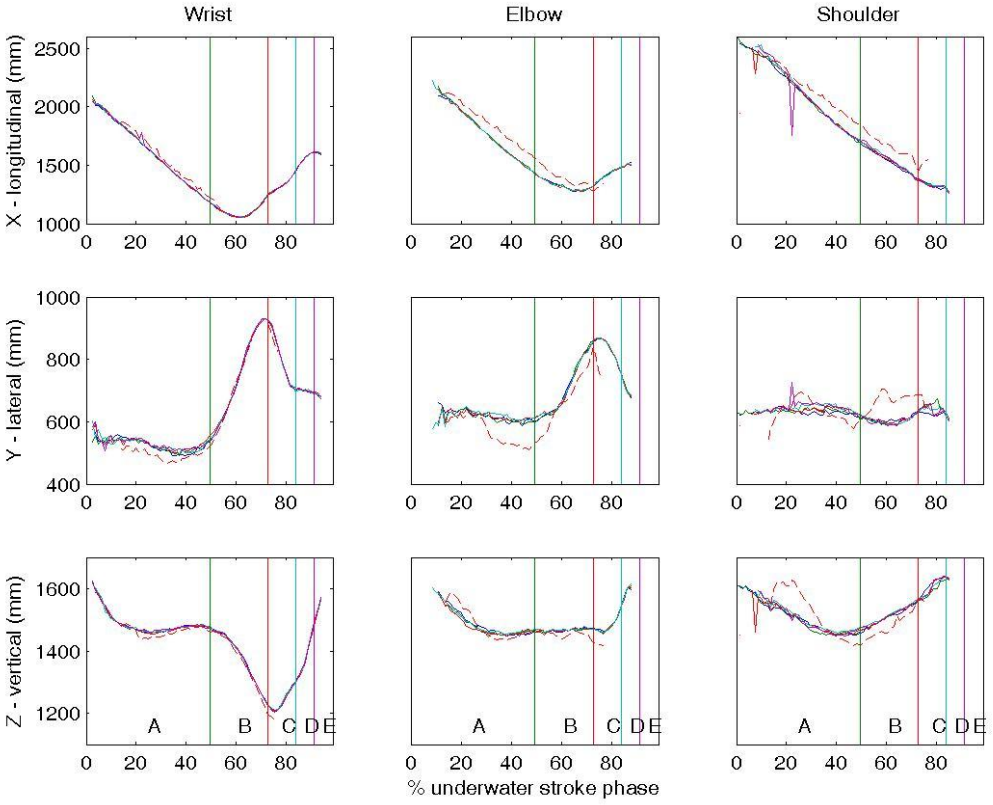



**Figure 3-19 Trajectories reconstructed through 5 digitizations by the same operator employing a commercial software (solid lines) and with the markerless technique (red dotted line). Phases A-E are explained in the text.**

RMSD values between markerless and average manual-tracking trajectories are reported in Table 1, for each coordinate, along with maximum SD values for manual-tracking reconstructed curves. While taking into account the wrist joint, the reported distance (14.4 to 34.9 mm) is comparable to the precision that can be achieved employing the commercial software (maximum SD: 40.9520 mm).

When considering both elbow and shoulder joints, it can be noticed that there is a relevant systematic error along the longitudinal (X) direction (see

Table 3-3). Velocity signals (Figure 3-21 for shoulder, Figure 3-22 for elbow, Figure 3-23 for wrist) appear to be noisier for markerless than for manual digitization; nonetheless, there is agreement in average values, especially for the wrist.

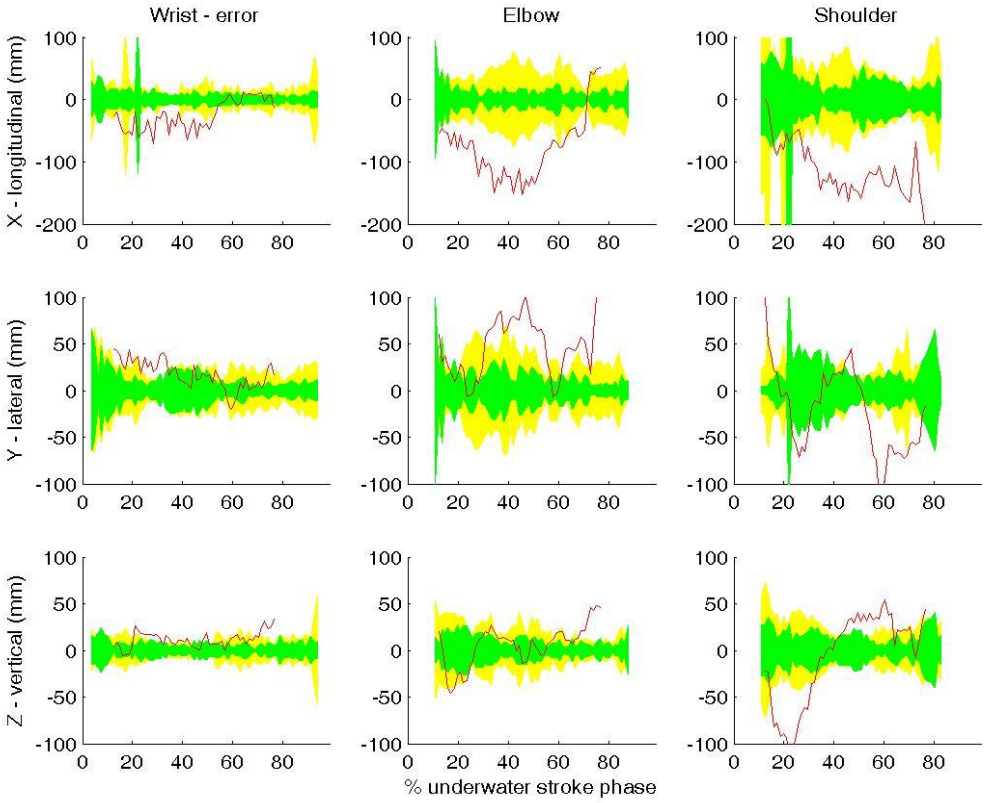



**Figure 3-20 Difference between trajectory coordinates estimated with markerless and averaged manual-tracking techniques (red line). ±3 standard deviation curves calculated on manual-tracking trajectories are shown in yellow (5 operators) and green (same operator)**

| | | RMSD (mm) | max SD (inter operator) | max SD (intra operator) |
|---|---|---|---|---|
| Shoulder | X | 120.7 | 129.6127 | 203.0957 |
| | Y | 50.4 | 21.8433 | 40.4617 |
| | Z | 48.6 | 24.1740 | 17.4831 |
| Elbow | X | 92.8 | 25.6515 | 41.4366 |
| | Y | 56.5 | 23.3312 | 30.2839 |
| | Z | 22.3 | 17.6926 | 8.5519 |
| Wrist | X | 34.9 | 40.9520 | 40.0076 |
| | Y | 23.0 | 22.6055 | 21.8843 |
| | Z | 14.4 | 19.4421 | 7.9039 |

**Table 3-3 RMSD values between markerless and average manual-tracking trajectories (left column); maximum SD values for manual-tracking reconstructed curves are reported in middle (inter-operator) and right (intra-operator) columns**

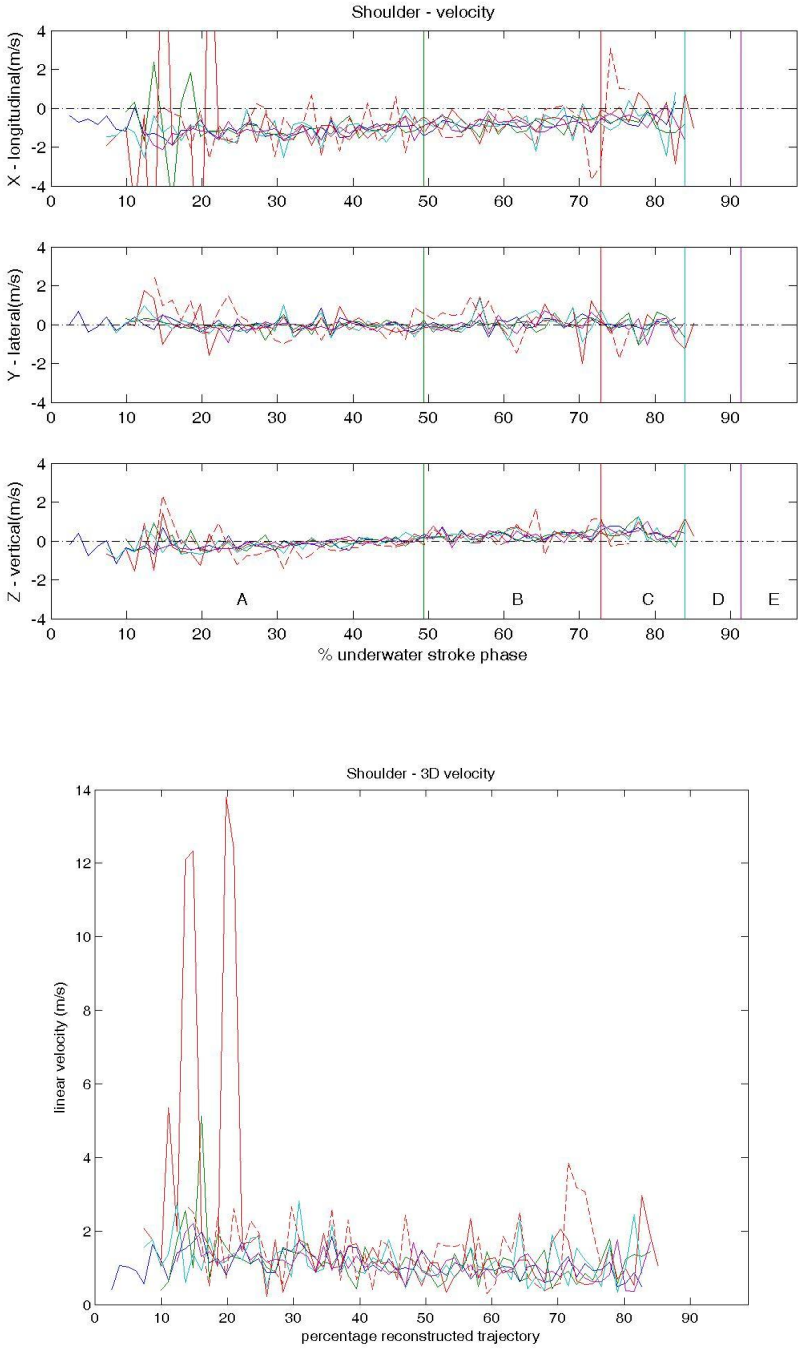



**Figure 3-21 Velocity of shoulder joint center calculated from trajectories reconstructed through digitization by 5 operators employing a commercial software (solid lines) and with the markerless technique (red dotted line)**

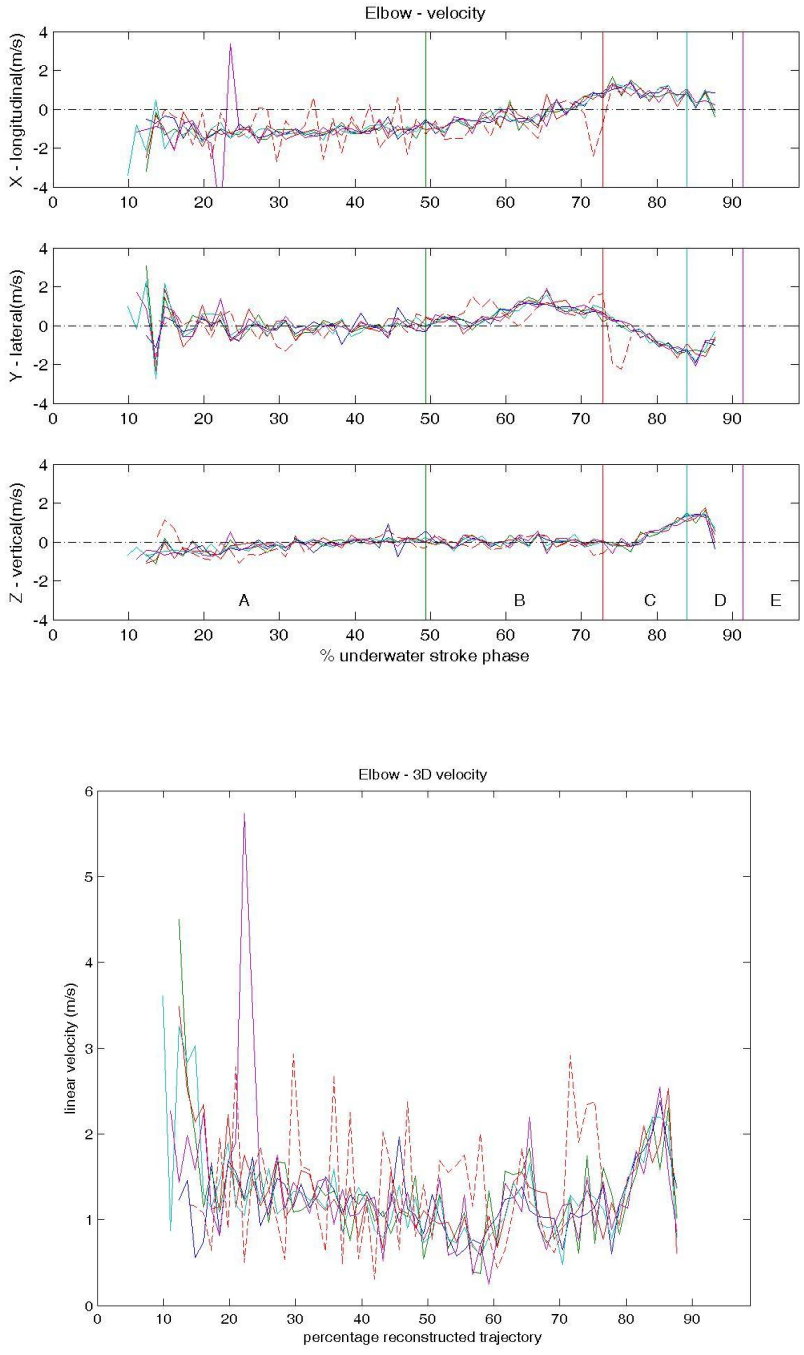
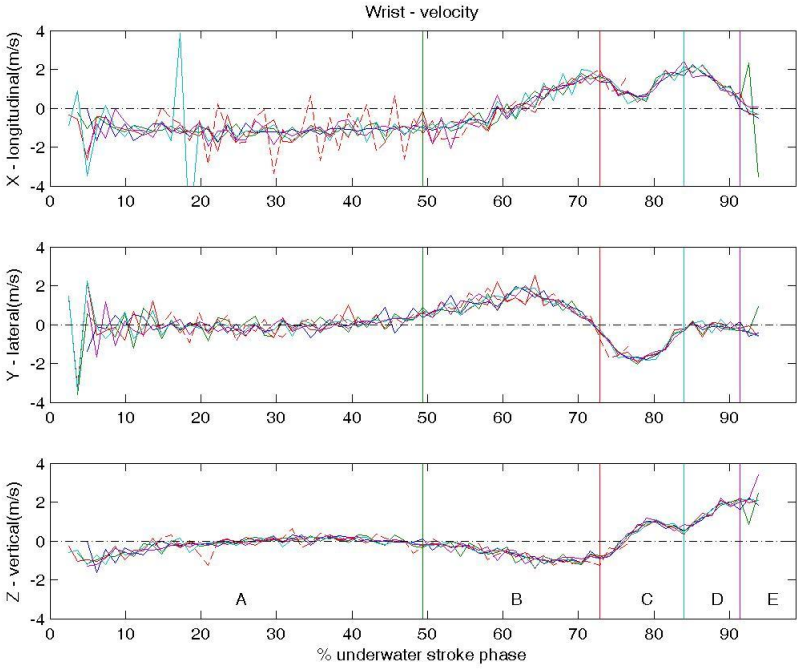



**Figure 3-22 Velocity of elbow joint center calculated from trajectories reconstructed through digitization by 5 operators employing a commercial software (solid lines) and with the markerless technique (red dotted line)**



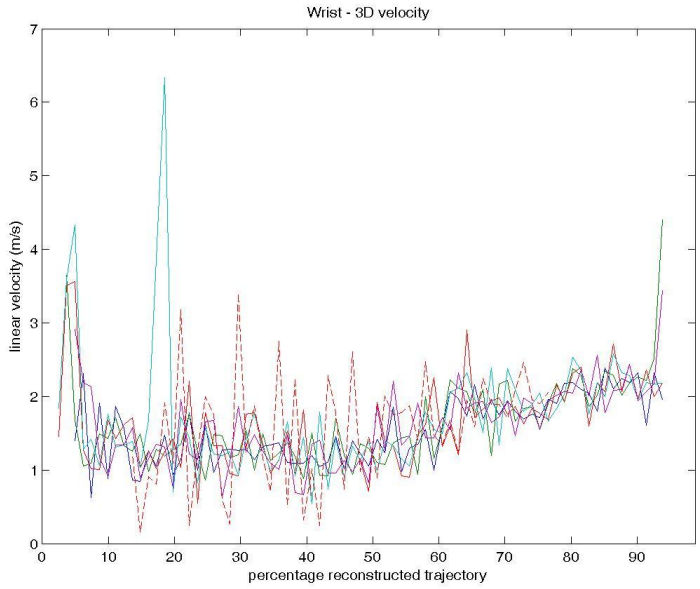




**Figure 3-23 Velocity of wrist joint center calculated from trajectories reconstructed through digitization by 5 operators employing a commercial software (solid lines) and with the markerless technique (red dotted line)**

## *3.5 DISCUSSION*

Upper arm 3D kinematics during a front crawl stroke was reconstructed by means of an automatic markerless technique that was tailored to operate properly underwater. Common out-of-shelf subaqueous cameras were employed. Main modifications regard the calibration procedure and the advanced image analysis algorithms that were adopted. The kinematic properties of the model were specialized for the analysis of arm kinematics. The hand was chosen as root segment of the kinematic chain instead of the trunk in order to allow joint trajectory reconstruction in the initial phase of the stroke, when only the arm is in view of all cameras. Joint trajectories estimation accuracy was evaluated in terms of RMSD with respect to trajectories obtained with a conventional 3D reconstruction technique, implemented in commercially available software. Accuracy at wrist joint level is deemed to be sufficient; this is most important for the technical analysis of the stroke, as hand trajectory is commonly used to identify and characterize stroke phases [54]. The stroke of the subject of this study presents a *downsweep* phase (phase B in figures) characterized by a downward displacement of 26 cm and a contemporary outward movement of 40 cm. In this same phase, the reaching of *catch* position can be identified as occurring at 60% of the underwater stroke period by the sign change in velocity along the X axis. Similarly, beginning of *insweep* phase is clearly indicated by sign change of velocity along the Y axis (73% of underwater stroke period). Larger errors occur instead for elbow and shoulder joints. Systematic difference in the sagittal plane can be due to the rigid-body assumption that is introduced in the definition of the kinematic model for markerless analysis. In this context, the significant surface deformations that occur on the body during the execution of the movement are not taken into consideration, instead higher translational freedom is allowed between body segments. Higher translations however can lead to artifacts in the identification of the joint centers. This is particularly exacerbated, at shoulder level, by the presence of “phantom volumes” in the visual hulls reconstructed for some part of the stroke. This phenomenon occurs because the limited number of

views causes spurious volumes to be recognized as occupied by the subject, indeed they are explained by the available silhouettes. The swimming pool environment, which limits possible camera placement positions, aggravates this drawback; for example, placing cameras on the bottom of the pool reduces excessively the reconstruction volume, and cameras cannot be positioned too close to the water-air interface, because the waves caused by the swimmer's movement would affect the quality of the images in a significant way.

Further limitations of the proposed method lie in the requirement that the whole body under investigation must be in view at all times, therefore reconstruction for the initial and final phases of the stroke, when the arm is partially out of water, is not currently possible. Information from out-of-water cameras cannot be easily integrated with the underwater views because of the issues related to the water-air interface, i.e. waves and foam.

The results obtained with the markerless technique are however encouraging, as this method allows 3D kinematics estimation in an automatic way, reducing processing time and costs. Wide-scale studies are therefore made possible, which can investigate differences in stroke patterns among a group of subjects, or identify performance indices for athletes' evaluation. In addition, the relaxation on the marker drawing requirement allows this technique to be adopted not only in controlled conditions, but also during competitions. Since the swimmer's volume is reconstructed, more complete analyses are possible, also in terms of drag quantification.

Further studies should concentrate on the investigation of a deformable arm model, as well as an hybrid approach in which a small set of skin features are tracked and employed for pose estimation, in order to provide more robust and accurate estimation of all three joint trajectories and trunk position.

# 4 COMPARISON OF MARKERLESS AND MARKER-BASED MOTION CAPTURE TECHNOLOGIES THROUGH SIMULTANEOUS DATA COLLECTION DURING GAIT

## *4.1 BACKGROUND*

Gait analysis is the systematic study of human walking, using the eye and the brain of experienced observers, augmented by instrumentation for measuring body movements, body mechanics, and the activity of the muscles [55]. In actual practice, gait analysis is used in orthopedic hospitals and clinics to diagnose pathologies, design surgical operations, plan treatments for individuals with conditions affecting their ability to walk. One of the conditions for which it has been employed extensively, and has proven its efficacy, is cerebral palsy in children [56][57][58]. Other conditions that are amply present in literature include Parkinson's disease [59], multiple sclerosis [60], Paget's disease [61], persons that have suffered anterior cruciate ligament injuries [62], even patients whose ambulatory deficiency is secondary to systemic diseases, such as diabetes [63]. Analysis of gait generally involves the investigation of different aspects of movement, such as kinematics, kinetics, motor control. In the following, we will only focus on kinematics, i.e. the description of the motion of body segments. We will neglect the analysis of the forces that generate this motion (kinetics), and the physiological processes through which the human body performs the desired motion (motor control, muscle activation), as they are independent of the motion capture technology adopted, which is the focus of this thesis.

In this context, an exhaustive description of the experiences that led to the diffusion of commercial marker-based stereophotogrammetric systems for kinematics evaluation in clinical gait analysis has been written recently by Sutherland [64]. Since the establishment of this technology, many protocols for data collection (marker placement, number of trials) and techniques for data analysis (biomechanical model, pose estimation) have been proposed. The

versatility of such instrument for kinematics reconstruction however soon called for an effort for the standardization of nomenclature and kinematics variables' definition, as well as for an advancement in estimation reliability [65][66]. Recommendations and guidelines have been compiled by eminent researchers and by international societies such as the International Society of Biomechanics for the definition of segment's anatomical frame of reference from identifiable anatomical landmarks [67][68], and for the description of joint angular movement [68]. On general terms, longitudinal (or vertical), medio-lateral and anterior-posterior axes are defined for each segment; relative orientation of two segments is described, according to Grood and Suntay's convention [69], by Cardan angle decomposition of the rotation matrix that represents the relative orientation between the two segments. In this way, a "joint coordinate system" can be defined for each joint: rotations around its axes can be expressed in clinically relevant terms and easily interpreted. For example, for the knee joint, the three angles of flexion-extension (on the sagittal plane), ab-adduction (frontal plane) and internal-external rotation (transverse plane) can be defined.

The widespread adoption of stereophotogrammetric technology lead researcher to investigate and quantify their reliability regarding both reconstruction of marker trajectories [70] and subsequent estimation of kinematic variables. In order to obtain repeatable and significant results, not only is the accurate identification of anatomical landmarks needed [71], but also the relative movement that can occur between a marker and the underlying bone must be taken into consideration [72]. Even though no ultimate solution has been found to overcome this last problem, different methods have been proposed for convenient marker placement and optimal pose estimation. Two main markerset typologies can be identified: on one hand, markers can be placed directly on landmarks, as well as mounted on wands secured to the segment, and not directly on the skin [73]; on the other, clusters of technical markers [67], placed on the body segment as to minimize the artifacts due to soft tissue [74], can be used. Pose estimation has also evolved from

geometrical methods to least-square optimal estimators. Furthermore, some techniques have been proposed for soft tissue artifacts compensation, such as the "multiple calibration" procedure [75], in which calibration of anatomical landmarks relative to the technical cluster is repeated for different body postures, and some interpolation scheme is devised to estimate the instantaneous cluster model configuration, or the "global optimization" approach [76].

In the last few years, a growing interest has been shown by the biomechanics community in novel markerless technologies, developed mainly in the fields of computer vision and for the entertainment industry. The advantages that such methods could provide to the gait analysis field would be mainly the reduction in preparation time of the subjects and to the absence of markers that could modify the naturalness of a subject's movement. As is common to all new technologies, there is still the need for validation and standardization of the biomechanical models they comprise. Some efforts in this sense are made by the computer vision community, with the creation of public datasets that include marker data (HumanEva [26]). By admission of the authors however, recommendations for marker placement have not been strictly followed, as markers were attached to loose-fitting clothes. Furthermore, the conventional marker set that has been employed provides less repeatable results than cluster-based marker sets with 6-degrees-of-freedom biomechanical models. Finally, evaluation of algorithms' performance has been made from errors in joint centers ("virtual markers") position, which do not represent the convention in clinical use. Description of functional joint angles, based on the precise anatomy of the subject and consistent with biomechanical societies' recommendations, has been so far neglected by markerless systems' developers; nonetheless, it is essential for the application of the latter in the clinical field.

In this thesis, a procedure has been investigated for comparison of a state-of-the-art marker-based technique and a silhouette-based markerless approach on lower limb joint angle estimation. Data has been acquired simultaneously with a

commercial stereophotogrammetric system, saving to file the grayscale videos that are used for reconstruction of markers' 3D trajectory. In order to calculate functional joint angles with the markerless technique, technical frames of reference of relevant segments have been registered to anatomically-based ones.

The present chapter is organized as follows: a description of the experimental setup is given first, detailing the motion capture system and the related equipment. The way that data thus acquired is processed by the markerless approach is then explained. Marker-based analysis is also described in detail, along with the registration procedure that has been developed to perform the comparison with the markerless technique. Finally, some experimental results are reported and discussed.

## *4.2 EXPERIMENTAL SETUP*

An 8-camera SMART-D stereophotrogrammetric optoelectronic system (BTS S.p.A) was employed to acquire experimental data. Acquisition rate was set to 200Hz for marker data and 100 Hz for image data (one image frame every two was saved to file). Resolution of the CCD digital cameras was 640x480 pixels. Infrared illuminators, that emit light impulses of 880 nm wavelength, were mounted around each camera; infrared filters were placed in the optics of the latter. Calibration was performed following manufacturer's recommendations (Thor2 calibration system, *http://www.btsbioengineering.com/BTSBioengineering/Kinematics/BTSSMARTD/BTS_SMARTD.html)*: a rigid wand on which three markers are mounted (Figure 4-1) is swept through the volume of interest, in a dynamic acquisition, for simultaneous calibration of intrinsic parameters, and relative position, of the video-cameras.

**Figure 4-1 Rigid wand with three markers, employed for intrinsic calibration of the cameras**

A three-axes calibration grid (Figure 4-2) was placed on the ground and acquired for determination of the global frame of reference.

**Figure 4-2 Calibration grid with 9 markers, employed for extrinsic calibration of the cameras**

Position of the eight cameras is reported in Figure 4-3.

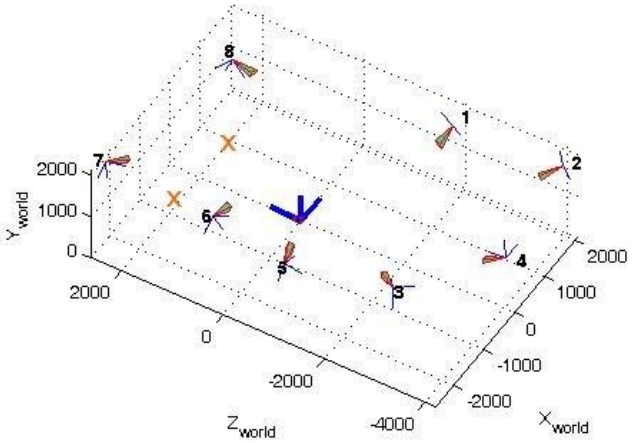



**Figure 4-3 Cameras' position as resulting from extrinsic calibration (coordinates in mm). Orange crosses correspond to additional infra-red illuminators**

Two additional infra-red illuminators were placed closed to the ground in positions indicated by the orange crosses, as to increase contrast between the subject and the floor. A pointer, on which two markers are mounted at known distance, is used for anatomical calibration (Figure 4-5).

**Figure 4-4 Retroreflective passive marker**

Forty-eight 10-mm-diameter spherical markers (Figure 4-4) have been used; they were attached to the subject's skin or thigh-fitting clothes by means of special bi-adhesive tape.

**Figure 4-5 Pointer used for anatomical calibration**

*Smart Tracker* software provided by the manufacturer was employed for marker tracks' reconstruction and labeling. A toolbox for Matlab (Mathworks, Inc.) was also provided for extraction from data files of calibration parameters and video data, to be input to the markerless system.

## *4.3 DATA PROCESSING*

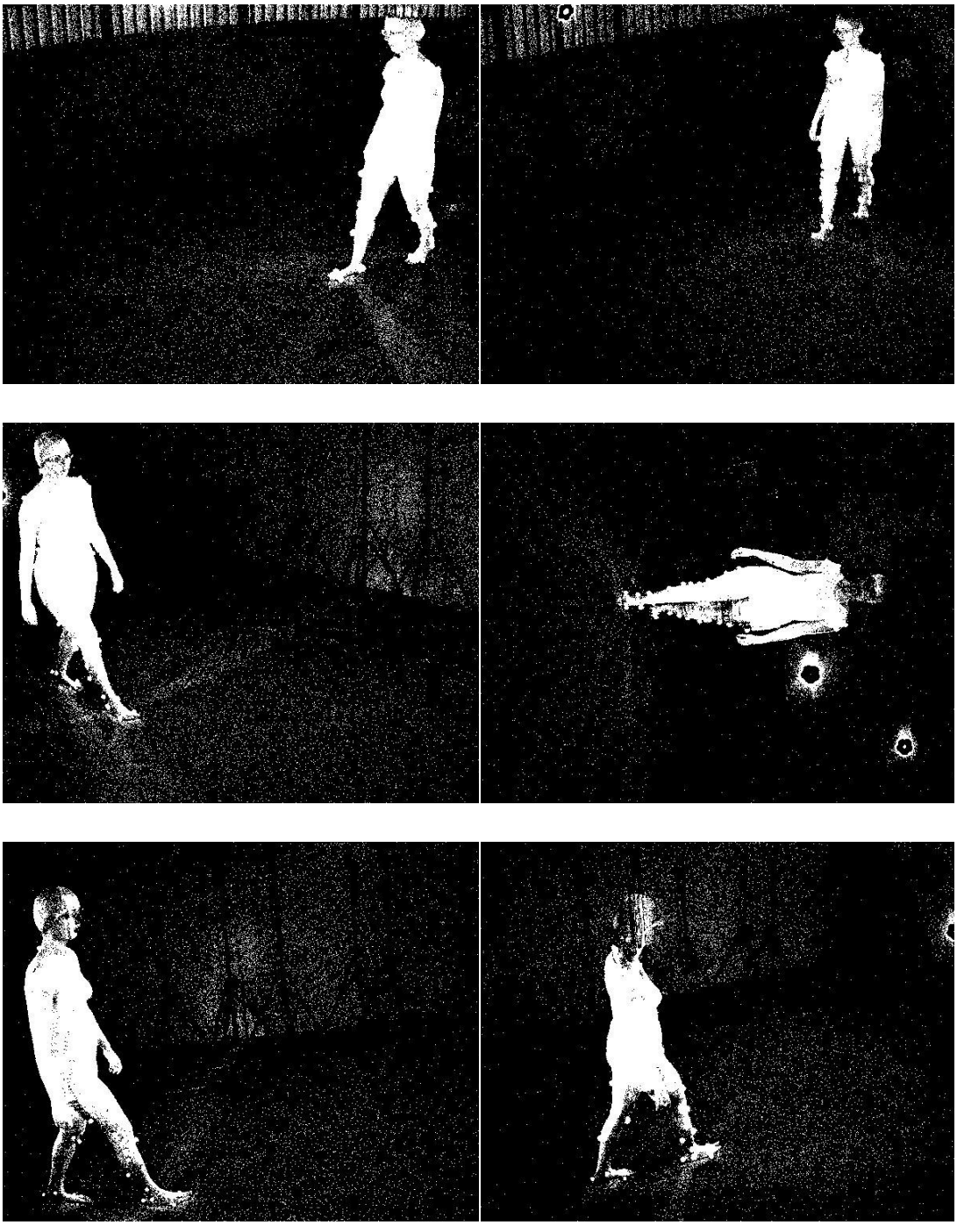
### Markerless

The present section includes the various steps through which 3D segmental kinematics is extracted, through markerless analysis, from grayscale videos acquired with a commercial motion capture system. An example of acquired frame from multiple cameras is shown in Figure 4-6.

**Figure 4-6 Frame from one of the acquired video sequences (infra-red grayscale cameras)**

*Background subtraction*

Foreground/background segmentation is in general more difficult for grayscale images, since the chromatic component of an image has often most discriminative power than sole luminance. We try to take advantage however of the fact that this type of images represent, at each pixel, the intensity of light in the sole infrared band of the electromagnetic spectrum; we assume that the subject's skin and clothes will present greater response to infrared illumination than the surrounding background scene. The foreground is extracted by simple subtraction of a reference background image from each frame of the video sequence. Pixels of this "distance image" that present values greater than a fixed threshold (usually set in the 2 - 4 range) are set as foreground (Figure 4-7). The main problems within this technology and this approach lie in the limited robustness to the presence of shadows cast by the subject on the floor and on themselves.



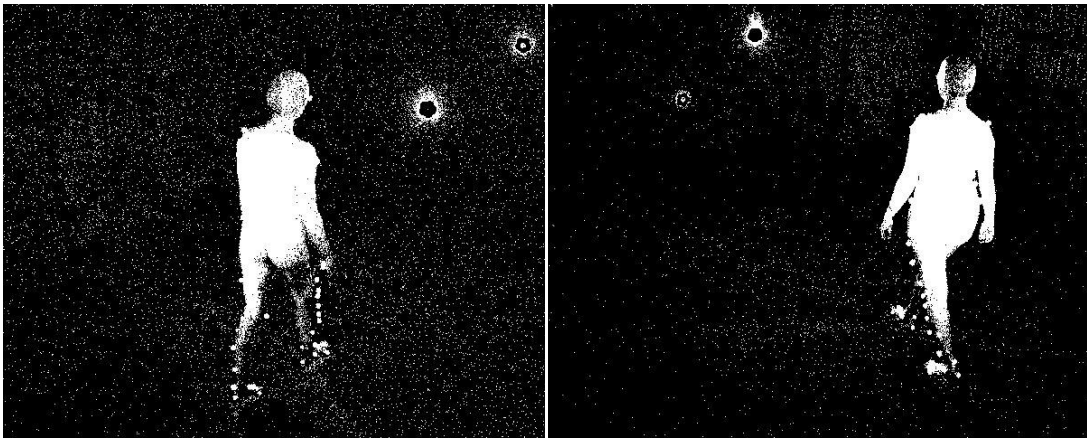


**Figure 4-7 Example of background subtraction for the frames in Figure 4-6**

*Visual hull creation*

Visual hulls are created from silhouettes as indicated in section 2.3. Since the calibration parameters provided by the acquisition system are consistent with the definition proposed in section 3.2, calibration functions from OpenCV mentioned in 3.3 can be used in this context as well. The volume of interest has been partitioned into voxels of size 0.01 m; the same algorithms described in section 3.3 are applied. An example of a visual hull reconstructed for gait analysis is shown in Figure 4-8.

**Figure 4-8 Visual hull reconstructed from a frame of an acquired gait trial**

*Model definition*

The adopted model definition does not depart from that described in Figure 2-6. The kinematic relationship between the segments is clarified by Figure 4-9; the pelvis is chosen as root of the kinematic tree.

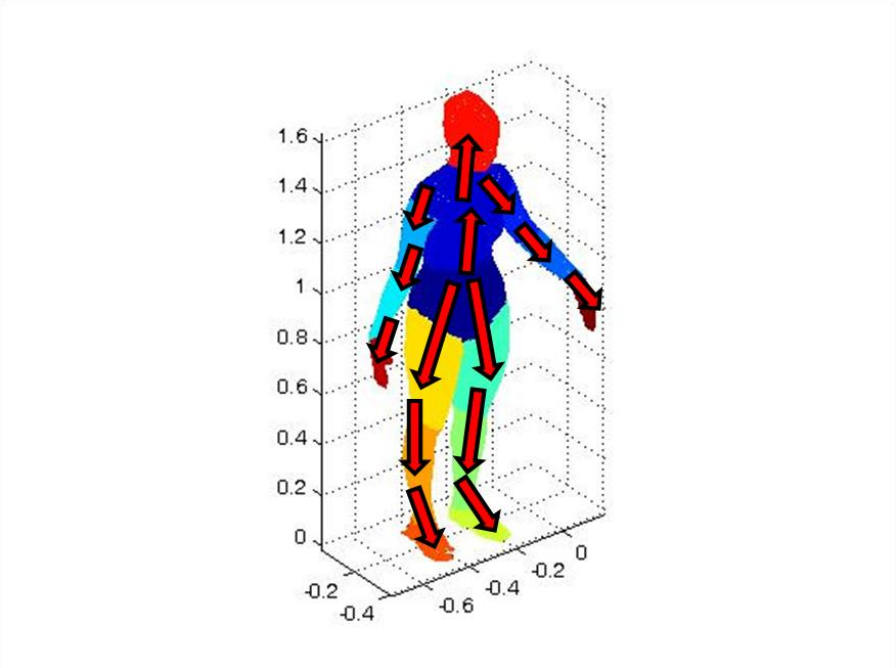



**Figure 4-9 Scheme of the model kinematic tree: each arrow point from "parent" segment to "child" segment**

A slight modification has been introduced in the definition of segments' frames of reference, so that they could resemble more closely the anatomically-based ones employed for marker-based analyses (see following paragraph). The longitudinal (Y) axis for thigh and shank segments is the axis connecting their parent and child joint centers. Antero-posterior (X) axis is the component of global antero-posterior axis that is orthogonal to Y axis, and Z axis is perpendicular to the others. For pelvis segment, axes are parallel to global ones. Foot segments' axes instead have been based on a principal component analysis (PCA) of the relative vertices on the mesh (Figure 4-10).

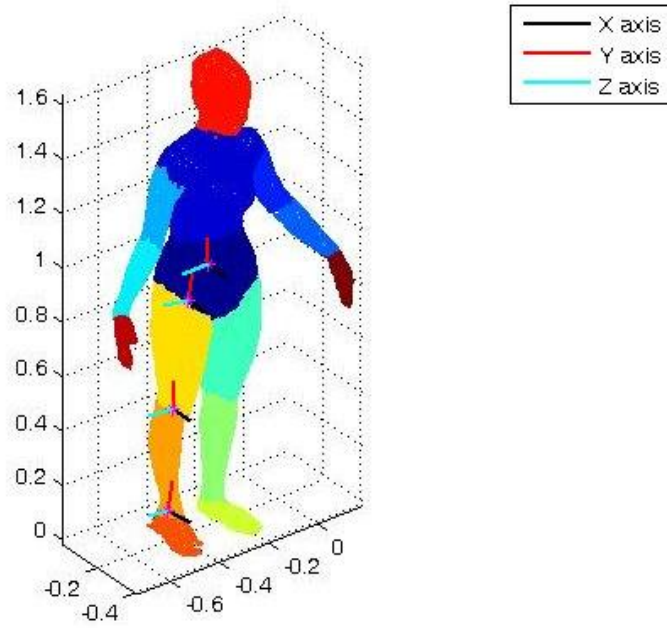



**Figure 4-10 Orientation of markerless technical frames for the right leg**

The sequence of visual hulls is tracked employing the articulated-ICP algorithm described in section 2.5 with a data-to-model approach (visual example shown in Figure 4-11); the position and orientation ) in space of each body segment's embedded frame of reference are obtained.

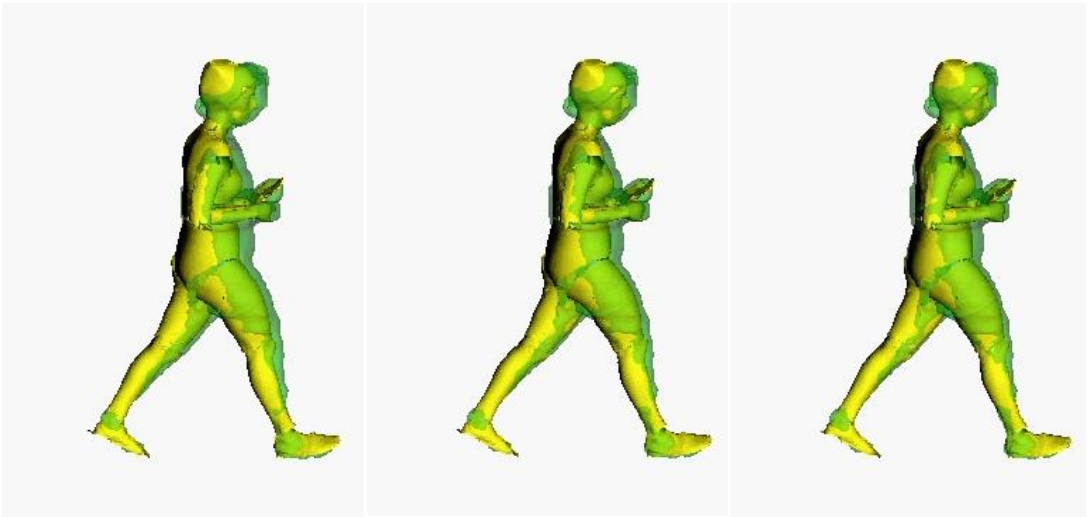


**Figure 4-11 Example of nine iterations of the articulated-ICP algorithm for matching of the model (yellow) to visual hull data (green)**

## Markerbased

The protocol that has been employed as gold standard for segmental kinematics' measurement is a modified version of the IORgait protocol proposed by Istituti Ortopedici Rizzoli [77], and has been actively used in clinical studies [78][79].

Anatomical calibration of some anatomical landmarks, either using a pointer or directly a marker, with respect to technical frames is involved.

*Marker placement*

Retroreflective markers are placed on the skin of the subject in the preparatory phase of the acquisition. Some of these markers are only needed in the initial calibration phase, and could be removed prior to the beginning of the walking trials' acquisition, while others need to be securely attached to the skin and remain in place for the whole duration of the experiment. The list of all positioned markers on each body segment is the following:

- *Trunk*
    - spinous process of C7 cervical vertebra (C7)
    - left acromion process(LA)
    - right acromion process (RA)
    - spinous process of L5 lumbar vertebra (L5)
- *Pelvis*
    - cluster of 4 markers on left side of pelvis
    - cluster of 4 markers on right side of pelvis
- *Thigh*
    - cluster of 4 markers, placed laterally
    - prominence of the lateral epicondyle (LE)
    - prominence of the medial epicondyle (ME)
- *Shank*
    - Cluster of 4 markers, placed laterally
    - proximal tip of the head of the fibula (HF)
    - most anterior border of the tibial tuberosity (TT)
    - prominences of the lateral (LM) and medial (MM) malleolus
    - prominences of the lateral (LM) and medial (MM) malleolus
- *Foot*

- aspect of the Achilles tendon insertion on the calcaneous (CA)
- dorsal margin of the first metatarsal head (FM)
- dorsal margin of the second metatarsal head (SM)
- dorsal margin of the fifth metatarsal head (VM).

Clusters are quasi-planar, and their longest principal axis is oriented toward the relevant anatomical landmarks [74].

*Anatomical calibration*

Anatomical landmarks on pelvis, thighs and shanks are calibrated with respect to the corresponding cluster of technical markers. This means that, defining a technical system of reference on the cluster , the position of the anatomical landmark in this system of reference is found:

4.1

The position of the landmark in the global frame of reference, , can be given by either the position of a marker placed directly on it (in this case, the landmark is assumed to coincide with the center of the spherical marker), or by the reconstructed position of the tip of the pointer.

Anatomical calibration is performed as follows:

- *pelvis:* the two most anterior and the two most posterior margins of the iliac spines (ASIS, PSIS) are calibrated through the use of the pointer on the corresponding side (left/right) cluster
- *thigh:* the most lateral prominence of the great trochanter (GT) is calibrated through the use of the pointer; the epicondyles (ME, LE) are calibrated as markers in a static acquisition
- *shank:* all four landmarks (HF, TT, LM, MM) are calibrated as markers in a static acquisition.

*Marker tracking and anatomical landmarks' reconstruction*

During a walking trial, the position of markers is measured, then reconstructed by the stereophotogrammetric system (Figure 4-12).

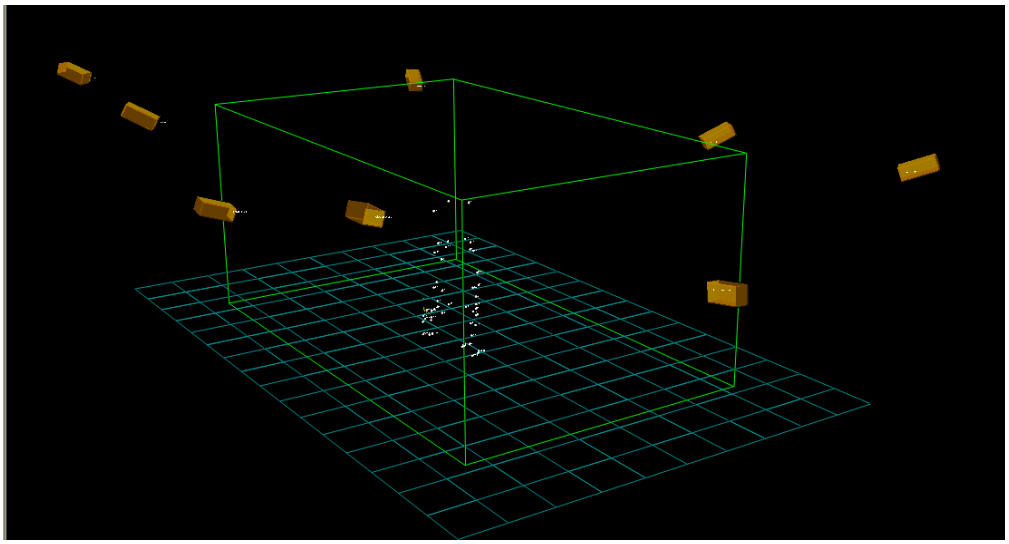


**Figure 4-12 Markers' position reconstructed in 3D space by the stereophotogrammetric system. Position of the cameras is also shown.**

In order to extract skeletal kinematics, however, each reconstructed track must be associated to the actual marker that generated it. The marker model is therefore defined in the tracking software to facilitate this labeling procedure: markers that have not been labeled are signaled, and desired topological connections between markers are shown for easier identification (Figure 4-13).

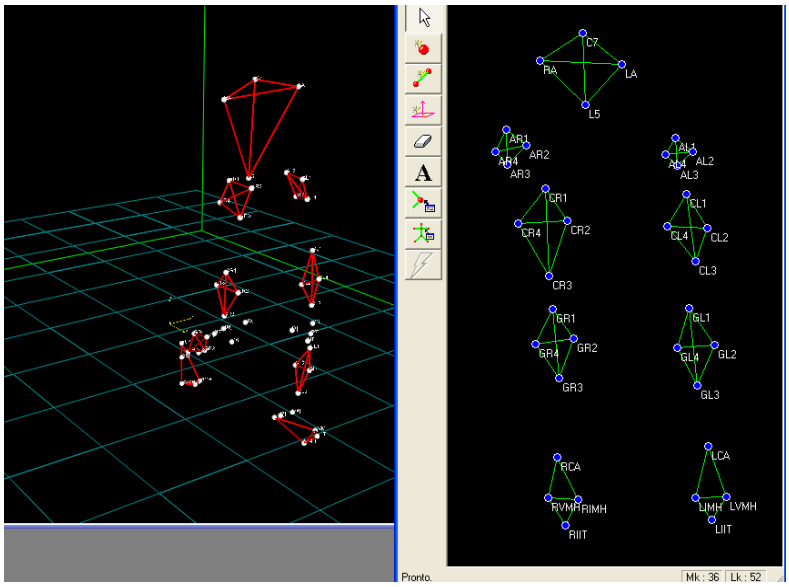



**Figure 4-13 Labeling of reconstructed markers (left) according to the marker model (right)**

As a result, trajectories of anatomical landmarks of trunk and feet, and of cluster points on pelvis and legs, are obtained. The pose of each cluster's embedded frame of reference is estimated at each frame through an optimal least-square procedure based on SVD decomposition of cross dispersion matrix [80],[74]. Knowledge of a segment's pose allows to determine the instantaneous position of all relative anatomical landmarks, as

4.2

An example of all anatomical landmarks reconstructed for one frame, and of the relative modified IORgait full-body protocol, is shown in Figure 4-14.

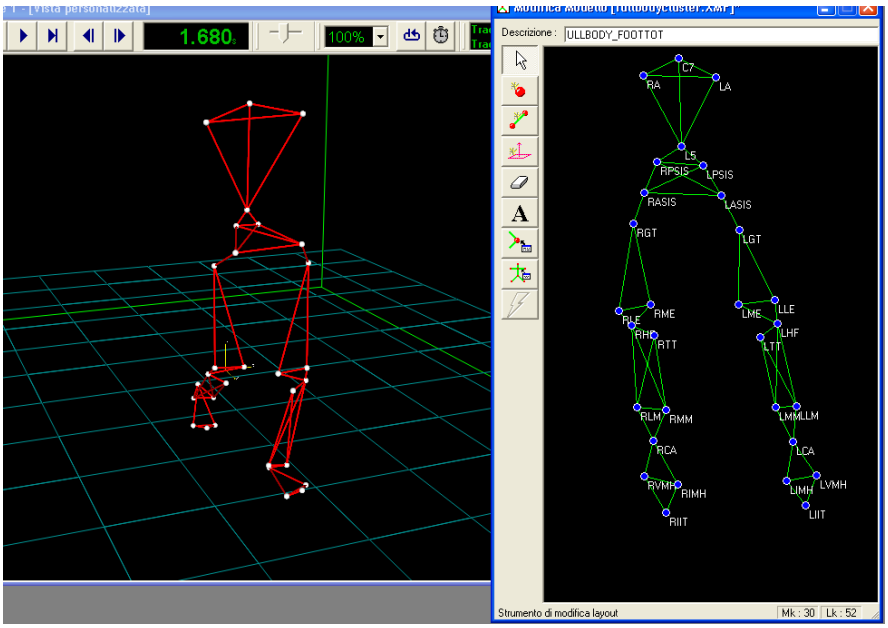



**Figure 4-14 Reconstructed anatomical landmarks' position (left) according to the kinematic model of the protocol (right)**

*Joint angles calculation*

Anatomical frames of reference defined in the IORgait protocol are based on the proposal by Cappozzo [67] and mostly consistent with relevant international recommendations [68]. Only lower limbs' kinematics will be considered in the following, as it is the established clinical practice [66]. A full description of frames' origin ( ) and axes is provided here for clarity.

- *Pelvis:*
  - · : The origin is at the midpoint between the ASISs.
  - · : The line connecting the right and left ASISs, and pointing to the right.
  - · : The line parallel to a line lying in the plane defined by the two ASISs and the midpoint of the two PSISs, orthogonal to axis, and pointing anteriorly.
  - · : The line perpendicular to both and , pointing cranially.
- *Thigh:*

· : The origin is located at the the midpoint between the lateral and medial epicondyles (LE and ME).

· : The line joining the origin and the hip center of rotation (HJC), and pointing cranially. HJC is assumed to coincide with the centre of the acetabulum, which is reconstructed by a geometrical prediction method based on the location of ASISs and PSISs [81].

· : The line perpendicular to the axis, lying in the quasi-frontal plane defined by the axis and the two epicondyles, pointing to the right.

· : The line perpendicular to both and axis, pointing anteriorly.

- Shank:

· : The origin is located at the midpoint of the line joining the lower ends of the malleoli (MM and LM).

· : The malleoli and the head of the fibula landmarks (HF) define a plane which is quasi-frontal. A quasisagittal plane, orthogonal to the quasi-frontal plane, is defined by the midpoint between the malleoli and the tibia1 tuberosity (IT). The axis is defined by the intersection between the above-mentioned planes with its positive direction proximal.

· : The line that is perpendicular to , and lies in the quasi-frontal plane; its positive direction is from left to right.

· : The line perpendicular to both and axis, pointing anteriorly.

- Foot

· : The origin is located at the calcaneus landmark (CA).

· : The calcaneus and the first and fifth metatarsal heads (FM and VM) define a plane which is quasitransverse. A quasi-sagittal plane, orthogonal to this latter plane, is defined by the calcaneus

landmark and the second metatarsal head (SM). The axis is defined by the intersection of these two planes and points anteriorly.

· : The line that lies in the quasi-transverse plane and is perpendicular to ; its positive direction is from left to right.

· : The line perpendicular to both and axis; its positive direction is dorsal.

From the pose of anatomical frames of reference, joint angles are calculated to describe the relative orientation of two adjacent body segments. The "Grood&Suntay" angle decomposition is followed [69], and three non-orthogonal rotation axes are identified, which form a "joint coordinate system". For hip (pelvis – thigh joint) and knee (thigh-shank joint), rotations are defined as follows (first name denotes positive sign):

- flexion/extension (Flex/Ext) is the relative rotation about axis e1, taken as the medio-lateral axis (Z) of the proximal segment
- internal/external (Intra/Extra) is the relative rotation about the vertical axis (Y) of the distal segment (axis e3)
- abduction/adduction (Abd/Add) is the relative rotation about a "floating" axis (e2) orthogonal to the previous two, which points anteriorly.

For ankle joint, nomenclature is as follows:

- dorsiflexion/plantarflexion (Dors/Plan) is the relative rotation about the medio-lateral axis (Z) of the shank (axis e1)
- internal/external rotation (Int/Ext) is the relative rotation about the vertical axis (Y) of the foot (axis e3)
- inversion/eversion (Inv/Ev) is the relative rotation about the "floating" axis (e2) orthogonal to the previous two, which points anteriorly.

### Registration between the two systems

In order to compare the performance of markerless and marker-based pose estimation systems in terms of clinically relevant joint angles, the same anatomical frames of reference must be defined for both systems. The procedure that has been followed exploits the anatomical calibration performed in the marker protocol to give a precise anatomical connotation to the markerless model. This procedure requires the tracking with both systems of a static acquisition of the subject (Figure 4-15).

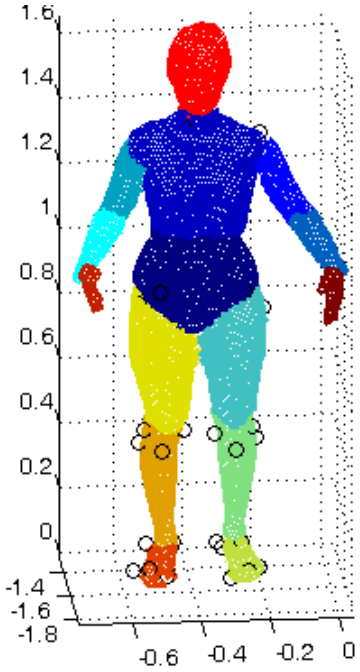



**Figure 4-15 Static acquisition tracked with the markerless system and contemporary marker-based reconstruction of anatomical landmarks' position (black circles)**

Pose estimation of segments' embedded frames of reference is then obtained and compared: for a segment , the relationship between its pose estimated with marker-based and markerless techniques can be represented by a 4x4 transformation matrix

4.3

Assuming that this relationship is only representative of the difference in technical frame definitions, we exploit it in dynamic acquisitions to obtain an estimate of the marker-based technical frame from the markerless estimation

4.4

This allows to obtain the global position of relative anatomical landmarks from the pose estimated with markerless:

4.5

In this context markers have been exploited for calibration, but not for pose estimation during the dynamic acquisition. Segmental kinematics based on the anatomical landmarks' trajectories thus reconstructed can be compared with those obtained through marker-based tracking of the dynamic acquisition:

Since no cluster of markers has been placed on the feet, calibration of anatomical landmarks has been performed on the anatomical frame of reference in the static acquisition.

An example of the position of anatomical landmarks reconstructed with markerless and marker-based techniques is shown in .

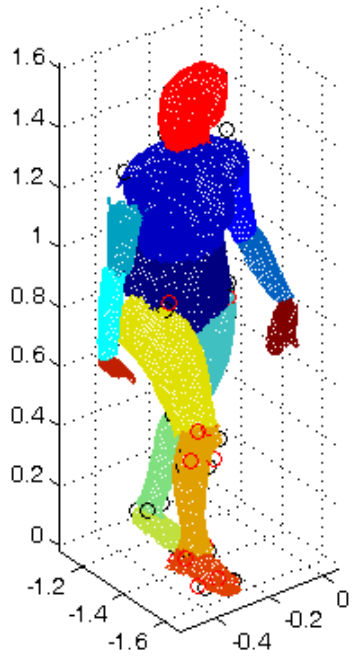



**Figure 4-16 Example of subject's pose estimated with the markerless system. Anatomical landmarks reconstructed for the right leg with markerless (red circles) and marker-based (black circles) techniques are shown.**

## *4.4 EXPERIMENTAL RESULTS*

### Experiment details

In order to test the applicability of the novel procedure for co-registration and comparison of markerless and marker-based gait analysis techniques, a healthy subject (female, age 26, BMI 20.9) was recruited. A laser scan of the subject was available, and was adopted as input for model creation in the automatic model generation procedure. The subject was wearing tight fitting clothes and a swim cap, and barefoot. After the pointer-based anatomical calibration acquisitions, video capturing from the infra-red cameras was activated. The subject was first asked to stand still in a reference (orthostatic) position, then to perform fifteen walking trials. Video acquisition of the sole background was also performed.

The position and orientation of all cameras during this experiment are shown in Figure 4-3. While placing the cameras, several requirements had to be taken into consideration. For markerless analysis, the cameras need to view the whole subject at all times, and from as most complementary views as possible. On the other hand, for markers' reconstruction, each of them must be in view in at least two cameras, so multiple cameras should be placed to each side of the subject. The resulting configuration is a compromise between these different demands. Cameras 4 (frontal view) and 5 (sagittal view) are mainly dedicated to markerless analysis, while the others, though still useful for visual hull reconstruction, were placed according to recommendations for marker visibility. Only cameras 1-6 however were used for visual hull creation, because background subtraction for cameras 7-8 (Figure 4-7) was too inaccurate after the first few frames, probably because of the their greater distance from the subject.

A gait cycle for the right leg was analyzed from each recorded trial. Joint angles calculated with marker-based and markerless technique after co-registration of technical systems were compared, and the difference was evaluated in terms of root mean squared distance (RMSD).

Comparison of joint angles calculated with the markerless technique with and without the co-registration with marker-based system is also shown.

**Results**

Hip, knee and ankle joint angles calculated with marker-based and markerless technique are shown respectively in Figure 4-17, Figure 4-18, and Figure 4-19.

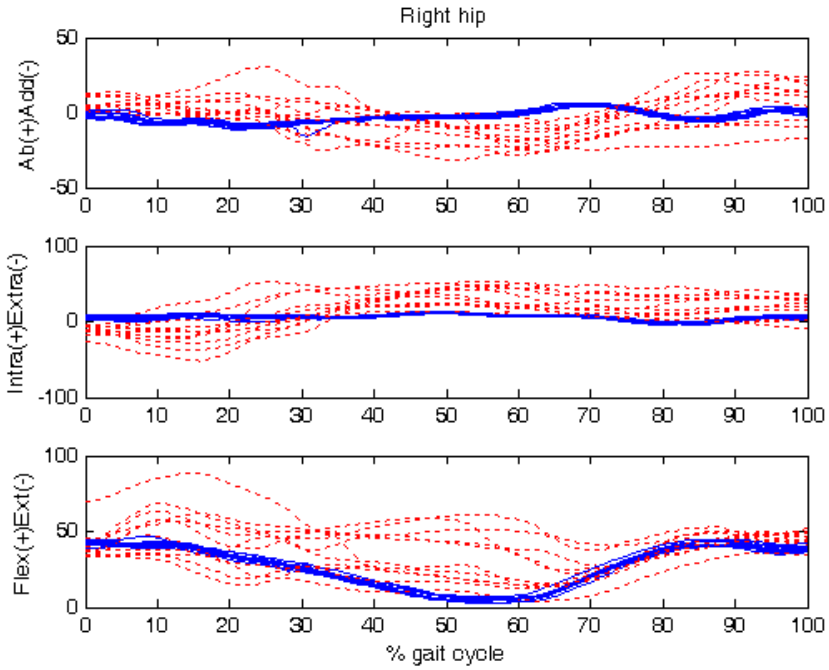



**Figure 4-17 Hip joint angles (degrees) estimated in 15 trials with marker-based (blue solid lines) and markerless (red dotted lines) techniques**

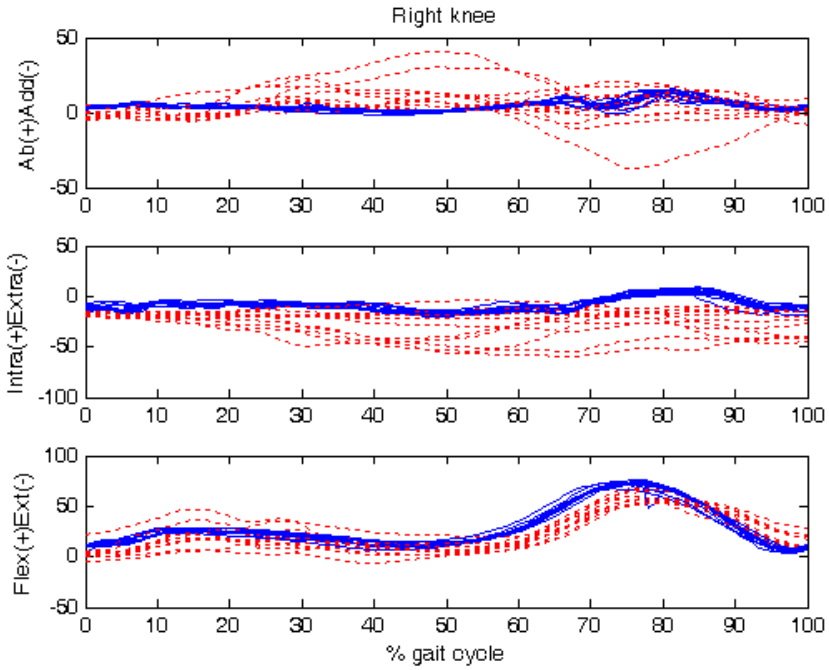



**Figure 4-18 Knee joint angles (degrees) estimated in 15 trials with marker-based (blue solid lines) and markerless (red dotted lines) techniques**

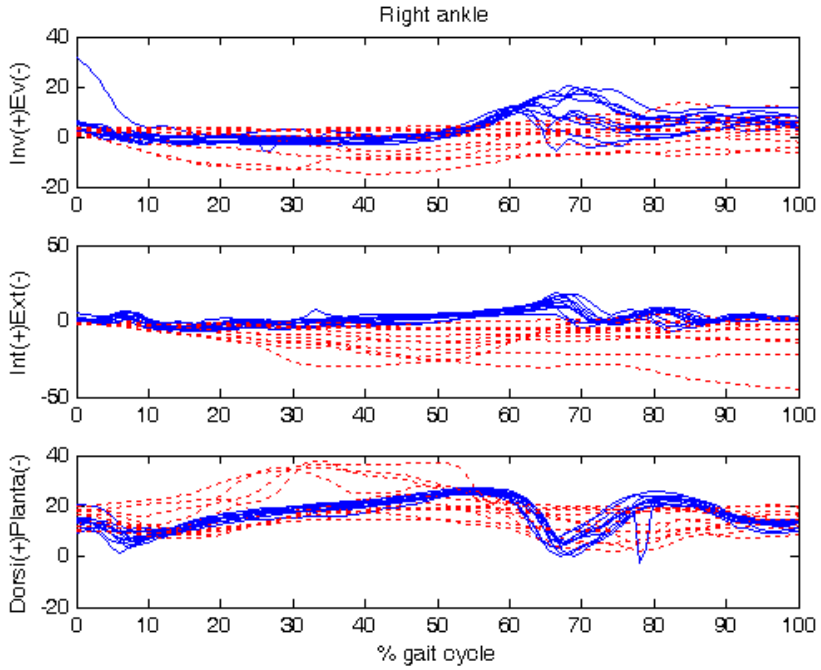



**Figure 4-19 Ankle joint angles (degrees) estimated in 15 trials with marker-based (blue solid lines) and markerless (red dotted lines) techniques**

A description of RMSD values is reported in **Errore. L'origine riferimento non è stata trovata.**. It comprises mean, standard deviation, minimum and maximum values calculated over all trials. Furthermore, mean RMSD values have been normalized to the corresponding range of motion calculated on a marker-based trial: the result is shown in the rightmost column of the table. Best results are achieved for knee flexion-extension angle, with an average RMSD error, 11.75 deg, corresponding to 18.35% of the range of motion. Motion on the sagittal plane is estimated better than on the other planes also for hip and ankle, with respectively 44.66% (17.62 deg) and 33.12% (7.17 deg) errors. Estimates for hip joint are the most incorrect; error on the transverse plane reaches 129.57% of the relative marker-based range of motion.

| | | Mean (deg) | STD (deg) | Min (deg) | Max (deg) | % range of motion |
|---|---|---|---|---|---|---|
| Hip | AbAdd | 14.1178 | 2.2859 | 9.5582 | 17.2638 | 91.5442 |
| | IntraExtra | 21.6101 | 9.2801 | 7.7223 | 34.6567 | 129.5661 |
| | FlexExt | 17.6177 | 8.4629 | 6.3787 | 28.9840 | 44.6654 |
| Knee | AbAdd | 9.9800 | 5.6696 | 4.0694 | 20.8053 | 73.1014 |
| | IntraExtra | 19.7928 | 9.4778 | 8.4926 | 35.7572 | 71.1698 |
| | FlexExt | 11.7473 | 2.4911 | 8.1255 | 15.9701 | 18.3551 |
| Ankle | InvEv | 7.0291 | 3.6313 | 3.3514 | 13.7928 | 54.7713 |
| | ExtInt | 12.8814 | 6.9739 | 5.3425 | 29.4322 | 88.2306 |
| | DorsiPlanta | 7.1715 | 1.8331 | 5.0275 | 11.0305 | 33.1221 |

**Table 4-1 Description of RMSD values between angles calculated with the two techniques. Average, standard deviation, minimum and maximum average values are reported. The average RMSD value is also given as percentage of the corresponding range of motion**

An example of joint angles calculated with the markerless technique with and without the co-registration with marker-based system is shown in Figure 4-20, Figure 4-21, and Figure 4-22. Similar behaviors are demonstrated for flexion/extension angles, but an offset is generally present. Cross-talk is also evident between flexion/extension and ab/adduction angles.

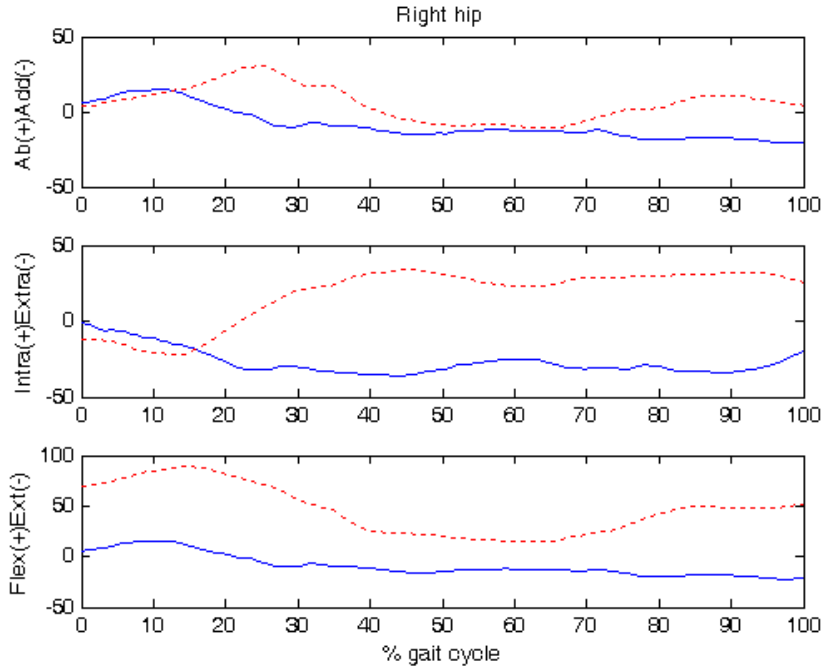



**Figure 4-20 Hip angles (degrees) for one gait trial calculated with markerless technique with (red dotted line) and without (blue line) the co-registration with marker-based system**

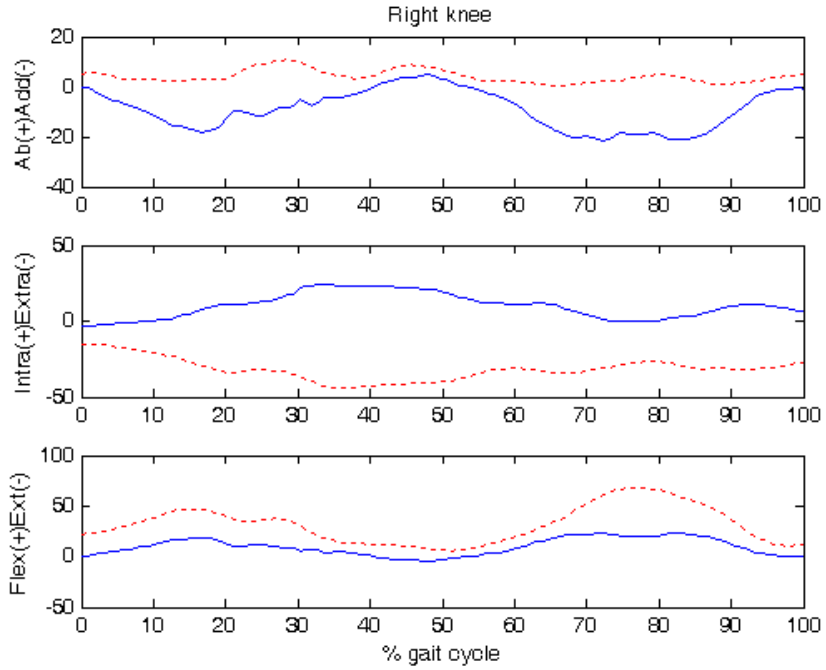



**Figure 4-21 Knee angles (degrees) for one gait trial calculated with markerless technique with (red dotted line) and without (blue line) the co-registration with marker-based system**

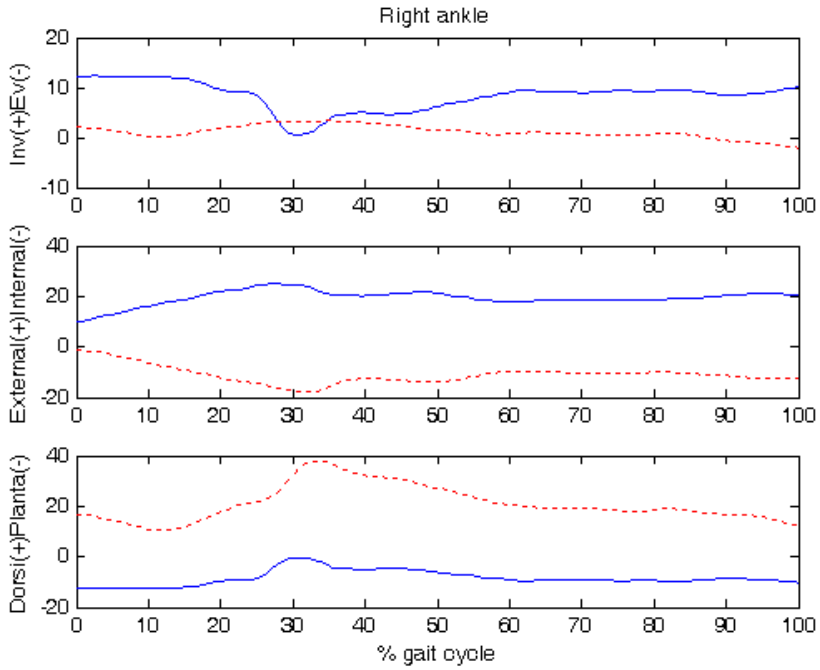



**Figure 4-22 Ankle angles (degrees) for one gait trial calculated with markerless technique with (red dotted line) and without (blue line) the co-registration with marker-based system**

## *4.5 DISCUSSION*

In the present chapter, a procedure for comparison of markerless and state-of-the-art marker-based techniques in joint angles' estimation has been presented. The actual results obtained from a gait analysis experiment seem to indicate that this markerless technology is not ready for clinical applications, as the level of accuracy and robustness is insufficient. The main problems identified within the markerless analysis were: excessive "phantom volume" artifacts at the level of the pelvis (an example is shown in Figure 4-23), that lead to mostly incorrect estimates for hip joint angles; an incorrect background subtraction that excludes the lower part of the leg from visual hull creation, caused by self-shadowing of that area in mistance-midswing phase; furthermore, rigid-body matching of the model surface of the foot to the highly deforming foot in the visual hull (caused mainly by midfoot-forefoot flexion) can easily yield to an estimate of ankle joint

angles which is different from that based on markers (which are not applied on the forefoot).

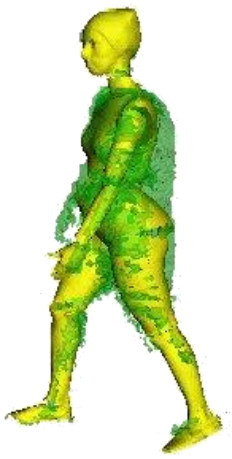


**Figure 4-23 Excessive pelvic obliquity in the matched model (yellow) caused by phantom volumes in the visual hull (green)**

Additional factors, intrinsic to the comparative nature of this experiment, may have affected the results of markerless estimation. For instance, camera placement was conditioned by the requirements imposed by markers' visibility. Furthermore, the presence of the markers attached to the skin of the subject deformed the silhouettes and consequently the visual hulls.

Nevertheless, the presented method can be used for further testing and developing of silhouette-based markerless techniques. Its main advantage is the possibility to use state-of-the-art marker-based data as gold standard, without any difference in calibration and identification of anatomical landmarks. The ability to perform different types of analysis with the same commercial system could be of use to gait laboratories, which could choose between one system or the other (or an hybrid version), according to the desired level of accuracy and acceptable patient discomfort.

# 5 CONCLUSIONS

In the present thesis, new frontiers of the application of markerless motion capture techniques have been studied. These techniques are only based on synchronized video sequences of the subject and allow noninvasive measurements of body segments kinematics. The main advantage over commercial marker-based motion capture systems is that they do not rely on skin based markers for tracking, capturing the subject's motion in a natural and easier way. Furthermore, markerless analysis can be performed in non-controlled environments where markers cannot be employed. For the latter reason, markerless kinematic analysis of swimmers has been investigated. Since motion capture systems based on retro-reflective skin markers cannot be employed underwater, the state of the art of swim kinematics analysis requires the manual digitization of feature points on video sequences of the subject. This is a time consuming task, which has often limited the analyses to the sagittal plane. In this context, an automatic markerless motion capture system was tailored to operate properly underwater. Common out-of-shelf subaqueous cameras were employed. Main modifications with respect to the systems presented in literature for laboratory acquisitions regard the calibration procedure and the advanced image analysis algorithms that were employed. The kinematic properties of the markerless model were specialized for the analysis of arm kinematics. Upper arm 3D kinematics during a front crawl stroke was reconstructed in terms of shoulder, elbow and wrist joint center trajectories. Only the phases where the arm is completely immersed in the water could be analyzed with the proposed technique. Joint trajectories estimation accuracy was evaluated in terms of RMSD with respect to trajectories obtained with a conventional 3D reconstruction technique, implemented in commercially available software. Accuracy is deemed to be sufficient for wrist joint, which is generally considered the most important for the technical analysis of the stroke. Larger errors occur for shoulder and elbow joints. A discussion of current limitations in the technique and in the comparison procedure has been provided. Further developments are suggested, which could render the technique robust and suitable for wide-scale studies of swimmers' motion.

On the other hand, in the clinical field, the application of markerless techniques is still debated. This is mainly due to a limited number of papers dedicated to the comparison with the state of the art of marker based motion capture, especially in term of repeatability and accuracy in the estimation of the three dimensional joint rotations. So far in the present thesis the application of markerless techniques to data acquired with a marker-based system has been investigated. The latter was employed to directly compare the joint kinematics estimated with the markerless technique with the one determined with traditional marker based systems, while excluding differences in calibration and synchronization of the cameras. In this context for marker placement the CAST protocol was adopted, which could be considered the most reliable in term of accuracy for joints rotation estimation. The actual results obtained from a gait analysis experiment seem to indicate that this markerless technology is not ready for clinical applications, as the level of accuracy and robustness is insufficient. The main problems that affect this technique were however identified and described, along with possible limitations in the experimental setup. Nevertheless, the presented method can be used for further testing and developing of silhouette-based markerless techniques. Indeed, a combined markerless and markerbased system could be considered a versatile solution for gait laboratories, which could perform different type of analyses based on the desired degree of accuracy and patients' discomfort.

# BIBLIOGRAPHY


[1] P.M. Mills, S. Morrison, D.G. Lloyd, and R.S. Barrett, "Repeatability of 3D gait kinematics obtained from an electromagnetic tracking system during treadmill locomotion.," *Journal of biomechanics*, vol. 40, Jan. 2007, pp. 1504-11.

[2] P. Picerno, A. Cereatti, and A. Cappozzo, "Joint kinematics estimate using wearable inertial and magnetic sensing modules.," *Gait & posture*, vol. 28, Nov. 2008, pp. 588-95.

[3] A. Ferrari, M.G. Benedetti, E. Pavan, C. Frigo, D. Bettinelli, M. Rabuffetti, P. Crenna, and A. Leardini, "Quantitative comparison of five current protocols in gait analysis.," *Gait & posture*, vol. 28, Aug. 2008, pp. 207-16.

[4] T.D. Collins, S.N. Ghoussayni, D.J. Ewins, and J. a Kent, "A six degrees-of-freedom marker set for gait analysis: repeatability and comparison with a modified Helen Hayes set.," *Gait & posture*, vol. 30, Aug. 2009, pp. 173-80.

[5] J.L. McGinley, R. Baker, R. Wolfe, and M.E. Morris, "The reliability of three-dimensional kinematic gait measurements: a systematic review.," *Gait & posture*, vol. 29, Apr. 2009, pp. 360-9.

[6] T.B. Moeslund and E. Granum, "A survey of computer vision-based human motion capture," *Computer Vision and Image Understanding*, vol. 81, 2001, pp. 231-268.

[7] T. Moeslund, A. Hilton, and V. Krüger, "A survey of advances in vision-based human motion capture and analysis," *Special Issue on Modeling People: Vision-based understanding of a person's shape, appearance, movement and behaviour*, vol. 104, 2006, pp. 90-126.

[8] R. Poppe, "Vision-based human motion analysis: An overview," *Computer Vision and Image Understanding*, vol. 108, Oct. 2007, pp. 4-18.

[9] J. Pansiot, *Markerless Visual Tracking and Motion Analysis for Sports Monitoring*, PhD Thesis, 2009.

[10] S. Corazza, A. Sheets, G. Abrams, M. Safran, and T.P. Andriacchi, "Tennis serve analysis using on-the-field markerless motion capture," *Proceedings of NACOB 2008: North American Congress on Biomechanics*, 2008.

[11] B. Rosenhahn, C. Schmaltz, T. Brox, J. Weickert, D. Cremers, and H.-P. Seidel, "Markerless motion capture of man-machine interaction," *2008*

*IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2008, pp. 1-8.

[12] L. Wang, W. Hu, and T. Tan, "Recent Developments in Human Motion Analysis," *Pattern Recognition*, vol. 36, 2003, pp. 585-601.

[13] P. Guan, A. Weiss, A.O. Balan, and M.J. Black, "Estimating human shape and pose from a single image," *2009 IEEE 12th International Conference on Computer Vision*, IEEE, 2009, pp. 1381-1388.

[14] M. Brand, "Shadow Puppetry," *ICCV '99: Proceedings of the International Conference on Computer Vision-Volume 2*, 1999.

[15] A. Agarwal and B. Triggs, "3D human pose from silhouettes by relevance vector regression," *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2004, p. II-882; II-888.

[16] N.R. Howe, "Silhouette Lookup for Automatic Pose Tracking," *Computer Vision and Pattern Recognition Workshop, 2004 Conference on*, 2004, pp. 15-22.

[17] A. Bharatkumar, K. Daigle, M. Pandy, Q. Cai, and J. Aggarwal, "Lower limb kinematics of human walking with the medial axis transformation," *IEEE Workshop on Non-Rigid Motion*, 1994, pp. 70-76.

[18] C.W. Chu, O.C. Jenkins, and M.J. Mataric, "Markerless kinematic model and motion capture from volume sequences," *Proceedings of 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2003, p. II-475-II-482.

[19] S. Corazza, L. Mündermann, and T.P. Andriacchi, "Model-free markerless motion capture through visual hull and laplacian eigenmaps," *Summer Bioengineering Conference ASME*, 2005.

[20] I. Miki, M.M. Trivedi, E. Hunter, and P.C. Cosman, "Human Body Model Acquisition and Motion Capture Using Voxel Data," *AMDO '02: Proceedings of the Second International Workshop on Articulated Motion and Deformable Objects*, London, UK: Springer-Verlag, 2002, pp. 104-118.

[21] J. Starck and A. Hilton, "Spherical matching for temporal correspondence of non-rigid surfaces," *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, 2005, pp. 1387-1394 Vol. 2.

[22] S. Corazza, L. Mündermann, a M. Chaudhari, T. Demattio, C. Cobelli, and T.P. Andriacchi, "A markerless motion capture system to study musculoskeletal biomechanics: visual hull and simulated annealing approach.," *Annals of biomedical engineering*, vol. 34, Jun. 2006, pp. 1019-29.

[23] A. Blake and M. Isard, *Active Contours: The Application of Techniques from Graphics,Vision,Control Theory and Statistics to Visual Tracking of Shapes in Motion*, Secaucus, NJ, USA: Springer-Verlag New York, Inc, 1998.

[24] J. Deutscher, A. Blake, and I. Reid, "Articulated Body Motion Capture by Annealed Particle Filtering," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2000.

[25] R. Kehl, M. Bray, and L.V. Gool, "Full Body Tracking from Multiple Views Using Stochastic Sampling," *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005, pp. 129-136.

[26] L. Sigal, A.O. Balan, and M.J. Black, "HumanEva: Synchronized Video and Motion Capture Dataset and Baseline Algorithm for Evaluation of Articulated Human Motion," *International Journal of Computer Vision*, vol. 87, Aug. 2009, pp. 4-27.

[27] J. Jackman, *Bluescreen Compositing*, Focal Press, 2007.

[28] M. Piccardi, "Background subtraction techniques: a review," *IEEE International Conference on Systems, Man and Cybernetics, 2004*, 2004, pp. 3099-3104 vol.4.

[29] P. Kaewtrakulpong and R. Bowden, "An Improved Adaptive Background Mixture Model for Real- time Tracking with Shadow Detection," *Proc. 2nd European Workshop on Advanced Video Based Surveillance Systems*, 2001, pp. 1-5.

[30] R.C. Gonzalez and R.E. Woods, *Digital Image Processing*, Addison-Wesley Publishing Company, 1992.

[31] A. Laurentini, "The Visual Hull concept for silhouette base image understanding," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, 1994, pp. 150-162.

[32] W. Matusik, C. Buehler, L. McMillan, and S.J. Gortler, *An Efficient Visual Hull Computation Algorithm*, MIT LCS Technical Memo 623, 2001.

[33] L. Mündermann, S. Corazza, A.M. Chaudhari, E.J. Alexander, and T.P. Andriacchi, "Most favorable camera configuration for a shape-from-silhouette markerless motion capture system for biomechanical analysis," *SPIE-IS&T Electronic Imaging*, 2005.

[34] C. Bregler and J. Malik, "Tracking people with twists and exponential maps," *Proceedings of 1998 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1998.

[35] R.M. Murray, *A Mathematical Introduction to Robotic Manipulation*, CRC Press, 1994.

[36] S. Corazza, E. Gambaretto, L. Mündermann, and T.P. Andriacchi, "Automatic generation of a subject-specific model for accurate markerless motion capture and biomechanical applications.," *IEEE transactions on bio-medical engineering*, vol. 57, Apr. 2010, pp. 806-12.

[37] D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, J. Rodgers, and J. Davis, "SCAPE: Shape Completion and Animation of People. ," *SIGGRAPH Conference*, 2005.

[38] A.L. Bell, D.R. Pedersen, and R.A. Brand, "A comparison of the accuracy of several hip center location prediction methods," *J ournal of Biomechanics*, vol. 23, 1990, pp. 617-621.

[39] S. Corazza, L. Mündermann, and T. Andriacchi, "A framework for the functional identification of joint centers using markerless motion capture, validation for the hip joint.," *Journal of biomechanics*, vol. 40, Jan. 2007, pp. 3510-5.

[40] P.J. Besl and N.D. McKay, "A Method for Registration of 3-D Shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, 1992, pp. 239-256.

[41] J.L. Bentley, "Multidimensional binary search trees used for associative searching," *Communications of the ACM*, vol. 18, 1975, pp. 509-517.

[42] P.E. Gill and W. Murray, "Algorithms for the Solution of the Nonlinear Least-Squares Problem," *SIAM Journal on Numerical Analysis*, vol. 15, 1978, pp. 977-992.

[43] H.B. Nielsen, *Damping parameter in Marquardt's method*, Technical Report, Technical University of Denmark, 1999.

[44] M.J. Holthe and S.P. McLean, "Kinematic comparison of grab and track starts in swimming," *Biomechanics Symposia*, 2001, pp. 31-34.

[45] R.E. Schleihauf, L. Gray, and J. DeRose, "Three-dimensional analysis of swimming propulsion in the sprint front crawl stroke.," *Biomechanics and Medicine in Swimming*, 1983, pp. 173-184.

[46] V. Gourgoulis, N. Aggeloussis, P. Kasimatis, N. Vezos, A. Boli, and G. Mavromatis, "Reconstruction accuracy in underwater three-dimensional kinematic analysis.," *Journal of science and medicine in sport / Sports Medicine Australia*, vol. 11, Apr. 2008, pp. 90-5.

[47] Y. Ohgi, "Microcomputer-based acceleration sensor device for sports biomechanics -stroke evaluation by using swimmer's wrist acceleration," *Proceedings of IEEE Sensors*, 2002, pp. 699-704 vol.1.

[48] S.E. Slawson, L.M. Justham, A.A. West, P.P. Conway, and M.P. Caine, "Accelerometer Profile Recognition of Swimming Strokes," *The Engineering of Sport 7*, vol. 1, 2008, pp. 81-87.

[49] Y.H. Kwon, "Object plane deformation due to refraction in 2-dimensional underwater motion analysis," *Journal of Applied Biomechanics*, vol. 15, 1999, pp. 396-403.

[50] Y.H. Kwon, "A camera calibration algorithm for the underwater motion analysis," *Scientific Proceedings of the XVII International Symposium on Biomechanics in Sports*, 1999, pp. 257-260.

[51] S.G. Psycharakis, R. Sanders, and F. MiII, "A calibration frame for 3D swimming analysis," *Proceedings of ISBS 2005*, 2005, pp. 901-904.

[52] Z. Zhang and O.M. Way, "Flexible camera calibration by viewing a plane from unknown orientations," *Proceedings of the Seventh IEEE International Conference on Computer Vision*, 1999, pp. 666-673 vol.1.

[53] J.M. Lavest, G. Rives, and J.T. Lapresté, "Dry camera calibration for underwater applications," *Machine Vision and Applications*, vol. 13, Mar. 2003, pp. 245-253.

[54] E.W. Maglischo, *Swimming Fastest*, Human Kinetics, 2003.

[55] M.W. Whittle, *Gait Analysis: An Introduction*, Butterworth-Heinemann, 2007.

[56] J.R. Gage, *Gait Analysis in Cerebral Palsy*, Cambridge University Press, 1991.

[57] R.E. Cook, I. Schneider, M.E. Hazlewood, S.J. Hillman, and J.E. Robb, "Gait Analysis Alters Decision-Making in Cerebral Palsy," *Journal of Pediatric Orthopaedics*, vol. 23, 2003, pp. 292-295.

[58] S. Ounpuu, P. DeLuca, R. Davis, and M. Romness, "Long-term effects of femoral derotation osteotomies: an evaluation using three-dimensional gait analysis.," *Journal of pediatric orthopedics*, vol. 22, pp. 139-45.

[59] M. Morris, R. Iansek, J. McGinley, T. Matyas, and F. Huxham, "Three-dimensional gait biomechanics in Parkinson's disease: evidence for a centrally mediated amplitude regulation disorder.," *Movement disorders : official journal of the Movement Disorder Society*, vol. 20, Jan. 2005, pp. 40-50.

[60] M.G. Benedetti, R. Piperno, L. Simoncini, P. Bonato, a Tonini, and S. Giannini, "Gait abnormalities in minimally impaired multiple sclerosis patients," *Multiple Sclerosis*, vol. 5, Oct. 1999, pp. 363-368.

[61] J.C. Gainey, M.P. Kadaba, M.E. Wootten, H.K. Ramakrishnan, E.S. Siris, R. Lindsay, R. Canfield, and G.V. Cochran, "Gait analysis of patients who have Paget disease.," *The Journal of bone and joint surgery. American volume*, vol. 71, Apr. 1989, pp. 568-79.

[62] T.P. Andriacchi, "Functional analysis of pre and post-knee surgery: total knee arthroplasty and ACL reconstruction," *Journal of Biomechanical Engineering*, vol. 115, Nov. 1993, pp. 575-581.

[63] P.R. Cavanagh, J.A. Derr, J.S. Ulbrecht, R.E. Maser, and T.J. Orchard, "Problems with Gait and Posture in Neuropathic Patients with Insulin-Dependent Diabetes Mellitus," *Diabetic Medicine*, vol. 9, Jun. 1992, pp. 469-474.

[64] D.H. Sutherland, "The evolution of clinical gait analysis. Part II kinematics.," *Gait & posture*, vol. 16, Oct. 2002, pp. 159-79.

[65] A. Cappozzo, U. Della Croce, A. Leardini, and L. Chiari, "Human movement analysis using stereophotogrammetry. Part 1: theoretical background.," *Gait & posture*, vol. 21, Mar. 2005, pp. 186-96.

[66] M. Benedetti, F. Catani, A. Leardini, E. Pignotti, and S. Giannini, "Data management in gait analysis for clinical applications," *Clinical Biomechanics*, vol. 13, Apr. 1998, pp. 204-215.

[67] A. Cappozzo, F. CATANI, A. LEARDINI, M.G. BENEDETTI, and U. DELLACROCE, "Position and orientation in space of bones during movement: anatomical frame definition and determination," *Clinical Biomechanics*, vol. 10, Jun. 1995, pp. 171-178.

[68] G. Wu, S. Siegler, P. Allard, C. Kirtley, A. Leardini, D. Rosenbaum, M. Whittle, D.D. D'Lima, L. Cristofolini, H. Witte, O. Schmid, and I. Stokes, "ISB recommendation on definitions of joint coordinate system of various joints for the reporting of human joint motion--part I: ankle, hip, and spine. International Society of Biomechanics.," *Journal of biomechanics*, vol. 35, Apr. 2002, pp. 543-8.

[69] E.S. Grood and W.J. Suntay, "A joint coordinate system for the clinical description of three-dimensional motions: application to the knee.," *Journal of biomechanical engineering*, vol. 105, May. 1983, pp. 136-44.

[70] L. Chiari, U. Della Croce, A. Leardini, and A. Cappozzo, "Human movement analysis using stereophotogrammetry. Part 2: instrumental errors.," *Gait & posture*, vol. 21, Mar. 2005, pp. 197-211.

[71] U. Della Croce, A. Leardini, L. Chiari, and A. Cappozzo, "Human movement analysis using stereophotogrammetry. Part 4: assessment of anatomical landmark misplacement and its effects on joint kinematics.," *Gait & posture*, vol. 21, Feb. 2005, pp. 226-37.

[72] A. Leardini, L. Chiari, U. Della Croce, and A. Cappozzo, "Human movement analysis using stereophotogrammetry. Part 3. Soft tissue artifact assessment and compensation.," *Gait & posture*, vol. 21, Feb. 2005, pp. 212-25.

[73] R.B. Davis, S. Ounpuu, D. Tyburski, and J.R. Gage, "A gait analysis data collection and reduction technique," *Human Movement Science*, vol. 10, 1991, pp. 575-587.

[74] A. Cappozzo, A. Cappello, U. Della Croce, and F. Pensalfini, "Surface-marker cluster design criteria for 3-D bone movement reconstruction.," *IEEE transactions on bio-medical engineering*, vol. 44, Dec. 1997, pp. 1165-74.

[75] A. Cappello, A. Cappozzo, P.F. La Palombara, L. Lucchetti, and A. Leardini, "Multiple anatomical landmark calibration for optimal bone pose estimation," *Human Movement Science*, vol. 16, Apr. 1997, pp. 259-274.

[76] T.W. Lu and J.J. O'Connor, "Bone position estimation from skin marker co-ordinates using global optimization with joint constraints," *J Biomech*, vol. 32, Feb. 1999, pp. 129-134.

[77] A. Leardini, Z. Sawacha, G. Paolini, S. Ingrosso, R. Nativo, and M.G. Benedetti, "A new anatomically based protocol for gait analysis in children.," *Gait & posture*, vol. 26, Oct. 2007, pp. 560-71.

[78] Z. Sawacha, G. Gabriella, G. Cristoferi, A. Guiotto, A. Avogaro, and C. Cobelli, "Diabetic gait and posture abnormalities: a biomechanical investigation through three dimensional gait analysis.," *Clinical biomechanics*, vol. 24, Nov. 2009, pp. 722-8.

[79] S. Del Din, E. Carraro, Z. Sawacha, A. Guiotto, L. Bonaldo, S. Masiero, and C. Cobelli, "Impaired gait in ankylosing spondylitis.," *Medical & biological engineering & computing*, Jan. 2011.

[80] K.S. Arun, T.S. Huang, and S.D. Blostein, "Least-squares Fitting of Two 3-D Point Sets," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 9, 1987, pp. 698-700.

[81] M.E. Harrington, a B. Zavatsky, S.E.M. Lawson, Z. Yuan, and T.N. Theologis, "Prediction of the hip joint centre in adults, children, and patients with cerebral palsy based on magnetic resonance imaging.," *Journal of biomechanics*, vol. 40, Jan. 2007, pp. 595-602.

# LIST OF FIGURES

Figure 1-1 Electromagnetic motion capture system by JZZ Technologies, Inc ..... 2
Figure 1-2 Electromechanical system by MetaMotion (Gypsy 4) ..... 3
Figure 1-3 Inertial motion sensors (left) and sensing suit (right) by Xsens ..... 4
Figure 1-4 Subject wearing retroreflective markers ..... 4
Figure 1-5 Infrared camera and strobes (BTS S.r.l) ..... 5
Figure 1-6 Two-camera optoelectronic system; 3D reconstruction is performed from 2D views recorded from the cameras ..... 6
Figure 2-1 RGB space ..... 13
Figure 2-2 Example of original video frame (top) and the binary silhouette image (bottom) resulting from background subtraction ..... 14
Figure 2-3 Visual hull generation from silhouettes ..... 15
Figure 2-4 Example of visual hull. Phantom volumes are present, i.e. between arms ..... 16
Figure 2-5 Model mesh: detail of the triangles (left) and whole subject's mesh (right) ..... 17
Figure 2-6 Segmented model of a subject; each color represents a body segment. Joint centers' position (red circles) and segments' embedded frames of reference are shown. ..... 18
Figure 2-7 Effects of principal components' variation on human body shape space ..... 19
Figure 2-8 Automatic model generation pipeline ..... 21
Figure 2-9 Static visual hull of subject in Figure 2-6 (left) and model generated from it (right) ..... 22
Figure 3-1 Color analog camera by Tracer Technology Co. Ltd ..... 31
Figure 3-2 Grid employed for extrinsic calibration of the cameras ..... 32
Figure 3-3 Checkerboard panel employed for intrinsic calibration of the cameras ..... 32
Figure 3-4 Cameras' position as resulting from extrinsic calibration (coordinates in mm). Control points are shown in red ..... 36

Figure 3-5 Result of distortion compensation on an image. Straight lines do not longer appear as bended. ..... 37
Figure 3-6 Calibration grid (laboratory setup) ..... 38
Figure 3-7 Static acquisition of the subject: view from the three cameras..... 38
Figure 3-8 Frame from each synchronized view of the front crawl trial acquisition ..... 39
Figure 3-9 Example of original frame from a DV video (top), and the corresponding frames obtained through interpolation by bob-deinterlacing (bottom) ..... 40
Figure 3-10 Example of foreground extraction for a lateral-view camera ..... 42
Figure 3-11 Example of foreground extraction for a bottom-up view camera..... 43
Figure 3-12 Example of two visual hulls reconstructed from the video sequence. Orientation of global frame of reference is also shown..... 44
Figure 3-13 Segmented model of the subject (particular of upper body). Numbers indicate the sequence of the kinematic chain. ..... 45
Figure 3-14 Initial model configuration, with flexed shoulder. Orientation of global frame of reference is also shown. ..... 46
Figure 3-15 Example of three iterations of the articulated-ICP algorithm for matching of the model (yellow) to visual hull data green)..... 47
Figure 3-16 Shoulder trajectory reconstructed through digitization by 5 operators employing a commercial software (solid lines) and with the markerless technique (red dotted line). Phases A-E are explained in the text. ..... 49
Figure 3-17 Elbow trajectory reconstructed through digitization by 5 operators employing a commercial software (solid lines) and with the markerless technique (red dotted line). Phases A-E are explained in the text. ..... 50
Figure 3-18 Wrist trajectories reconstructed through digitization by 5 operators employing a commercial software (solid lines) and with the markerless technique (red dotted line). Phases A-E are explained in the text. ..... 50

Figure 3-19 Trajectories reconstructed through 5 digitizations by the same operator employing a commercial software (solid lines) and with the markerless technique (red dotted line). Phases A-E are explained in the text. ..... 51
Figure 3-20 Difference between trajectory coordinates estimated with markerless and averaged manual-tracking techniques (red line). ±3 standard deviation curves calculated on manual-tracking trajectories are shown in yellow (5 operators) and green (same operator)..... 52
Figure 3-21 Velocity of shoulder joint center calculated from trajectories reconstructed through digitization by 5 operators employing a commercial software (solid lines) and with the markerless technique (red dotted line)..... 54
Figure 3-22 Velocity of elbow joint center calculated from trajectories reconstructed through digitization by 5 operators employing a commercial software (solid lines) and with the markerless technique (red dotted line)..... 55
Figure 3-23 Velocity of wrist joint center calculated from trajectories reconstructed through digitization by 5 operators employing a commercial software (solid lines) and with the markerless technique (red dotted line)..... 56
Figure 4-1 Rigid wand with three markers, employed for intrinsic calibration of the cameras..... 63
Figure 4-2 Calibration grid with 9 markers, employed for extrinsic calibration of the cameras..... 63
Figure 4-3 Cameras' position as resulting from extrinsic calibration (coordinates in mm). Orange crosses correspond to additional infra-red illuminators..... 64
Figure 4-4 Retroreflective passive marker..... 64
Figure 4-5 Pointer used for anatomical calibration..... 65
Figure 4-6 Frame from one of the acquired video sequences (infra-red grayscale cameras) ..... 67
Figure 4-7 Example of background subtraction for the frames in Figure 4-6 ..... 69
Figure 4-8 Visual hull reconstructed from a frame of an acquired gait trial..... 69
Figure 4-9 Scheme of the model kinematic tree: each arrow point from "parent" segment to "child" segment ..... 70

Figure 4-10 Orientation of markerless technical frames for the right leg ..... 71
Figure 4-11 Example of nine iterations of the articulated-ICP algorithm for matching of the model (yellow) to visual hull data (green) ..... 72
Figure 4-12 Markers' position reconstructed in 3D space by the stereophotogrammetric system. Position of the cameras is also shown. ..... 75
Figure 4-13 Labeling of reconstructed markers (left) according to the marker model (right) ..... 76
Figure 4-14 Reconstructed anatomical landmarks' position (left) according to the kinematic model of the protocol (right) ..... 77
Figure 4-15 Static acquisition tracked with the markerless system and contemporary marker-based reconstruction of anatomical landmarks' position (black circles) ..... 80
Figure 4-16 Example of subject's pose estimated with the markerless system. Anatomical landmarks reconstructed for the right leg with markerless (red circles) and marker-based (black circles) techniques are shown. ..... 82
Figure 4-17 Hip joint angles (degrees) estimated in 15 trials with marker-based (blue solid lines) and markerless (red dotted lines) techniques ..... 84
Figure 4-18 Knee joint angles (degrees) estimated in 15 trials with marker-based (blue solid lines) and markerless (red dotted lines) techniques ..... 84
Figure 4-19 Ankle joint angles (degrees) estimated in 15 trials with marker-based (blue solid lines) and markerless (red dotted lines) techniques ..... 85
Figure 4-20 Hip angles (degrees) for one gait trial calculated with markerless technique with (red dotted line) and without (blue line) the co-registration with marker-based system ..... 87
Figure 4-21 Knee angles (degrees) for one gait trial calculated with markerless technique with (red dotted line) and without (blue line) the co-registration with marker-based system ..... 87
Figure 4-22 Ankle angles (degrees) for one gait trial calculated with markerless technique with (red dotted line) and without (blue line) the co-registration with marker-based system ..... 88

Figure 4-23 Excessive pelvic obliquity in the matched model (yellow) caused by phantom volumes in the visual hull (green)..... 89

# LIST OF TABLES

Table 3-1 Calibration error (average over the 12 calibration points in each camera) ..... 37
Table 3-2 Average reconstruction error on the 12 calibration points ..... 37
Table 3-3 RMSD values between markerless and average manual-tracking trajectories (left column); maximum SD values for manual-tracking reconstructed curves are reported in middle (inter-operator) and right (intra-operator) columns53
Table 4-1 Description of RMSD values between angles calculated with the two techniques. Average, standard deviation, minimum and maximum average values are reported. The average RMSD value is also given as percentage of the corresponding range of motion ..... 86