


Garment-based motion capture (GaMoCap): high-density capture of human shape in motion

Nicoló Biasi¹  · Francesco Setti² · Alessio Del Bue³ · Mattia Tavernini¹ · Massimo Lunardelli¹ · Alberto Fornaser¹ · Mauro Da Lio¹ · Mariolino De Cecco¹

Received: 6 August 2014 / Revised: 5 March 2015 / Accepted: 14 May 2015 / Published online: 16 July 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract This paper presents a new motion capture (MoCap) system, the garment-based motion capture system—GaMoCap. The key feature is the use of an easily wearable garment printed with colour-coded pattern and a generic multicamera setup with standard video cameras. The coded pattern allows a high-density distribution of markers per unit of surface (about 40 markers per 100 cm²), avoiding markers-swap errors. The high density of markers reconstructed makes possible a simultaneous reconstruction of shape and motion, which gives several concurrent advantages with respect to the state of the art and providing performances comparable with previous marker-based systems. In particular, we provide effective solutions to counter the soft-tissue artefact which is a common problem for garment-based techniques. This effect is reduced using Point Cluster Technique to filter out the points strongly affected by non-rigid motion. Uncertainty of motion estimation has been experimentally quantified by comparing with a state-of-the-art commercial system and numerically predicted by means of a Monte Carlo Method procedure. The experimental evaluation was performed on three different articulated motions: shoulder, knee and hip flexion-extension. The results shows that for the three

motion angles estimated with GaMoCap, the system provides comparable accuracies against a commercial VICON system.

Keywords 3D reconstruction · Motion capture · Structure-from-motion · Multi-camera systems · Soft-tissue artefacts

1 Introduction

The quest for accurate and reliable human body 3D reconstruction and motion capture has delivered many solutions that encompass different application scenarios ranging from medical, sports, military, robotics, film-making and videogames development. There exist several motion capture approaches that can be tuned to the specific analysis requirements and operational constraints with each system having its own advantages and drawbacks that made them popular in the mentioned application fields. In particular, marker-based active systems have mastered many disciplines thanks to their accuracy as provided by the easy-to-deploy optoelectronic cameras. However, being so popular, there are undoubtedly limitations that can restrict their application.

In practice, the setup stage of these systems requires the physical placement of the markers and this might be a tedious task if a full body shape has to be reconstruct. This of course also limits the number of points reconstructed thus providing a sparse representation of the 3D object/individual at study. Marker-less systems of course step over this limitation, however, at the expense of less accuracy in the 3D estimation. Moreover, the marker placement indeed poses the problem of the repeatability of the experiments, especially in clinical trials. The MOCAP technician has to be skilful enough to place markers in the very same position for each subject. This task is often not practicable when clinical conditions

The work was supported by the FP7-ICT-2009.7.2, Accessible and Assistive ICT, Grant 247765, IP project VERITAS.

✉ Nicoló Biasi
nicolo.biasi@unitn.it

- ¹ Department of Mechanical and Structural Engineering, University of Trento, via Mesiano, 77, 38121 Trento, Italy
- ² Istituto di Scienze e Tecnologie della Cognizione (ISTC), Consiglio Nazionale delle Ricerche (CNR), via alla Cascata, 56/C, 38121 Trento, Italy
- ³ Visual Geometry and Modelling Lab (VGM), Pattern Analysis and Computer Vision Department (PAVIS), Istituto Italiano di Tecnologia (IIT), Via Morego, 30, 16163 Genova, Italy



Fig. 1 A picture of the experimental setup together with a test subject wearing the colour-coded garment. Lamps and cameras (circled in red) are attached over the aluminium cage structure. Cameras are linked to the PCs that grab and process the video. The subject is standing in the middle of capture area and he can move freely inside the cage (colour figure online)

alter the placement of the marker (e.g. obesity or serious injuries).

Here we propose a system, GaMoCap, that provides an easier setup stage that allows a better repeatability of the tests together with a semi-dense 3D reconstruction of the shape. This places the GaMoCap between a marker-based and marker-less system with the advantage of retaining a high accuracy. The key factors of the system are the introduction of a garment-based MoCap that can be easily worn by the user and the implementation of novel Computer Vision algorithms that can detect a set of markers overlaid over the garment. Figure 1 shows the working setup with the garment worn by a user and the acquisition system.

The pattern is recognized using a standard camera system thus limiting the cost given by more expensive active systems. In particular, the GaMoCap also implements post-processing techniques able to deal with the soft-tissue artefacts (STA) that are related to the displacement given, for instance, by flexing muscles. This effect also degrades standard marker-based MoCap systems since the STA can inject an estimation bias when localising the joints of the articulated body. In order to proceed with a fair comparison we will first provide a review on the state of the art of MoCap systems and then detail the key features of GaMoCap by pointing out the improvements of the proposed solution.

1.1 Current MoCap state of the art

Advantages and disadvantages of MoCap technology have been analysed and discussed in several comprehensive surveys [16,30] and in this paper we review only the most relevant technologies for human motion distinguishing them

by the type of sensors employed and the reconstructed modality: dense or sparse. Magnetic motion capture systems detect the position and orientation of each body limb using a magnetic field (either the Earth's magnetic field or the field generated by a large coil), specific magnetic sensors and a fine calibration of the entire acquisition volume. With a relatively low number of markers, one for each limb, the system offers a sparse 3D output with good accuracy, medium update rates and no line-of-sight (self-occlusion) related problems. However, this technology is very expensive, has high power consumption, is sensitive to the presence of metallic objects in the environment and, most important, is susceptible to magnetic and electrical interferences. Mainly for these reasons, magnetic motion capture is only employed for the study of small and complex parts, e.g. hands [20].

Inertial motion capture technology [19] is based on miniature inertial sensors that allow to measure rotation of the joint angles using gyroscopes or accelerometers placed on each body limb. These systems can capture the full 3D body motion of a human body in real-time, are usually portable and they allow large capture volumes, since the system is completely wearable. The main drawbacks are the low resolution and accuracy, the sensitivity to electromagnetic noise and, most important, the fact that measurements drift by significant amounts over extended time periods.

Both magnetic and inertial systems allow to reconstruct the motion (i.e. the configuration and parameters of a skeletal model) of the subject but not its 3D shape, due to the reduced number of markers (transmitters/sensors) that can be applied. They also need to locate sensors on the body, which creates load effect encumbering movements.

Vision systems have a high potential thanks to the great improvements achieved in video sensing and image processing. It is possible to distinguish vision-based MoCap in two categories: marker-based and marker-less. Many successful commercial products (e.g. Vicon, Qualisys and Motion Analysis) belong to the first category. These systems recover human motion by triangulating a set of markers (either active or passive) placed on the subject skin or clothes. The main limitations are the self-occlusion problem and the marker matching problem. Regarding occlusions, as 3D is achieved by triangulation, one marker must be directly visible by more than one camera. In practice, for very complex motions, this increases the number of cameras to be deployed to increase the chance of a marker being visible by the sensors. Attaching a kinematic model may reduce the influence of such problem since the occluded marker position may be predicted by the location of the other markers. However, these models are in general piece-wise rigid and thus highly affected by STA.

Statistic priors to overcome this problem have also been proposed in the literature in terms of coarse-to-fine models [3], partially rigid models [8] and Monte Carlo Markov Chain [35]. On the other hand, the matching ambiguity prob-

lem refers to the correct assignment of a marker label in the multiple views. This problem becomes more relevant with a high number of markers (> 100) since all the markers have the same appearance, especially when using active near-infrared systems. Furthermore, the setup can be complex and the measurement procedure time-consuming, especially for clinical applications where markers must be placed one by one using a precise protocol [10].

In recent years, marker-less vision systems have proven to be a valid alternative to marker-based technologies [5, 17] providing a direct and often dense surface reconstruction of the subject. By using kinematics and dynamics models it is possible to estimate shape and motion by fitting the models to the estimated surfaces. Accuracy and reliability are critical points of these systems. In particular, since the resulting shape can be very complex, the method can generate protrusions, and it is not able to handle cavities. For these reasons hybrid systems [2] have been introduced; they combine information coming from silhouette, shading and texture to improve accuracy in the shape estimation process. Despite the relatively easy measurement procedure, the reduced accuracy still remains due to the fact that natural landmarks are less accurate than their artificial counterparts and to the complexity of human kinematics modeling, that also introduces uncertainties. Nevertheless, these systems are an excellent option for less-critical applications like gesture recognition and human motion pattern classification [34].

Another interesting approach is to turn the capture system from the inside-out, as described in [23], by placing the cameras on the subject limbs. Despite the innovative approach and the improved capturing space, the cameras on the limbs can obstacle free motions, although going toward camera miniaturization will partially solve this problem. One important interference factor is represented by the limbs motion that can induce camera rotations and cause accuracy deterioration. Being the cameras worn by the subject, this system is able to reconstruct only the subject motion but not its shape.

In November 2010, a low-cost vision system based on IR lighting patterns projection was made commercially available on the market, the Kinect. This system also provides a human pose detection algorithm by applying a classifier over the depth images as provided by the sensor [24].

Despite the fast segmentation of body limbs data, this system provides a low accuracy for clinical applications [29].

Because of the direct observation of soft tissue motion, i.e. subject's skin, all the above motion capture systems suffer from the STA problem. Inaccuracies introduced by STAs are much larger than stereophotogrammetric ones and are both systematic and random in nature. The pattern of the artefact is subject and task dependent which prevents an experimental generalization and compensation; the artefacts are highly correlated in time and frequency content with the acquired motion thus avoiding filtering techniques to be effective.

STAs are, therefore, considered the bounding error source in evaluating the skeletal motion by optoelectronic systems recording markers on the skin since the very early application of those methods for clinical, virtual reality or sport applications. STA negative influence on motion estimation have been investigated by many authors [14, 18, 25] and have been addressed in two ways: trying to avoid the interference or minimizing it. To completely avoid STA, it is possible but only with invasive techniques such as the use of radio-opaque markers implanted onto the bones [15] or even bones pin [13]. On the other way, it is possible to minimize the effect of STA by optimization techniques as described by [1].

1.2 Main contributions

The main aim of this paper was to present—and describe in detail—the GaMoCap system, a new hybrid system based on a wearable garment and a multi-camera setup able to estimate body shape and motion by means of a dense acquisition and tracking of points printed on a colour-coded pattern. This system provides the 3D localization of the markers with high density and accuracy. In particular, the GaMoCap provides an ideal trade-off between data completeness (about 3000 points) and accuracy (3D reconstruction error lower than 2 mm^3). The presented motion capture system provides both shape and motion estimation with an overall level of accuracy suitable for most applications.

The main contributions of the paper are:

2D colour-coded pattern to uniquely identify each marker: in this work we extend and definitely improve the colour-coding method first proposed by [21]; in particular, we introduce several constraints to guarantee three important properties: uniqueness, rotation invariance and compactness (see Sect. 2.1).

Easy and fully automatic calibration procedure: we propose a completely automatic calibration procedure based on the acquisition of images of a planar tool with colour-coded markers. The system only requires a minimum of 100 images for each camera to estimate both intrinsic and extrinsic camera parameters (see Sect. 2.3).

Automatic generation of human body model: by means of Structure-from-Motion techniques, our system does not need a human body model to be known a priori, but it is able to automatically recognize the number of body limbs acquired throughout the sequence and the joints within them. This is achieved with a 3-dimensional generalization of Local Subspace Affinity (LSA) algorithm (see Sect. 3).

Compensation of non-rigid deformations: by means of Points Cluster Techniques (PCT), the GaMoCap is able to compensate the effect of STAs in the estimation of joint location and joint angles (see Sect. 3.1).

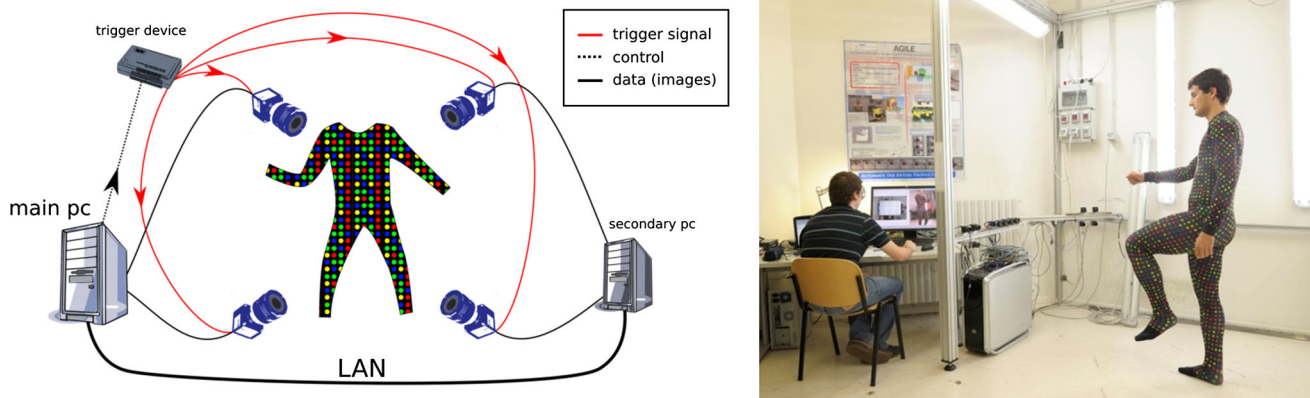


Fig. 2 Hardware setup scheme and an image showing the GaMoCap system

The paper is organized as follows: Sect. 2 describes the procedure for the 3D reconstruction of the human body; in particular it addresses the problems of colour-coded markers, camera calibration and image processing. Section 3 presents the methods proposed for the estimation of the joint angles and its uncertainty analysis. Section 4 shows the experiments conducted on three different motions and the comparison with the best performing commercial product (i.e. Vicon). Finally, Sect. 5 draws conclusions and some directions for future improvements.

2 Shape analysis

GaMoCap consists of a multi-camera setup and a garment composed of a dark suit and more than 3000 coloured circular markers as shown in Fig. 2 (right).

Figure 2 (left) shows the configuration of the hardware setup employed in GaMoCap. A main PC controls the triggering device that synchronizes the video acquisition from the cameras and it stores the acquired images from each pair of cameras. Each secondary PC controls two cameras and it communicates with the main unit. We use 8 triggerable USB Point Grey Chameleon cameras with a resolution of 1280×960 pixels and with a frame rate of 10 fps. The illumination lamps with mercury vapour at low pressure allow a high colour rendering and they are compliant with European Standard EN 12464-1. The system has been designed to be completely scalable in terms of capturing space and number of cameras to be able to perform either full body or single limb motion analysis with a unique MoCap system. The cost of the system for a full body motion capture is of about 10k€.

The key feature of the system is the wearable suit garment printed with colour-coded pattern. Each marker on the suit is identified by a unique ID, colour coded by its 8 neighbours. A detailed description of ID marker calculation and its properties is reported in Sect. 2.1.

The motion and shape of the subject are recorded as a sequence of synchronized images acquired by the multi-camera system. A 3-step procedure, reported in Sect. 2.2, is used to recover the 3D position of each marker over time. The trajectories of the markers are then used to analyse the motion of the subject, as described in Sect. 3.

The calibration of GaMoCap multi-camera system is performed with a new procedure (Sect. 2.3) based on the same colour-coded pattern printed on the garment. The use of the pattern as calibration grid, instead of the classic chessboard, makes the multicamera calibration more flexible and fully automatic, since it is able to deal with partial grid occlusions.

2.1 Marker ID calculation

The garment with colour-coded pattern was inspired by [21]. In that work each marker on the garment is identified by a unique ID, colour coded by its 8 neighbours. The ID uniqueness of markers allows to achieve a robust multicamera matching and tracking even with a high number of markers per cm^2 . This high density of markers (about 40 markers per 100cm^2) allows to reconstruct the shape of the subject, as well as its motion.

The colour coding described by [21] was originally developed for realistic simulation and visualization of textiles that points to applications ranging from virtual actors for the film industry to virtual prototyping of cloth design by means of texture replacement in the recorded video streams. In our work, this is the first time that the colour coding is applied for human motion capture.

The garment used has a dark background and a base of 4 colours for the markers: red, yellow, green and blue. HSV components of each colour have been chosen to maximize the reciprocal distance in the HSV space. Disposition and size of the markers depend on the application: a full-body analysis requires an increased size and spacing of the markers, while for a closed-up application the markers' size can be reduced

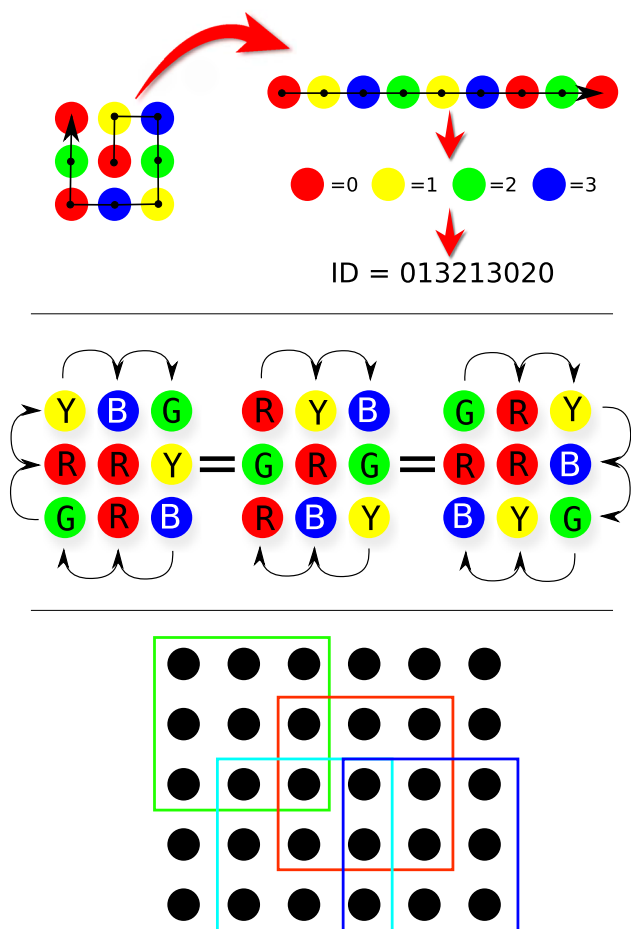


Fig. 3 Marker ID computation: graphical example (*top*), rotation invariance property (*center*) and sequences overlap (*bottom*)

and their density increased. For a full-body motion analysis markers have a diameter of 10mm and they are placed on a square grid with a 18-mm step.

Each marker on the garment has a unique colour-coded ID assigned detecting the colours of its 8 neighbours. This colour sequence of 9 elements (1 central + 8 neighbours) is then translated into a numerical one with a simple colour number assignment procedure: red = 1, yellow = 2, green = 3, blue = 4, Fig. 3 (top). The sequence of 9 elements is organized in blocks of 3 × 3 markers to obtain a compact disposition.

The proposed colour coding has 3 important properties:

- uniqueness* there are not multiple markers on the suit with the same colour sequence
- rotation invariance* a marker ID is always the same, even if the colours are all shifted forward (or backward) of N positions, Fig. 3 (centre)
- compactness* each sequences of 9 markers, organized in 3 × 3 blocks, overlaps with the neighbour sequences, Fig. 3 (bottom).

Thanks to these properties, it is possible to recover the correct value of the marker ID with a generic unknown camera-subject relative position and orientation and a high-density markers distribution.

The maximum number of ID that can be created with a sequence of 9 elements and 4 colours is $q^n = 4^9 = 262, 144$. However, not all the 262, 144 sequences are unique and rotation invariant as well. Thus, all the sequences which do not have these properties are deleted from the list of the possible sequences. The rotation invariance property of the sequence can be analysed as a cyclic shift of a sequence of m elements forward or backward of r positions. In more detail, the number of unique shift for a sequence of $m = 8$ elements (the central marker is fixed) and $q = 4$ colours can be computed as follows:

$$\frac{1}{d} s_d \text{ where } s_d = q^d - \sum_{i < d, i|n} s_i.$$

The number of strings with 8 different shifts is given by $\frac{1}{8} s_8 = \frac{65280}{8} = 8160$. Since every centre of the sequence can assume one of the 4 colours, there is a maximum of $4 \cdot 8160 = 32,640$ unique IDs. The compactness requires that the sequences, organized as 3 × 3 blocks, partially overlap with the neighbours, Fig. 3 (bottom). In order to arrange all the blocks, we have implemented a brute force algorithm: starting from a seed-block the algorithm tries to place all the other blocks. When the algorithm does not succeed to add more sequences, it undoes the last 3 sequences arranged and start again with a different sequence. The algorithm finishes when no other sequences can be arranged on the grid. With this algorithm we were able to create a grid of 90 rows and 90 columns for a total of $(rows - 2)(cols - 2) = 7744$ sequences. While this number of sequences is much lower than the theoretical one (32,640), it was enough for creating a high-density garment.

2.2 Image processing for shape reconstruction

Motion and shape of the subject are recorded as a sequence of synchronized images acquired by the multicamera system as shown in Fig. 4. The reconstruction of the subject's shape is made possible by a 3-step procedure that calculates the 3D position of each coloured marker on the garment captured by the multicamera system at each time instant. The steps for shape reconstruction are

1. markers' detection on each image plane
2. IDs' computation for each marker
3. markers' matching and triangulation.

Marker detection is based on a contour extraction method. We use the Canny edge detector to generate a binary image of edges and then use this as input for the contour detection



Fig. 4 Set of simultaneous images captured by different cameras

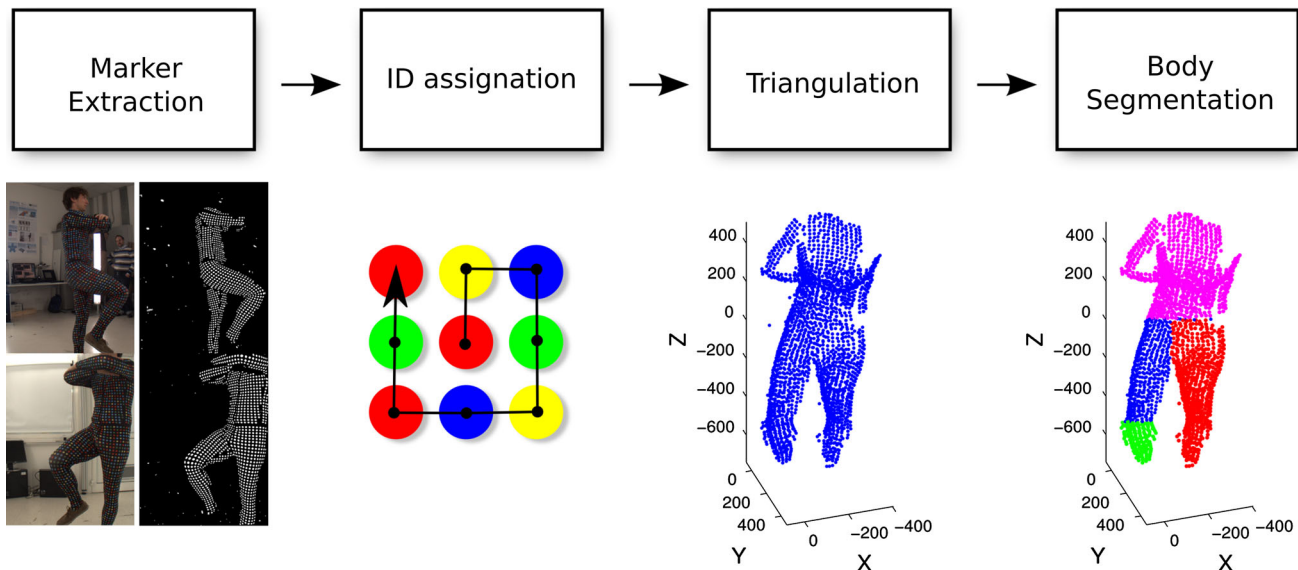


Fig. 5 The steps for shape reconstruction and motion segmentation

algorithm presented by Suzuki and Abe [26] as shown in Fig. 5.

Among all the contours extracted from the image, markers are identified through a filtering operation based on contour properties. Two distinctive properties are calculated for each contour, area and perimeter, allowing to distinguish markers from objects in the background not related to the subject. Each marker is then associated with the coordinates of its centre of mass.

The second step relies on the assignment of marker ID. For each marker, the colour and the relative position in the code sequence are calculated. The marker colour labelling procedure is facilitated by performing the analysis in the Hue colour space. This is also thanks to the stable illumination provided by the neon lamps that image the 4 colours with very different and stable Hue values. As described in Sect. 2.1, the ID of each marker is computed by associating the colour sequence of neighbour markers into a numerical one.

After the ID assignment stage, intra-image markers matching is a straightforward procedure: same marker will have same ID in different images. Since each marker is iden-

tified by the ID, the association procedure works well also with a high change in the point of view. The coordinates of each specific marker with a certain ID for all cameras are stored in a matrix to be then used for 3D reconstruction.

All the information gathered by cameras are then used to recover the 3D position of each marker using a multi-camera framework for triangulation. This was chosen instead of a multi-stereo approach that can provide a better accuracy, if uncertainty is properly estimated as in [6,7]. Differently, multi-camera provides smoother points trajectories while multi-stereo reconstruction tends to generate marker “jumps” from one stereo pair to another, which ends up with also a jump in the corresponding 3D position. An example of the trajectories generated by our system is in Fig. 6. The reduction of this effect is of paramount importance for the subsequent clustering step. There are also other reasons that led to the multi-camera solution: with the multi-camera it is possible to calibrate simultaneously all the cameras, thus speeding-up the procedure. Moreover, it does not require to place couple of cameras close to each other thus providing an easier configuration/dislocation of the cameras.

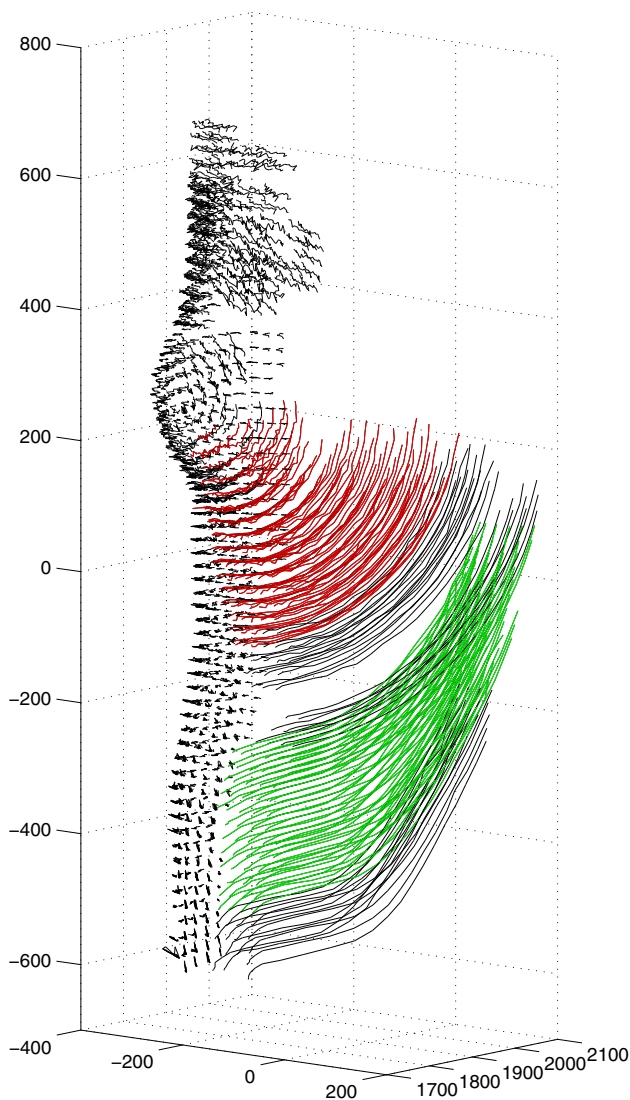


Fig. 6 The trajectories computed with our system during a whole movement. The point density is very high with respect to marker-based competitors, while the accuracy in 3D reconstruction is much higher than marker-less competitors

A first guess of the markers 3D coordinates is computed with a generalized mid-point method [11]. This method calculates the 3D position of the marker minimizing the squared distance between the optical rays passing through the same marker in two different images. Since a marker can be seen by more than two cameras, all the possible stereo combinations are taken into consideration and then the results of triangulations are fused with a weighted mean function. The weight of each triangulation result is proportional to the square of the inverse of the distance of the optical rays for that particular triangulation. A further Bundle Adjustment step [27] is applied to refine the 3D estimation by considering all the information from the cameras at each frame.

2.3 Multi-camera calibration

Multi-camera calibration is a procedure to recover camera's relative position and orientation with respect to a common reference frame. Relative camera pose calculation is achieved by acquiring with all the cameras a rigid object with a pre-defined geometry. The whole calibration of the GaMoCap system is achieved through a 2-step semi-automatic procedure based on Zhang's calibration method [33] and Bundle Adjustment refinement.

The first step consists in capturing a series of about 100 synchronized images of the calibration target with each camera. The calibration target is a rigid plastic plane where the same grid of circular colour-coded marker of the garment is printed out, see Fig. 7. Reciprocal marker distances between rows and columns are constant and equal to 68 mm.

In this step user intervention is only necessary to move the calibration target, changing position and orientation, between each camera view (Fig. 7 shows some key-frames).

The second step consists in a data elaboration procedure. In this step the 3D position and orientation of the calibration plane are recovered for each frame. While the original implementation of Zhang's method uses a chessboard as a target, we adopt a grid built with the same colour-coding method used on the garment. This allows us to avoid problems of ambiguity in plane pose reconstruction when the target grid is partially missed in some images. With this pattern the pose of the plane can be recovered by recognising at least 3 markers with relative IDs. The use of this strategy results in a speed up of the calibration since an image of a partially occluded calibration pattern is still a valid calibration image.

The data elaboration procedure consists in

- compute intrinsic parameters for each camera and a first guess of extrinsic parameters
- run Bundle Adjustment to refine both intrinsic and extrinsic parameters.

Figure 8 shows the relation between a point expressed in a world reference frame and its coordinates on the image plane.

This relation can be written in a compact form as

$$x = \mathbf{K} [\mathbf{R}|t] X \quad (1)$$

where x are the coordinates of the point in image reference frame, \mathbf{K} is the matrix of intrinsic parameters, \mathbf{R} and t are, respectively, the rotation matrix and translation vector between world and camera reference frame. The calibration procedure aims to identify the parameters that characterize these transformations for all cameras. The camera model used is a finite projective model.

This camera model is defined by the calibration matrix:

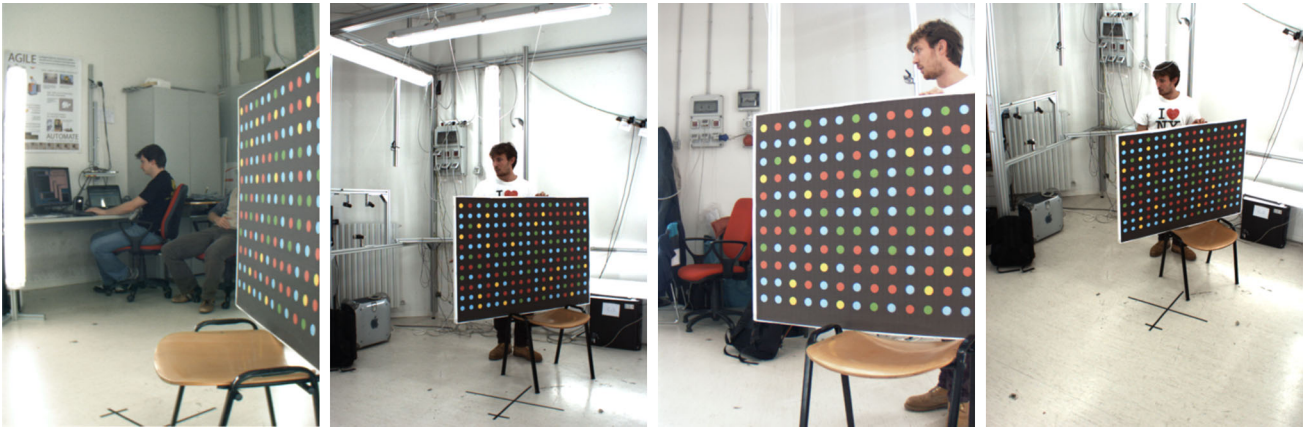


Fig. 7 Set of images captured by different cameras during calibration procedure. The planar calibration pattern is printed with a colour-coded pattern

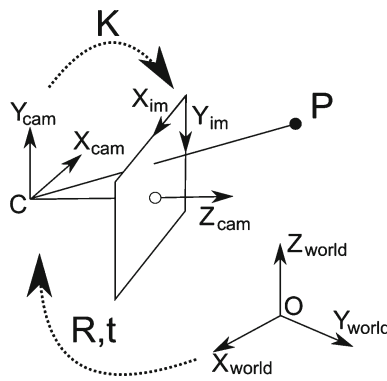


Fig. 8 Transformations between the three reference frames: world, camera and image. *Dotted arrows* represent the two transformations ($[R|t]$ and K) between the three reference frames (world, camera and image)

$$K = \begin{bmatrix} f_x & s & x_0 \\ 0 & f_y & y_0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{2}$$

where f_x and f_y are the focal lengths of the camera, respectively, in the x and y directions, expressed in pixel; s is the skew factor that is taken into account only with non-rectangular pixels; and x_0 and y_0 are the coordinates of the principal point, i.e. the point of intersection between the principal axis and the image plane.

This camera model is used to define the relations between a 3D point defined in the camera reference frame and its projection into image plane:

$$x = K [I_3|0] X_{cam}, \tag{3}$$

where X_{cam} is a vector that contains the homogeneous coordinates of a generic point $[x_{cam}, y_{cam}, z_{cam}, 1]^T$ expressed in the camera reference frame. Compensation of real lens dis-

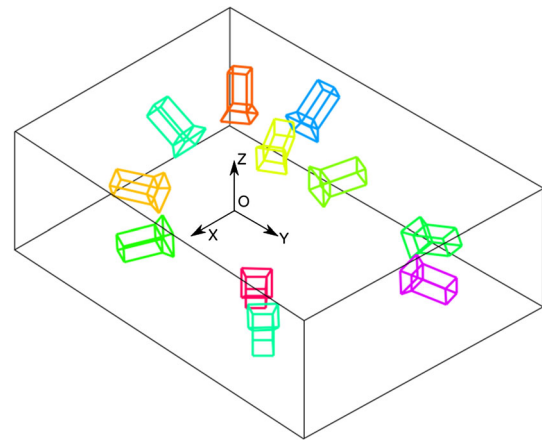


Fig. 9 Reference frames of the 11 cameras with respect to the world reference after calibration procedure

tortion is taken into account using a radial-tangential model identified by 5 parameters (3 tangential and 2 radial).

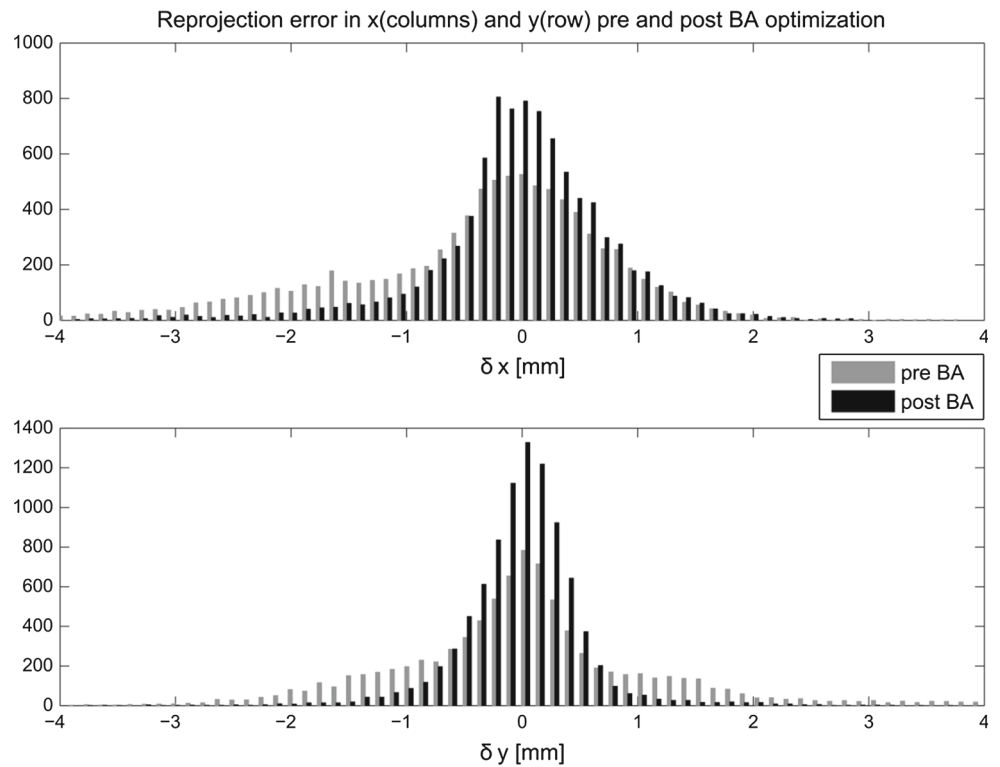
Extrinsic parameters are defined as rotation and translation of the cameras reference frame with respect to the world reference frame.

$$X_{cam} = [R|t] X \tag{4}$$

where R and t are, respectively, the rotation matrix and translation vector between world and camera reference frame, and X are the homogeneous coordinates $[x, y, z, 1]^T$ of the point P expressed in the world reference frame.

Extrinsic parameters are successively refined with a Bundle Adjustment technique while intrinsic parameters are kept fixed. After the calibration procedure, the mean reprojection error achieved in our setup was lower than 1 pixel as reported in Fig. 10. Extrinsic parameter calculation allows to identify the relative position and orientation of each camera with respect to the world reference frame, Fig. 9.

Fig. 10 Reprojection errors of markers along columns (x) and rows (y) of image before and after Bundle Adjustment optimization



3 Motion analysis

To perform motion analysis of a dynamic body, it is necessary to process the 3D point cloud obtained from the previous triangulation step. For human motion analysis, this is necessary to compute the number of limbs, their connection (joint position) and the angles between them. Given the unique IDs of the markers, generating trajectories followed by each point is straightforward. Given these trajectories, the first step of motion analysis is the automatic segmentation of different body parts.

Although the coded markers could allow us to work with a pre-defined limb segmentation based on a mapping of each marker's ID on the garment, we prefer to work with an automatic algorithm to perform this task. The main reason of this choice is the fact that different subjects can wear the garment in different ways—e.g. a short man could need to fold the final parts of the garment's legs—and thus some areas of the garment, especially close to the joints, might be subject to mistakes, in particular markers located around knees and elbows.

As pointed out in [28], among the algorithms for motion segmentation, one of the best performing for articulated and degenerate motions is the Local Subspace Affinity (LSA) analysis [32]. LSA is a general framework for motion segmentation exploiting spectral analysis to define the data clusters which refer to different motion subspaces. It is based on local subspace fitting in the surrounding of each trajectory followed by spectral clustering. The basic idea is that

trajectories of points belonging on the same rigid body lie on a subspace of dimension 3. Thus, we can build an affinity matrix between each pair of point trajectories. Once the affinity matrix is built, a standard clustering algorithm is applied, K -means in our case, to group the trajectories compatibles with different limbs. This procedure has been proved very effective to work in the image space, i.e. directly from the trajectories of points in the image plane, and able to handle degenerate motions as well. In this work we use a generalization of this technique to work in a Euclidean 3D space [22]. In our particular case, the main advantage of LSA consists in recovering the structure of an articulated body (human skeleton) without the need of a priori model.

This procedure clusters the 3D point trajectories into different body limbs. Figure 17 shows a result of the motion segmentation. In order to run this algorithm, a unique marker ID is an important feature since it provides a direct method to track markers during the various frames and thus build a reliable trajectory matrix.

Once 3D points are clustered into limbs, it is possible to calculate reciprocal limbs rotation with a simple cloud registration based on Singular Value Decomposition (SVD) [12].

The compensation in the estimated angle's given non-rigid motion and STA was achieved using Point Cluster Techniques (PCT) [1]. This technique is included in an optimization phase successive to LSA segmentation to filter out the points affected by STAs and thus obtaining a better angle at joints estimation, see Figs. 17 and 18.

Uncertainty of angle estimation was experimentally quantified by means of a Vicon system and numerically predicted by means of a Monte Carlo (MC) procedure (Sect. 4.2).

3.1 Angle estimation

The rotation of a limb in each frame is computed by estimating the rotation of the reconstructed cloud of points relative to that limb. The point cloud rotation, with respect to the first frame, is computed using a Singular Value Decomposition (SVD) technique as proposed by [12]. Let p_i and p_{i+1} be two points sets relative to the same limb at frame i and $i + 1$. These two points sets are related by the equation: $p_{i+1} = \mathbf{R} \cdot p_i + t + n_{i+1}$ where \mathbf{R} is a 3×3 rotation matrix, t is a 3×1 translation vector, and n_{i+1} a noise vector.

The Least Squares solution of \mathbf{R} and t is based on the SVD of the cross correlation matrix C computed as $C = \tilde{p}_i \cdot \tilde{p}_{i+1}^T$, where $\tilde{p}_i = p_i - \dot{p}_i$, $\tilde{p}_{i+1} = p_{i+1} - \dot{p}_{i+1}$ and \dot{p}_i, \dot{p}_{i+1} are, respectively, centroids of the point sets p_i and p_{i+1} .

The rotation matrix \mathbf{R} is obtained as $\mathbf{R} = \mathbf{U}\mathbf{V}^T$, where \mathbf{U} and \mathbf{V} are the result of the SVD decomposition of $\mathbf{C} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$.

Relative rotation between two limbs, A and B, at i -th frame is calculated by composing the direct cosine rotation matrix of the limb A with the inverse of the limb B as

$$\mathbf{R}_i^{A \rightarrow B} = \mathbf{R}_i^A \cdot (\mathbf{R}_i^B)^{-1} \quad (5)$$

In order to obtain the rotation angle between two limbs, the $\mathbf{R}_i^{A \rightarrow B}$ cosine rotation matrix is expressed with the Rodrigues' rotation vector representation. In this representation the direction of rotation vector defines the axis of rotation and the magnitude is the value of the angle.

The main limitation with this angle estimation technique is the LSA rigid motion assumption. Markers on the suit in correspondence to the joints are stretched and heavily affected by STAs during motion. This non-rigid motion of the points at the joints affects the LSA motion segmentation and thus angles estimation procedure. Since these markers have trajectories compatible with both limbs, LSA will produce a less accurate segmentation. Figure 17 shows an example where the points of torso and leg at hip proximity are mixed up. This problem is addressed in the next section.

4 Experiments

In order to prove the good performances of our system, we tested single parts of the processing software and the complete system. In particular we present a synthetic experiment for angle estimation, and three basic motions have been

analysed for real data, together with an uncertainty analysis for angle estimation.

4.1 Synthetic data

To show the impact of STA effect, a toy example is presented simulating wrongly segmented points in the angle estimation procedure. In this experiment two sets of points relative to two limbs, generically called A and B, are created. Points relative to limb B are rotated of an angle from 0° to -90° around point O , while points of limb A are kept fixed. These data are used in two trials.

In trial 1 the points sets are considered perfectly segmented: points relative to the first limb are labelled as A and points relative to second limb as B. The angle of relative rotation between B and A is calculated with the SVD technique previously described. In trial 2 a segmentation error is introduced by assigning the label A to two points relative to limb B at joint proximity as shown in Fig. 11. After this operation the angle of relative rotation between B and A is estimated. The results of the angle estimated in these two trials are shown in Fig. 12.

It is noticeable that the segmentation in trial 1 leads to a perfect estimation of the angles, while the wrong segmentation of points in trial 2 reduces the angle estimated of a maximum of about 10 %.

In order to reduce the effect of wrong limbs segmentation, especially close to the joint position, the points which are stretched and heavily affected by STAs during motion are identified with a technique, called PCT [1]. The main idea of PCT is to assign a weight factor to each point and optimize it to minimize the inter-frame variation of the inertia tensor of the points cloud. This procedure aims to reduce the effects of STA decreasing the contribution of points behaving as "weak" rigid bodies. The points that move from the reference frame with a non-rigid transformation will receive a low weight. In our work the weights are used to filter out the points considered affected by STAs by setting a threshold computed as the 50 % of the mean of the weights of the points evaluated for each frame. If over the threshold, a point is considered affected by STA and thus rejected as a non-rigid one.

4.2 Angle uncertainty analysis

Uncertainty is estimated experimentally using a VICON system as the reference and numerically quantified by means of a Monte Carlo Method (MCM) procedure, according to the GUM standard [30 100:2008 4]. The experimental comparison of GaMoCap and VICON, that is generally accepted as a very reliable system, is able to provide only a rough uncertainty estimation for two reasons: the VICON, due to STAs, can only have a level of accu-

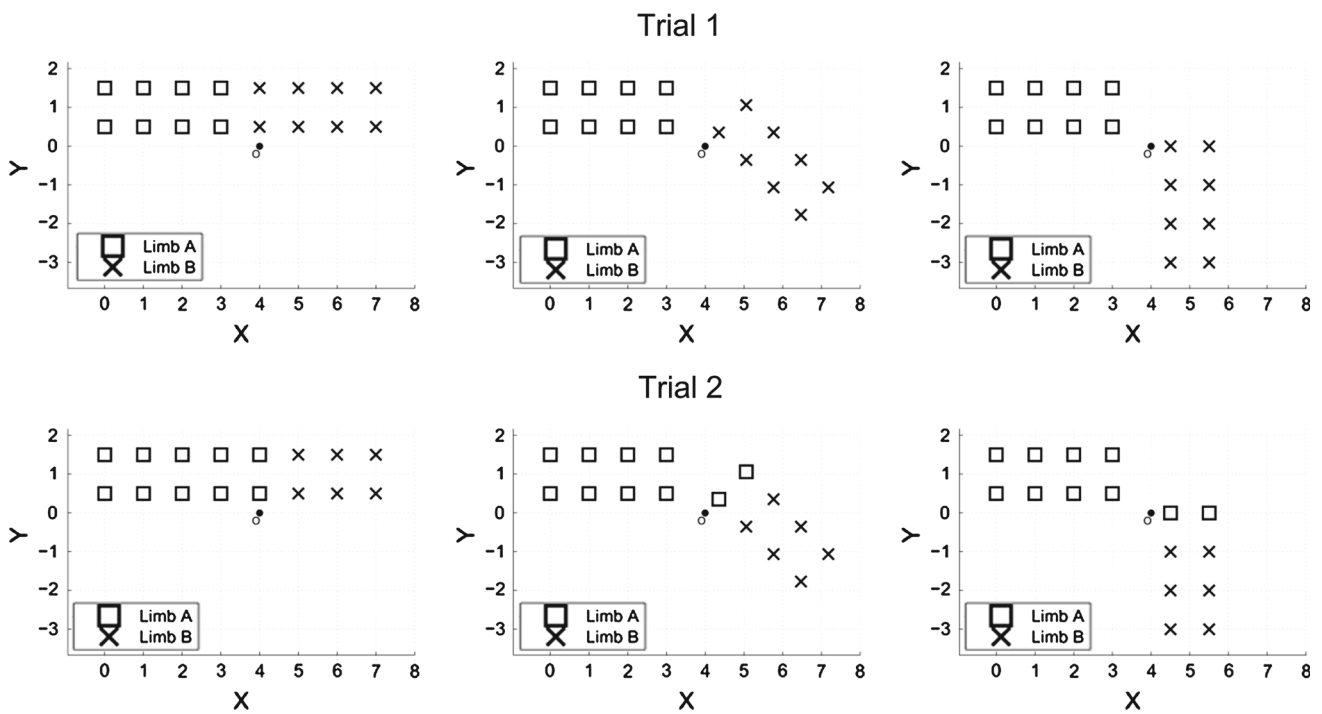
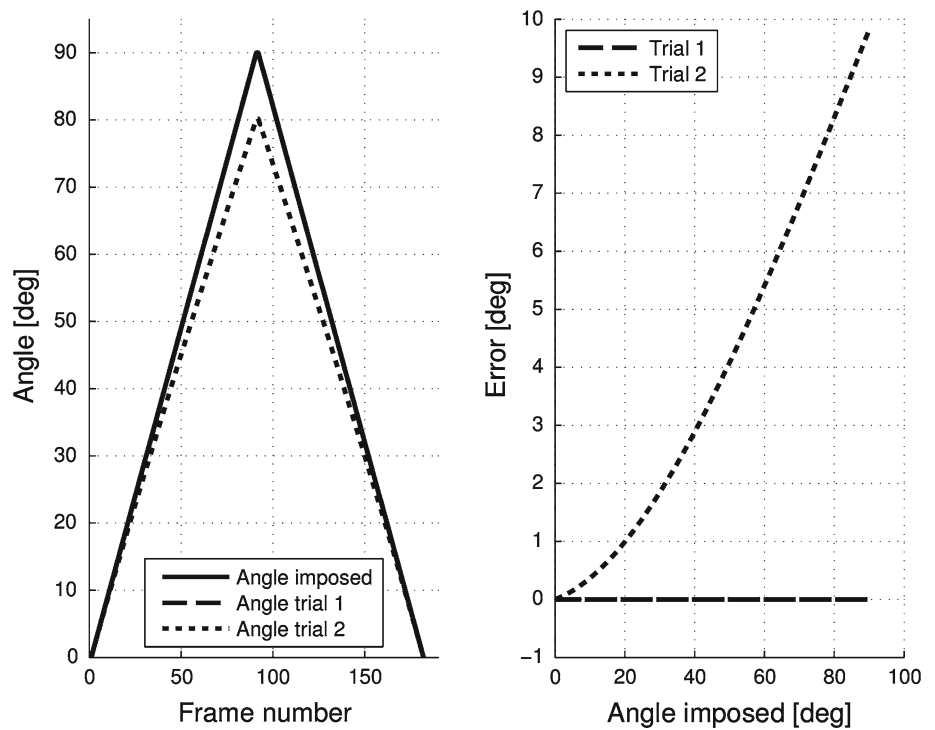


Fig. 11 Trial 1 perfect points segmentation. Trial 2 wrong segmentation of point at joint proximity

Fig. 12 Angle estimated in trial 1 and 2 versus the ground truth angle. Error is calculated as the difference between the angle of the trial and the angle imposed



racy comparable to GaMoCap; furthermore, to achieve a complete statistics through experiments, a massive experimental campaign is required that is out of the scope of the present paper. The implemented MCM procedure quantifies in a more complete statistical way the system accuracy. To our knowledge such experimental/numerical procedure

for LSA and PCT has never been performed in the literature.

The evaluation of the uncertainty of the angle estimated is achievable through an uncertainty propagation stage, [30 100:2008 4]. This can be implemented in several ways. The two mainly used are

- uncertainty propagation based on replacing an explicit model by its first-order Taylor series approximation—the law of propagation of uncertainty [30 5.1.2];
- numerical methods [30 G.1.5] that implement the propagation of distributions, specifically using Monte Carlo Method (MCM).

In practice, only for simple cases where an explicit and close to linearity model is available can the law of propagation of uncertainty be applied. Due to strong non-linearities and the procedural complexity of the model that relates inputs (marker 3D coordinates) to the output (angle between bodies) without an explicit formulation, MCM is chosen.

The main stages of the MCM numerical procedure for uncertainty propagation are:

Formulation we define the inputs (X_i) and output quantity (Y) and the development of a model relating them. On the basis of available knowledge, we assign a PDFs—Gaussian (normal), rectangular (uniform), etc.—to the inputs X_i . Then we assign a joint PDF to those input variables X_i that are not independent. We then finally define of the number M of trials of Monte Carlo.

Propagation we propagate the PDFs for the X_i through the model to obtain the PDF for Y . This step can be summarized with the following sequence: generation of M vectors, by sampling from the assigned PDFs, as realizations of the (set of N) input quantities X_i ; for each of such vectors, forming the corresponding model value of Y , yielding M model values; use of the M model values to estimate the expected value (y) of Y and the standard uncertainty $u(y)$ associated with y .

Summarizing use of the propagated PDF for Y to obtain the expectation and standard deviation of the output quantity Y .

Figure 13 shows the block diagram of the procedure for angle estimation of GaMoCap. Input variables are represented in triangles, the output variable in ellipse and the functional blocks in rectangles. Input variables are the coordinates of markers extracted from images and the intrinsic and extrinsic calibration parameters. The output variable corresponds to the measured quantities, i.e. the angle between limbs. Coordinates of markers extracted from images influence the triangulation process, 3D point cloud computation and consequently the evaluation of the angle between the limbs. Intrinsic and extrinsic camera parameters obtained during the calibration phase also influence the position of the markers extracted (through distortion coefficients) and the triangulation results. Covariance matrix of intrinsic and extrinsic parameters is recovered from [4].

The marker ID assignment is not represented in Fig. 13 and it is not considered as an input variable of the process

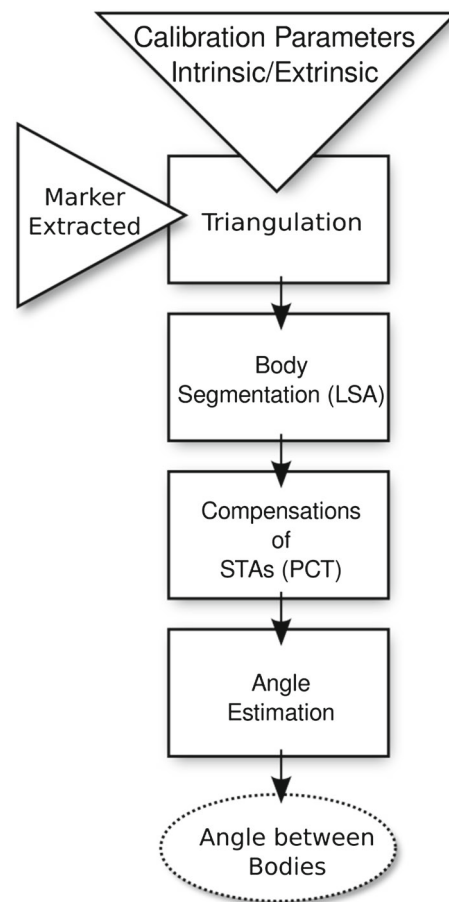


Fig. 13 Block diagram of the procedure for angle estimation

because the colour recognition process is considered very robust and rarely affected by errors.

The formulation step in MCM methods consists also in the selection of the trials number. This is fixed at 100 after a convergence study. The first step of propagation is the input data generation. In this step a hybrid procedure is adopted, in which we use synthetic, obtained by PDF sampling, and real data, obtained directly by image elaboration. The use of real data is necessary to overcome the PDF generation process for marker extraction. The PDF distribution of markers extracted is very difficult to estimate because of the great number of variables involved in the estimation procedure of markers center: light condition, image focus, marker inclination with respect to the camera, marker projection on the CCD, etc. The number of markers in each image (~ 1000), multiplied by the number of cameras (4) and by the number of frames for each motion analysed (100), generate a number of samples (10^5) that provide a good statistics of this influencing factor. The calibration parameter sampling, due to variables correlation, is performed from the covariance matrix obtained by calibration algorithm [4].

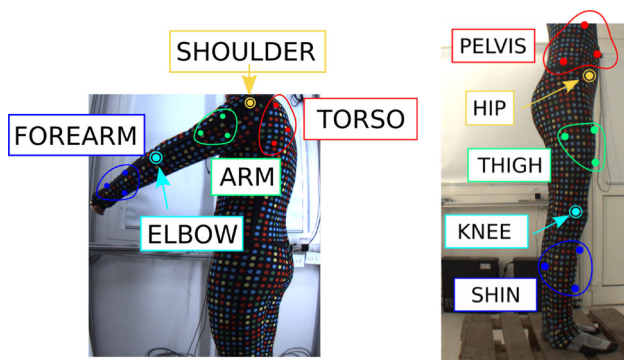


Fig. 14 Vicon marker disposition on GaMoCap suit for motion A (left) and motions B and C (right)

The second step of propagation consists in the model evaluation for 100 times with the input data previously generated. As last step, the 100 output values (angle between limbs) obtained numerically through the model are used to estimate the mean and uncertainty of the angles, with a level of confidence set at 65%. The uncertainty range corresponding to this confidence level is obtained according to GUM as the minimum.

4.3 Real data

In order to evaluate the GaMoCap accuracy, results in terms of estimated angles were compared to the commercial MoCap system VICON.¹ The VICON system is an optoelectronic MoCap based on triangulation of passive retroreflective markers and active infrared cameras. Performances in terms of 3D reconstruction accuracy of VICON were extensively studied during the years [9,31] confirming a high accuracy and repeatability of the results.

VICON, compared to GaMoCap, achieves a higher capture frequency (>100Hz) but a lower number of reconstructed markers. The maximum number of markers for the VICON system is limited by issues related to the tracking algorithm. This problem does not occur in GaMoCap because each marker has a unique ID.

The skeletal configuration can be retrieved with VICON using the following procedure: a triplet of markers is placed on each body limb, which is then considered as a rigid body, see Fig. 14 for the markers disposition.

We performed experiments on three very different types of motions, one of them involving the arms (A–shoulder flexion/extension) and two for the legs (B–knee and C–hip flexion/extension). All motions were acquired a frequency of 10Hz and last about 10s. Figure 15 shows the three motions analysed.

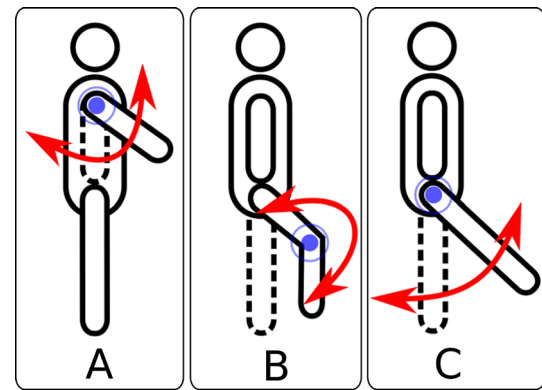


Fig. 15 Motions analysed for angle estimation. A shoulder flexion/extension, B knee flexion/extension, C hip flexion/extension

Next sections will present the uncertainty analysis on one sequence for each kind of movement. Angles at joints estimated with the GaMoCap are compared with angles estimated with VICON system. Figure 14 shows the disposition of VICON markers employed, respectively, in motion A (left) and in motions B and C (right). Values of uncertainty of angles for GaMoCap, computed as described in Sect. 4.2, are also reported and discussed.

4.3.1 Motion A

Figure 16 (left) shows the comparison of the angle estimated with LSA, LSA+PCT and VICON. The uncertainty bands of angles estimated with MCM for LSA and LSA+PCT are shown on the plot. Error bands, relative to standard uncertainty, show that the three methods lead to compatible results.

Figure 16 (right) shows the standard uncertainty of the angle estimated with LSA and LSA+PCT with respect to VICON. LSA+PCT have a standard uncertainty of 3.8° . LSA shows a standard uncertainty which increase linearly with the angle estimated. The standard uncertainty can be expressed as linear expression as follow: $3.3^\circ + 0.08 \cdot AV$. Where AV is the value of the angle estimated with VICON, expressed in degree. This effect is more evident for motions B and C.

4.3.2 Motion B

Figure 17 shows segmentation results for motion B with LSA (left) and LSA+PCT (right) methods. It is noticeable that many points in correspondence to the joints (within circles) are wrongly segmented with LSA. This is due to the fact that the points at joints do not move with a rigid motion. LSA+PCT is not affected by this effect since points close to the joints have been pruned. In Fig. 17 (left) LSA+PCT points highlighted are the one used to calculate the angle.

¹ <http://www.vicon.com/>.

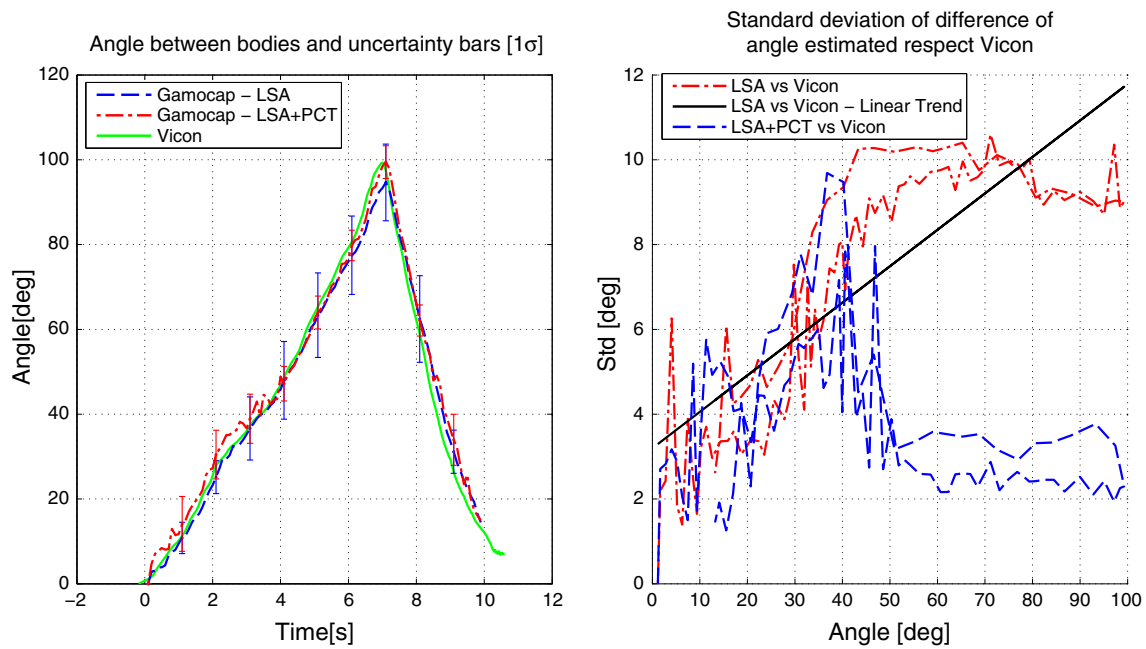
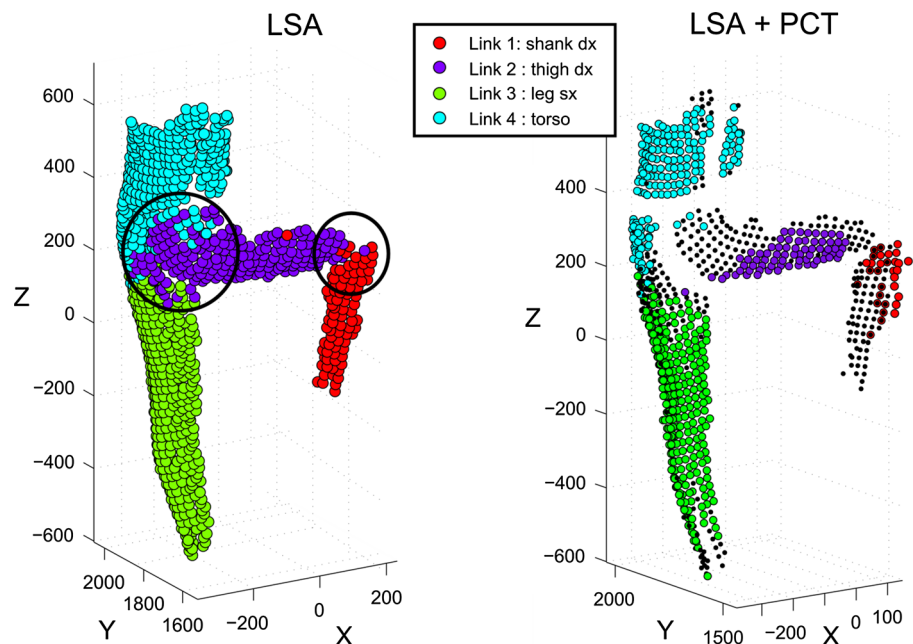


Fig. 16 Comparison of angle obtained for motion A with GaMoCap (LSA and LSA+PCT) and VICON (*left*) and standard deviation of the difference of the angle estimated with LSA and LSA+PCT respect to VICON (*right*) for Motion A

Fig. 17 Motion B. Knee flexion/extension. Comparison of results of segmentation operation with LSA (*left*) and LSA+PCT (*right*)



Finally, Fig. 18 shows angles calculated with the three methods.

It is noticeable that LSA+PCT is very close to VICON angle estimation while LSA deviates strongly in terms of lower angle estimation query (-20° with respect to VICON) and wider bands of error. These two effects are justified by the wrong segmentation of points close to the joints provided by LSA and the non-determinism of trajectory clustering

operation in LSA. Non-rigid motions of points in proximity of joints introduce an error of angle underestimation as described in Sect. 3.1.

Wider error bars are justified by the fact that clustering step of points in LSA is performed with a K -means approach. In this algorithm, the points chosen as seed for segmentation are selected randomly at each trial of MCM. This generates a segmentation slightly different in each trial, especially

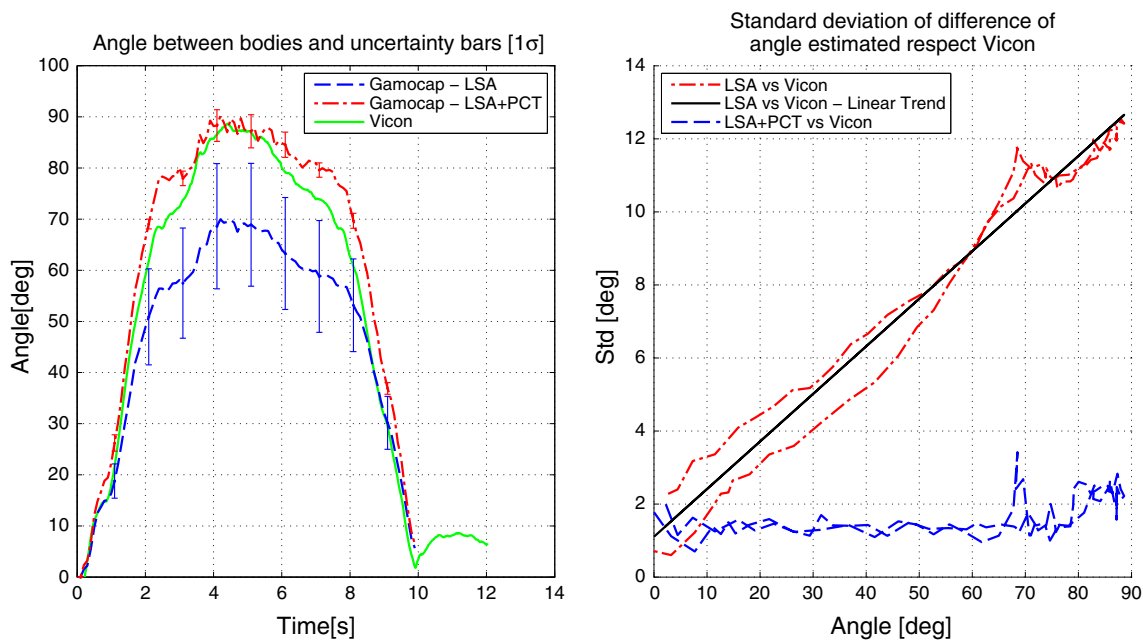
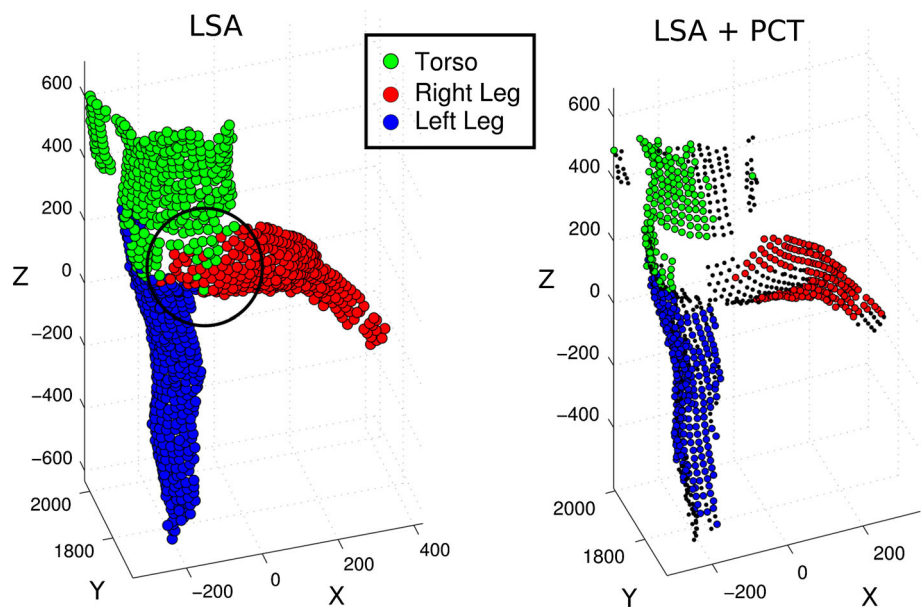


Fig. 18 Left comparison of angle obtained for motion B with GaMoCap (LSA and LSA+PCT) and VICON. Right standard deviation of the difference of the angle estimated with LSA and LSA+PCT respect to VICON for Motion B

Fig. 19 Motion C. Hip flexion/extension. Comparison of results of segmentation operation with LSA (left) and LSA+PCT (right)



for the points at the proximity of joints. The different segmentation of these points in each MCM trial causes a wider error bands in angle estimation. Figure 18 (right) shows the standard uncertainty of the angle estimated with LSA and LSA+PCT with respect to VICON. LSA+PCT have a standard uncertainty of 1.6°, and this value remains constant with the angle estimated. LSA shows a standard uncertainty which increases linearly with the angle estimated. The standard uncertainty can be expressed as linear expression as follows: $1.1^\circ + 0.13 \cdot AV$. Where AV is the value of the angle estimated with VICON, expressed in degree.

4.3.3 Motion C

Figure 19 shows segmentation results for motion C with LSA (left) and LSA+PCT (right) methods.

Figure 20 (left) shows how LSA underestimates the maximum angle at hip of about 14° with respect to VICON. Figure 20 (right) shows the standard uncertainty of the angle estimated with LSA and LSA+PCT with respect to VICON. LSA+PCT have a standard uncertainty of 0.8°. This value remains constant with the angle estimated. LSA instead has a standard uncertainty that increases linearly with the angle

Table 1 Average difference between angles estimated with LSA and LSA+PCT with respect to VICON for 10 subjects and 3 different type of motions

Acq.	Motion type	LSA vs Vicon	LSA+PCT vs VICON
1	A	4.6°	1.1°
2	A	3.0°	2.7°
3	A	16.0°	0.8°
4	B	7.9°	1.4°
5	B	21.6°	0.3°
6	B	5.3°	2.9°
7	C	2.8°	0.3°
8	C	4.2°	3.1°
9	C	5.6°	0.1°
10	C	3.6°	1.1°
Avg.	–	7.46°	1.16°

estimated. The standard uncertainty can be expressed as linear expression as follows: $0.2^\circ + 0.05 \cdot AV$. Where AV is the value of the angle estimated with VICON, expressed in degrees.

The difference of repeatability for motion A and B compared with C is explained by a greater number of points reconstructed on which the angle is calculated. Table 1 shows further experimental validation provided by a test on ten different subjects reporting the results of the comparison of the proposed method— named LSA+PCT—and LSA with respect to VICON. We can appreciate how the average error

of GaMoCapis very limited, about 1.16° , and the poorest result in our test is still a very good achievement, about 3° . Moreover, the PCT introduced to handle STA provides an average improvement in performances of about 6.3° with peaks of more than 20° in sequence 5 (Table 1).

5 Conclusions

In this article we presented a new multicamera garment-based motion capture system, the GaMoCap. The key feature of this system is a wearable garment with a colour-coded pattern. The pattern allows a high-density distribution of markers on the target surface without introducing mismatches. The high number of markers makes possible a simultaneous reconstruction of motion and shape of the subject that wears the garment.

Analysis of the motion is performed without a prior knowledge of the kinematics model of the subject with Local Subspace Affinity. LSA allows, through the segmentation of the trajectories of the markers, to identify the limbs and joint position of the subject. The unique ID of each marker on the garment ensures to build outlier-free trajectory matrices to segment the motion.

The procedure of angle estimation was analysed in particular to show how soft tissue artefacts introduce an error into limb segmentation that leads to an underestimation of the joint angles. This effect was reduced using Point Cluster Techniques to filter out the points affected by non-rigid motion. This technique is particularly suitable for GaMoCap

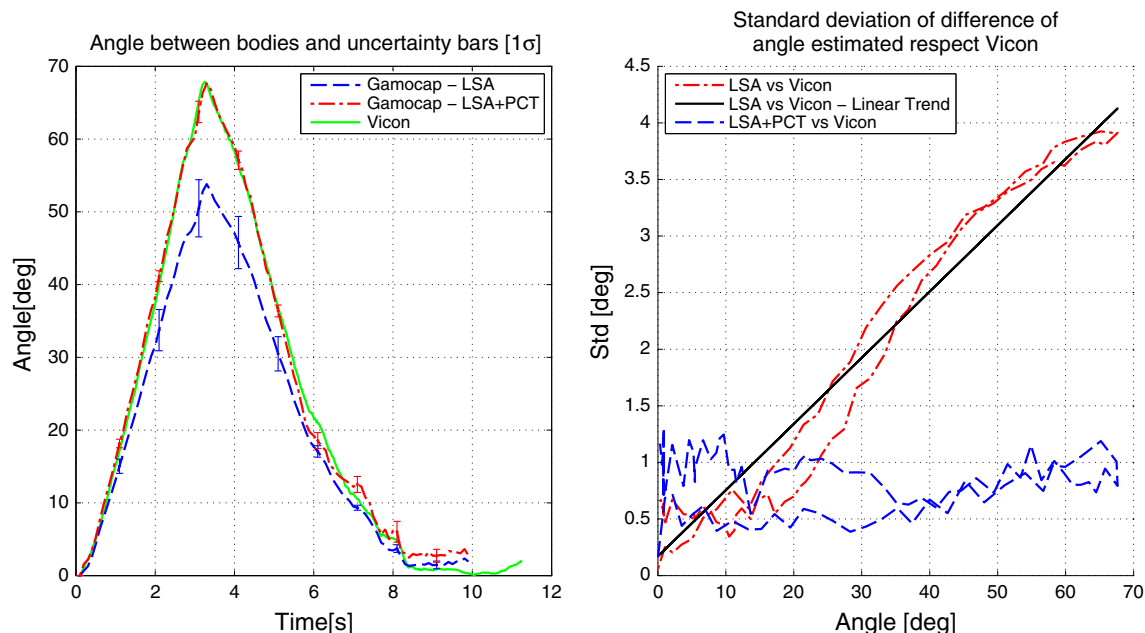


Fig. 20 Left comparison of angle obtained for motion C with GaMoCap (LSA and LSA+PCT) and VICON. Right standard deviation of the difference of the angle estimated with LSA and LSA+PCT respect to VICON for Motion C

because of the high number of points reconstructed and distributed on the limbs.

A detailed uncertainty analysis of the angle estimation method was performed. Uncertainty of angle estimated with GaMoCap was estimated experimentally by means of the VICON reference and numerically quantified by means of the Monte Carlo Method procedure. The MCM proposed used an hybrid approach in which synthetic, obtained by sampling the probability density function, and real data, obtained directly by image processing, were used.

The VICON system was chosen as reference because of the high accuracy and repeatability of the results but, due to STA, it can only have a level of accuracy comparable to GaMoCap.

The experimental validation shows how LSA+PCT gives results comparable with VICON system. The standard uncertainties of the angles estimated have a value of 3.8° for the shoulder, 1.6° for the knee and 0.8° for the hip. LSA, due to STAs, underestimates the angle of a maximum of 20° compared with VICON for the knee and a maximum of 14° for the hip. Furthermore, results show how the uncertainty of the angle estimated with LSA+PCT is stable during the motion while LSA increases linearly with the angle.

Acknowledgments The authors would like to thank Patrick Olivier, Guy Schofield and Dave Green of Culture Lab, Newcastle University (<http://di.ncl.ac.uk/>) for providing the VICON system and the support during data acquisition.

References

- Andriacchi, T.P., Alexander, E.J., Toney, M.K., Dyrby, C.O., Sum, J.: A point cluster method for in vivo motion analysis: Applied to a study of knee kinematics. *J. Biomech. Eng. Trans. ASME* **120**(6), 743–749 (1998)
- Ballan, L., Cortelazzo, G.M.: Multimodal 3d shape recovery from texture, silhouette and shadow information. In: International Symposium on 3D Data Processing, Visualization, and Transmission, pp. 924–930. IEEE, Chapel Hill, NC (2006). doi:[10.1109/3DPVT.2006.99](https://doi.org/10.1109/3DPVT.2006.99)
- Bartoli, A., Gay-Bellile, V., Castellani, U., Peyras, J., Olsen, S., Sayd, P.: Coarse-to-fine low-rank structure-from-motion. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8 (2008)
- Bouquet, J.Y.: Camera calibration toolbox for matlab. http://www.vision.caltech.edu/bouquetj/calib_doc/index.html (2010)
- Corazza, S., Gambaretto, E., Mündermann, L., Andriacchi, T.P.: Automatic generation of a subject-specific model for accurate markerless motion capture and biomechanical applications. *IEEE Trans. Biomed. Eng.* **57**(4), 806–812 (2010)
- De Cecco, M., Baglivo, L., Parzianello, G., Lunardelli, M., Setti, F., Pertile, M.: Uncertainty analysis for multi-stereo 3d shape estimation. In: IEEE International Workshop on Advanced Methods for Uncertainty Estimation in Measurement (AMUEM), pp. 22–27. IEEE, New York (2009)
- De Cecco, M., Pertile, M., Baglivo, L., Lunardelli, M., Setti, F., Tavernini, M.: A unified framework for uncertainty, compatibility analysis, and data fusion for multi-stereo 3-d shape estimation. *IEEE Trans. Instrum. Meas.* **59**(11), 2834–2842 (2010)
- Del Bue, A., Llado, X., Agapito, L.: Non-rigid metric shape and motion recovery from uncalibrated images using priors. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 1191–1198 (2006)
- Dorociak, R.D., Cuddeford, T.J.: Determining 3-D system accuracy for the VICON 370 system. *Gait Posture* **3**(2), 88 (1995)
- Gorton, G.E.I., Hebert, D.A., Gannotti, M.E.: Assessment of the kinematic variability among 12 motion analysis laboratories. *Gait Posture* **29**(3), 398–402 (2009)
- Hartley, R.L., Sturm, P.: Triangulation. *Comput. Vis. Image Underst.* **68**(2), 146–157 (1997)
- Kanatani, K.: *Statistical Optimization for Geometric Computation: Theory and Practice*. Elsevier, New York (1996)
- Lafortune, M.A., Cavanagh, P.R., Sommer, H.J.I., Kalenak, A.: Three-dimensional kinematics of the human knee during walking. *J. Biomech.* **25**(4), 347–357 (1992)
- Leardini, A., Chiari, L., Della Croce, U., Cappozzo, A.: Human movement analysis using stereophotogrammetry: part 3. Soft tissue artifact assessment and compensation. *Gait Posture* **21**(2), 212–225 (2005)
- Lundberg, A., Svensson, O.K., Bylund, C., Selvik, G.: Kinematics of the ankle/foot complex—part 3: influence of leg rotation. *Foot Ankle Int.* **9**(6), 304–309 (1989)
- Moeslund, T.B., Hilton, A., Krüger, V.: A survey of advances in vision-based human motion capture and analysis. *Comput. Vis. Image Underst.* **104**(2), 90–126 (2006)
- Mündermann, L., Corazza, S., Andriacchi, T.P.: The evolution of methods for the capture of human movement leading to markerless motion capture for biomechanical applications. *J. NeuroEng. Rehab.* **3**(1), 1–11 (2006)
- Reinschmidt, C., van den Bogert, A.J., Nigg, B.M., Lundberg, A., Murphy, N.: Effect of skin movement on the analysis of skeletal knee joint motion during running. *J. Biomech.* **30**(7), 729–732 (1997)
- Roetenberg, D., Luinge, H., Slycke, P.: Xsens MVN: full 6DOF human motion tracking using miniature inertial sensors. In: Tech. rep, Xsens Motion Technologies BV (2009)
- Sandilands, P., Choi, M.G., Komura, T.: Capturing close interactions with objects using a magnetic motion capture system and a RGBD sensor. In: Kallmann, M., Bekris, K. (eds.) *Motion in Games. Lecture Notes in Computer Science*, vol. 7660, pp. 220–231. Springer, Berlin (2012)
- Scholz, V., Stich, T., Magnor, M., Keckeisen, M., Wacker, M.: Garment motion capture using color-coded patterns. In: *ACM SIGGRAPH Sketches*. ACM, New York (2005)
- Setti, F., De Cecco, M., Del Bue, A.: A multi-view stereo system for articulated motion analysis. In: Richard, P., Braz, J. (eds.) *International Conference on Computer Vision Theory and Applications (VISAPP 2010)*, pp. 367–372. INSTICC Press Angers, France (2010)
- Shiratori, T., Park, H.S., Sigal, L., Sheikh, Y., Hodgins, J.K.: Motion capture from body-mounted cameras. *ACM Trans. Graph.* **30**(4), 31:1–31:10 (2011)
- Shotton, J., Fitzgibbon, A.W., Cook, M., Sharp, T., Finocchio, M., Moore, R., Kipman, A., Blake, A.: Real-time human pose recognition in parts from single depth images. In: Cipolla, R., Battistato, S., Farinella, G.M. (eds.) *Machine Learning for Computer Vision, Studies in Computational Intelligence*, vol. 411, pp. 119–135. Springer, Berlin (2013)
- Stagni, R., Fantozzi, S., Cappello, A., Leardini, A.: Quantification of soft tissue artefact in motion analysis by combining 3d fluoroscopy and stereophotogrammetry: a study on two subjects. *Clin. Biomech.* **20**(3), 320–329 (2005)
- Suzuki, S., Abe, K.: Topological structural analysis of digitized binary images by border following. *Comput. Vis. Graph. Image Process.* **30**(1), 32–46 (1985)

27. Triggs, B., McLauchlan, P.F., Hartley, R.I., Fitzgibbon, A.W.: Bundle adjustment—a modern synthesis. In: Triggs, B., Zisserman, A., Szeliski, R. (eds.) *Vision Algorithms: Theory and Practice*. Lecture Notes in Computer Science, vol. 1883, pp. 298–372. Springer, Berlin (2000)
28. Tron, R., Vidal, R.: A benchmark for the comparison of 3-d motion segmentation algorithms. In: *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–8 (2007)
29. Weiss, A., Hirshberg, D.A., Black, M.J.: Home 3D body scans from noisy image and range data. In: *International Conference on Computer Vision (ICCV)*, pp. 1951–1958. IEEE, New York (2011)
30. Welch, G., Foxlin, E.: Motion tracking: no silver bullet, but a respectable arsenal. *IEEE Comput. Graph. Appl.* **22**(6), 24–38 (2002)
31. Windolf, M., Götzen, N., Morlock, M.: Systematic accuracy and precision analysis of video motion capturing systems—exemplified on the vicon-460 system. *J. Biomech.* **41**(12), 2776–2780 (2008)
32. Yan, J., Pollefeys, M.: A general framework for motion segmentation: independent, articulated, rigid, non-rigid, degenerate and non-degenerate. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) *Computer Vision—ECCV 2006*. Lecture Notes in Computer Science, vol. 3954, pp. 94–106. Springer, Berlin (2006)
33. Zhang, Z.: A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(11), 1330–1334 (2000)
34. Zhou, H., Hu, H., Liu, H., Tang, J.: Classification of upper limb motion trajectories using shape features. *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* **42**(6), 970–982 (2012)
35. Zhou, H., Li, X., Sadka, A.H.: Nonrigid structure-from-motion from 2-d images using markov chain monte carlo. *IEEE Trans. Multimed.* **14**(1), 168–177 (2012)



Nicoló Biasi received his M.S. degree in Mechatronic Engineering from the University of Trento, Italy in 2010 and his Ph.D. degree in Space Sciences, Technologies and Measurements from CISAS, Padua, in 2013. His research and development includes Computer Vision, human shape and motion reconstruction, pattern recognition. He is a co-founder of Robosense, start-up of the University of Trento.



Francesco Setti received his Ph.D. in Mechanical Measurements in 2010 from the University of Padua and his MS in Mechatronics Engineering in 2006 from University of Trento. He was Marie Curie fellow at the Queen Mary University of London under the supervision of Prof. Lourdes de Agapito (2010–2011) and then Post-Doc research fellow at the Institute of Cognitive Science and Technology of the Italian National Research Council. He is currently

Post-Doc research fellow at the Department of Computer Science of the University of Verona. He has co-authored more than 20 papers in international peer-reviewed conferences and journals.



Alessio Del Bue received the Laurea degree in Telecommunication engineering in 2002 from University of Genova and his Ph.D. degree in Computer Science from Queen Mary University of London in 2006. He was a researcher in the Institute for Systems and Robotics (ISR) at the Instituto Superior Tecnico (IST) in Lisbon, Portugal. Currently, he is leading the Visual Geometry and Modelling (VGM) lab in the department of Pattern Analysis and Computer



Mattia Tavernini received his B.S. degree in Industrial engineering, his M.S. degree in Mechatronic engineering from the University of Trento, Trento, Italy, and a Ph.D. degree in Space Sciences, Technologies and Measurements from the University of Padova in 2006, 2009 and 2013, respectively. He is co-founder of Robosense Srl, His interests include mobile robots localization and control, 3D data processing and sensor fusion.



Massimo Lunardelli graduated in Mechanical Engineering at Padua University (Italy) in 2006. From 2006 to 2012 he cooperated with University of Trento in several research and technology support projects. His main focus are machine vision as well as quality control and 3D reconstruction.



Alberto Fornaser received his B.S. degree in Industrial engineering, his M.S. degree in Mechatronic engineering and Ph.D. degree in Mechatronic systems from the University of Trento, Italy, in 2009, 2010 and 2014, respectively. His technical interests include robotics, image processing, 3D processing and object recognition. Co-founder and CEO of Robosense, start-up of the University of Trento.



Mauro Da Lio received the Laurea degree in mechanical engineering from the University of Padova, Italy, in 1986. He is Full professor of mechanical systems with the University of Trento, Italy. His earlier research activity was on modeling, simulation and optimal control of mechanical multibody systems, in particular vehicle and spacecraft dynamics. More recently his focus shifted to the modeling of human sensory-motor control, in particular, drivers and motor impaired

people. Prior to academic career he worked for an offshore oil research company in underwater robotics (an EUREKA project). He was/is involved in several EU framework programme 6 and 7 projects (PReVENT, SAFERIDER, interactIVe, VERITAS, adaptIVe, and NoTremor).



Mariolino De Cecco Associate Professor of Mechanical Measurements and Robotics. CoI of OSIRIS (Rosetta—ESA Cornerstone Mission, launched in February 2003). Unit manager of two European research projects VERITAS (EU—FP VII) and AGILE (Eurostars 2008). Coordinator of the H2020 project EMBRACEme that was positively evaluated being currently in the reserved lists. Authored more than 100 journal/conference papers. Co-funder of Robosense,

start-up of the University of Trento.