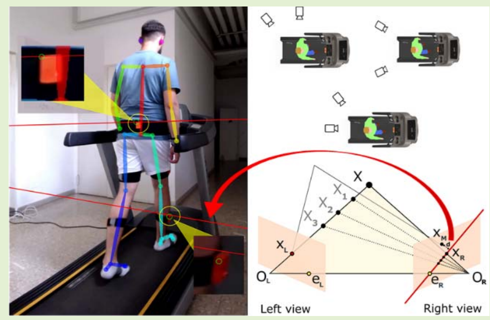
Validation of a 3D Markerless System for Gait Analysis Based on OpenPose and Two RGB Webcams

Erika D'Antonio, *Graduate Student Member, IEEE*, Juri Taborri, Ilaria Mileti, Stefano Rossi, and Fabrizio Patané

***Abstract*—Markerless human motion capture could be a potential tool for studying body biomechanics. Despite the extensive previous studies, the design of an accurate markerless system able to perform a quantitative clinical gait analysis is still an open challenge. Our purpose was to characterize the performance of a low-cost markerless system, consisting of the open-source library OpenPose, two webcams and a linear triangulation algorithm. The system was validated in terms of 3D kinematic gait analysis, by comparison with inertial sensors. Two synchronized videos of six healthy subjects were recorded in three webcam configurations, in walking and running sessions on a treadmill. Sagittal joint angles were compared between the two systems, to evaluate the kinematic performance of the markerless system. Results showed that the angular paths, provided by OpenPose, were close to the ones computed by inertial sensors in all camera positions, but OpenPose inaccuracy affected the computation of joint angles parameters producing absolute errors up to 14.0°± 1.8°. The observed differences depended on OpenPose ability in the recognition of joints centers and on camera configurations. In the optimal condition (joint visible from both cameras during the whole trial) our system accuracy resulted equal to 1.6 °± 0.3 °. This system could be useful in those applications where a real-time body tracking is required, i.e. video games or virtual rehabilitative trials. When a quantitative analysis is needed, the low level of accuracy may be overcome, by both increasing the cameras' number and implementing an optimization triangulation algorithm.**



***Index Terms*— Accuracy, gait analysis, inertial sensors, markerless, motion capture, neural network.**

I. Introduction

THREE-DIMENSIONAL (3D) gait analysis, which concerns the systemic study of human locomotion during walking, plays an important role in the diagnosis and treatment of patients with neuromotor disorders or, more in general, affected by gait abnormalities [1], [2]. Gait analysis consists in a first visual examination of the patient's gait and a subsequent objective measurement of spatio-temporal, kinematic and/or kinetic parameters [3], in order to quantify the magnitude of deviations from normal gait [4]. To date, optoelectronic systems, which record the movement of markers placed on reference joints by means of multiple cameras, represent the most accurate and precise technology for the assessment of joint kinematics [5]–[7]. These systems are considered as the gold standard in quantitative clinical gait analysis assessment, but their use is restricted to laboratories and research environments, because marker tracking requires high costs, technical skills as well as long preparation time. The latter is due to the reflective markers' placement on the user's skin, which is a delicate procedure, since small variations in their position could induce large changes in the computation of joint angles [8], [9].

Inertial Measurement Units (IMUs) represent a viable alternative to optoelectronic systems, for the measurement of body angular kinematics [10]. Placing IMUs on the human segment

Manuscript received April 23, 2021; accepted May 13, 2021. Date of publication May 17, 2021; date of current version July 30, 2021. This work was supported by the University of Rome Niccolò Cusano. The associate editor coordinating the review of this article and approving it for publication was Prof. Yudong Zhang. *(Corresponding author: Erika D'Antonio.)*

Erika D'Antonio is with the Assistive Robotics and Interactive Exosuits (ARIES) Laboratory, Institute of Computer Engineering (ZITI), University of Heidelberg, 69120 Heidelberg, Germany (e-mail: erika.dantonio@ziti.uni-heidelberg.de).

Juri Taborri and Stefano Rossi are with the Department of Economics Engineering Business Organization (DEIM), University of Tuscia, 01100 Viterbo, Italy (e-mail: juri.taborri@unitus.it; stefano.rossi@unitus.it).

Ilaria Mileti and Fabrizio Patané are with the Mech. Meas. and Microelectr. Laboratory (M3lab), Department of Engineering, University Niccolò Cusano, 00166 Rome, Italy (e-mail: ilaria.mileti@unicusano.it; fabrizio.patane@unicusano.it).

Digital Object Identifier 10.1109/JSEN.2021.3081188

links and estimating their 3D orientation in the space it is possible to evaluate the 3D angular kinematics through the application of *ad-hoc* biomechanical model. Despite their low cost with respect to the optoelectronic system and the absence of limitation of the testing environment to a controlled laboratory [11], they also require technical skills and a long execution time in order to perform a correct sensor alignment with the body segment [12]. Furthermore, the presence of ferromagnetic disturbances could affect sensor orientation, leading to a wrong computation of the joint angles [13].

The limitations of marker-based systems could be potentially overcome by video-based markerless systems [14]–[16]. Markerless techniques involve no need to apply markers on subjects' skin, making the experimental session faster and easier. In addition, they allow to perform a gait analysis even outside of the laboratory, which is particularly useful for those sport applications that require a biomechanical analysis, as well as for the quantitative assessment of patients with neuromotor dysfunction, during daily life activities [2]. Markerless technologies for motion capture have advanced dramatically in recent years, spacing from depth cameras, such as Microsoft's Kinect or Stereolabs Zed camera, to integrated multicamera devices [17]. Although the recent technical developments in markerless hardware and software, their use in biomechanical or clinical applications is still scarce. Several authors [18], [19] analyzed the performance of Kinect camera, showing that its accuracy in kinematic gait analysis is not sufficient for precise and accurate biomechanical analyses. In addition, the short range of depth cameras, the lower framerate and the need of light-controlled environment preclude their application in sport biomechanics [20]–[22].

Multicamera systems combined with stereo reconstruction algorithms for 3D pose estimation could be a valid alternative to depth cameras [23]. Corazza *et al.* [15][24] and Sandau *et al.* [14] have achieved high accuracy with markerless motion capture compared to marker-based systems. However, the robustness of multicamera approach was strictly related to the number of cameras used, *i.e.* generally greater than seven, and to the biomechanical model used for human body representation. Several previous approaches, which modelled human body as rigid segments, had problems with the accurate identification of joints and the resulting 3D kinematics [17], [25], [26].

Recent progress in the field of computer vision, advanced image features and modern machine learning could provide a solution for reducing the required number of cameras, increasing the number of people who can be simultaneously tracked and avoiding the articulated modelling of the human body [27], [28]. In particular, open-source Convolutional Neural Networks (CNN) have emerged as potential method for 2D single and multi-person pose estimation, such as OpenPose (OP) [29] and PoseNet [30]. These networks, which are trained with data obtained from monocular images, detect firstly the joints or limbs and then aggregate them into the human body. In particular, OP is able to identify, by means of a commercial webcam, not only the joint centers but also the orientation of vectors between two joints. By combining multiple calibrated 2D images, OP can perform a 3D tracking of human body that could allow the computation of the 3D lower limb kinematics. Despite the potential of OP, the design of a reliable markerless system for the capture of human body movement is still an open challenge [17]. Previous authors analyzed the performance of OP in the computation of lower limbs angles with a single camera [31], [32], by comparing the results with marker-based technologies. However, only Zago *et al.* [33] characterized OP performance in the computation of spatiotemporal gait parameters, when multiple points of view are integrated. The system composed by two cameras, OP and a triangulation algorithm was validated by means of a comparison with an optoelectronic motion capture system. They showed that OP underestimated the step length of about 1.5 cm while no significant differences were found for the measurement of the swing and stance times. The feasibility of performing a gait analysis with a low cost markerless system was confirmed, but the accuracy in the 3D detection of anatomical landmarks and in the 3D estimation of joint angles during human gait need to be characterized.

In our previous works [34], [35] we developed a low-cost markerless system, consisting of OP tracking, two webcams and a linear triangulation algorithm. The system kinematic accuracy has been characterized by a comparison with IMUs in a walking session on a treadmill, at comfortable speed and in a single camera configuration. The purpose of the present study, instead, is to generalize the results by a deepening kinematic characterization analyzing different camera positions and locomotion activities, on a larger number of subjects. In particular, we recorded two synchronized videos of six healthy subjects in three webcam configurations, in a walking session, and in a running session, both performed on a treadmill. The gait kinematics was compared with the one obtained by inertial sensor data, in order to compute system accuracy in the detection of sagittal angles of knee, hip and ankle. Since in clinical gait analysis the majority of assessment occurs in the sagittal plane, which is also the one in which typical walking movements are mainly performed [36], the 3D coordinates over time of anatomical landmarks have been used to compute the joint angles in that specific anatomical plane [37]. That procedure allowed us also to assess OP ability in the detection of body anatomical landmarks.

II. Methods

A. Participants and Experimental Design

This methodological study was designed to evaluate the metrological performance in kinematic gait assessment of a low-cost system composed by OP, two webcams and a linear triangulation algorithm. The study involved six healthy subjects (age 27.6 ± 3.4 (Mean ± STD), 1 female). On the day of testing, participants were healthy and did not suffer from any lower limb musculoskeletal injuries or neurological problems that could compromise the data. The volunteers were informed about the aims and benefits of the study and they signed a written informed consent for the participation in the experiment. The protocol was in accordance with the Helsinki Declaration.

The workflow required for the estimation of 3D body joint coordinates and the validation of system kinematic accuracy

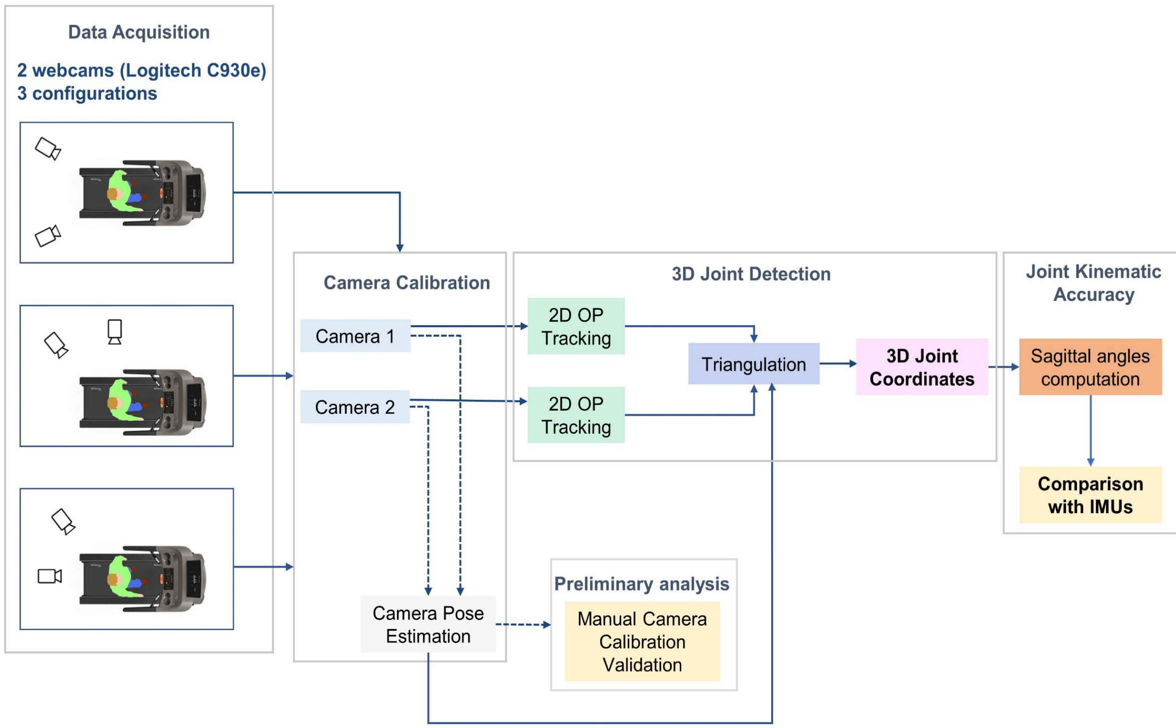


Fig. 1. Workflow for the estimation of 3D joints coordinates and validation of kinematic accuracy.

is described in Fig.1. It consisted of five main steps: Data Acquisition, Camera Calibration, 3D Joint Detection, Manual Camera Calibration Validation and Joint Kinematic Accuracy. Data acquisition comprised the technical requirements and the process that enables video capture. Camera calibration allowed to estimate the camera pose parameters necessary for the reconstruction of the 3D space. Manual camera calibration validation concerned the estimation of triangulation accuracy. 3D joint detection was the step in which the 2D points provided by OP were triangulated in order to obtain the 3D coordinates of the body joints. Finally, we estimated the kinematic accuracy by performing a comparison with IMUs. The five steps are described in detail in the following sections.

B. Data Acquisition

The experimental setup consisted of two identical high-definition webcams (Logitech brio 4k stream edition) placed around a motorized treadmill. The cameras were synchronized and operated at 60 frames per second, with a resolution of 720 × 1280 pixel. The cameras' frame rate was chosen in order to not limit the acquisition of dynamic movement considering that the 99 % of the signal power during gait is preserved for frequencies up to 15 Hz [38], and that previous authors [31], [33], [39] showed how a frame rate equal to 30 Hz is yet suitable for performing gait analysis. In the current study, we increased the camera frame rate to 60 Hz in order to be suitable also for the running activity. The contrast and brightness were automatically selected by the software provided by the manufacturer. In order to investigate the contribution of the cameras' point of view on the estimated gait parameters, the webcams were placed in three different positions (Fig.1):

- Back-Back (BB): cameras placed 1 m apart symmetrically behind the treadmill, pointing the back of the subject, reproducing the classic stereoscopic camera configuration.
- Latero-Lateral (LL): both cameras placed 1 m apart laterally to the treadmill, pointing the left side of the subject, like the point of view of 2D single camera applications.
- Back-Lateral (BL): one camera pointing the back of the subject and one pointing his left side. The distance between the two cameras optical centers was equal to 1 m.

The three configurations were chosen to (i) optimize the visibility of body joints, minimizing the triangulation errors, as described in the Manual Camera Calibration Validation section; (ii) maximize the volume occupied by the subjects and, (iii) reproduce the camera configurations usually adopted with depth cameras [18], [40], [41] or 2D acquisitions [42], [31].

In all positions, cameras were anchored on aluminum bars perpendicular to the walking direction at a height of 1 m, framing the subject vertically.

An IMU system (MTws Xsens Technologies, Enschede, The Netherlands) was used as the reference system for the estimation of kinematic accuracy. This decision was driven by previous studies showing a comparable reliability of IMUs with respect to the optoelectronic system for measuring kinematic gait parameters [11], [43]. When combined with sensor fusion algorithms, the IMU's accuracy in measuring sagittal angles ranges between 0.2° and 2.9° [44].

Before the experimental session, the IMUs were aligned outside at least 3 m away from metallic objects in order to avoid ferromagnetic disturbances. Seven IMUs were placed on shank, thigh and foot, for both sides, and on mid pelvis by means of suitable elastic belts in order to avoid relative movements between each sensor and the relative body segment. Each subject was asked to perform the IMUs Functional Calibration procedure consisting in a standing and sitting task, lasting 10 s, for body-to-sensors alignment [45]. The IMUs sampling frequency was set at 60 Hz.

For each camera configuration two different locomotion activities were recorded: (i) **Walking** at comfortable speed (3.5 km/h); (ii) **Running** at a speed of 7.5 km/h. The time duration of each trial was 60 s.

C. Camera Calibration

The camera set used in this study can be described by the epipolar geometry [46], which is essentially the geometry of the intersection of the image plane with the epipolar planes containing the baseline. This model consists of two kinds of parameters, called Intrinsic and Extrinsic parameters, encoded by the Essential matrix and the Fundamental matrix. Intrinsic parameters represent the optical center and focal length of the camera and determine the linear projection of 3D points on the image plane. The Extrinsic parameters represent the location of the cameras in the 3D scene and relative position of one camera in relation to the other. The estimation of these parameters, which allows the calibration of the stereoscopic system, was performed within Matlab (vR2016b, The Mathworks Inc., Natwick, USA) by means of the Stereo Camera Calibration App, which uses the calibration algorithm of Zhang Z. [47]. A planar 7×9 grid checkerboard pattern, consisting of black and white 30 mm $\times$ 30 mm squares was framed by the two cameras while moving the pattern into the working volume. For each camera configuration, 45 paired images of the checkboard were acquired and then imported in the Matlab toolbox (Fig. 2). The paired pictures for which the reprojection error was larger than 1.5 pixel (0.23 cm) were removed by the calibration. Furthermore, we acquired a paired picture of the checkboard pattern placed on the treadmill surface and oriented such that the global coordinate system X-Z plane was parallel to the floor.

D. 3D Joints Detection

Videos, synchronously acquired by the two webcams, were processed within OP to extract the 2D position, in pixel, of lower limbs joints, for each frame in JSON format. OP is an open source processing framework for real-time multi-person 2D joints detection, released by a research group from the Carnegie Mellon University [29]. OP is written in C++ programming language, using OpenCV and Caffe and it can be freely adapted to various operating systems. OP takes as input colour images from simple webcams and, by using a two-branch multi-stage convolutional neural network (CNN) [48], produces as output the 2D location of 18 anatomical keypoints for each person in the image. First, a feedforward network simultaneously predicts a set of 2D confidence maps of body part locations and a set of 2D vector fields that encode the location and orientation of limbs over the image domain [28].

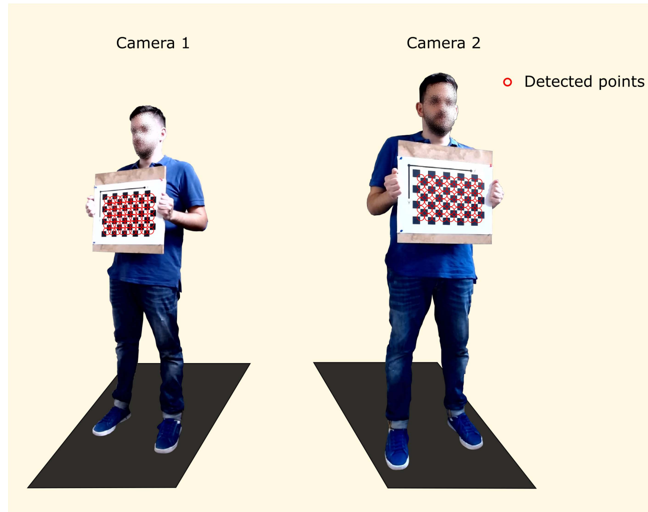


Fig. 2. Example of a paired picture of the checkboard pattern used for the stereo camera calibration.

Each confidence map is a 2D representation of the probability that a particular body part occurs at each pixel's location. Furthermore, for each pixel OP provides a 2D vector representing the position and orientation of the limb. Finally, the confidence maps and the affinity fields are associated by greedy inference to output the 2D keypoints for all the people in the image. The CNN is trained on COCO (Common Objects in Context) keypoint dataset [49] plus a foot dataset composed of 15,000 annotations out of the COCO dataset. These datasets include RGB images representing several real-world scenarios [28]. The 2D data provided by OP from each video image were triangulated using the calibration outcomes, in order to obtain the 3D cartesian coordinates of anatomical landmarks with respect to the reference system of one camera. This computation was performed using the Matlab Computer Vision System Toolbox Matlab (vR2016b, The Mathworks Inc., Natwick, USA).

E. Manual Camera Calibration Validation

A 3D joint is placed on a ray going from each camera center of projection to the location of the point on the image plane. The 3D joint coordinates are provided by the intersection of each camera ray. Due to the error in estimating the camera intrinsic and extrinsic parameters, the two back-projected rays do not, in general, intersect exactly at the real joint position. Therefore, the joint location is typically approximated by finding the midpoint between rays [50], [51]. The accuracy in 3D coordinates estimation is based both on OP ability in detecting body joints position on the two image planes and on the errors in the camera model. If the two rays from the camera centers of projection to the target point are parallel, they cannot intersect and no solution can be found. If the two cameras are placed perpendicularly, the triangulation error is negligible [52] but the target cannot be viewed from the two cameras. Based on Rahimian and Kearney [53], to assure that all the joints were visible by the two webcams with camera rays sufficiently non-parallel, we selected the convergence angle between the two cameras in a range from 40° and 50°, minimizing also the triangulation errors. Indeed, when 2D coordinates provided

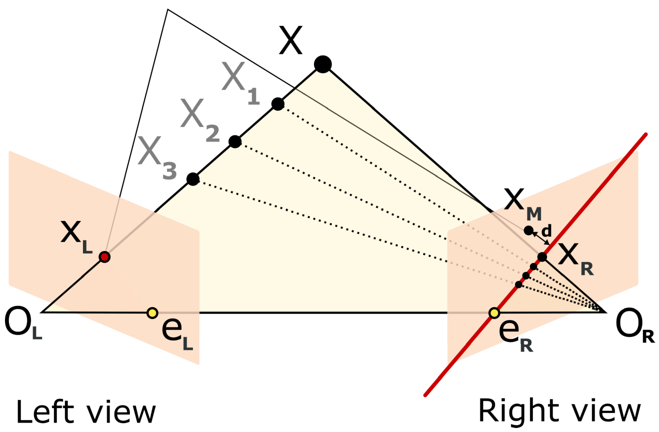


Fig. 3. Epipolar geometry and reprojection error *d*. X is the point in the 3D world; X_R and X_L are the real point position on the two camera planes; X_m represents the point provided by OP from one camera. e_R *and* e_L represents the epipolar lines.

by OP from one camera do not match with the ones obtained from the other camera, linear triangulation fails. As a result, we have an error in the detection of body joints coordinates that can be characterized in terms of the reprojection error (d), computed as the distance between the matched point (X_m) provided by OP and its epipolar line (e_R) obtained by the fundamental matrix (Fig.3). The manual camera calibration validation was performed firstly estimating the accuracy in the measurement of anatomical landmarks positions and then differentiating the contribution of triangulation errors. This evaluation has been performed on a single subject since it depends on the abilities of OP, which is trained on monocular images with different scenarios and it has been shown to maintain efficiency with several body poses, even as the number of people increases [29]. The accuracy in body joint detection was defined as the Root Mean Square Error (RMSE), in a complete gait cycle, of the reprojection error d:

$$\varepsilon_{ep} = RMSE\,(d) \tag{1}$$

The pixel coordinates of the top left vertex of the IMU sensor placed on mid pelvis and on right shank were manually acquired, in the BB camera configuration, from the frames of a complete gait cycle to analyze the amount of error due the stereoscopic reconstruction.

F. Kinematic Accuracy

The estimation of OP kinematic accuracy consists in a first computation of 3D joint angles, which represent the rotations between two body segments, encoded in the joint rotation matrices. Given two body segments b_i and b_j and the ground reference frame $\mathrm{CS_g}$, we can express the rotation matrix ${}^{b_j}\boldsymbol{R}_{b_i}$ between the coordinate systems CS_{b_i} and CS_{b_j} relative to the two segments, as follows:

$${}^{b_i}\boldsymbol{R}_{b_j} = \left({}^{g}\boldsymbol{R}_{b_i}\right)^{T}\,{}^{g}\boldsymbol{R}_{b_j} \tag{2}$$

The 3D joints coordinates, obtained after the triangulation of the 2D points provided by OP, were originally expressed in a reference system CS_c located in the optical center of one of the cameras, oriented as the camera itself. This coordinate system

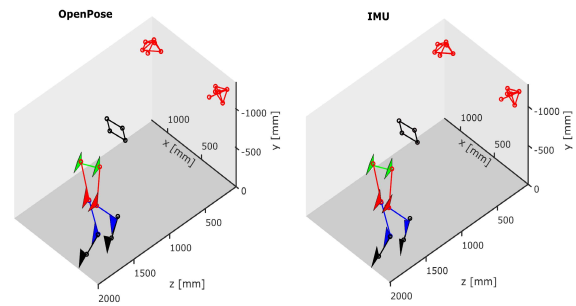


Fig. 4. Example of the biomechanical models, obtained for OP (Left) and IMUs (Right) data, for a generic video frame n, for the camera configuration BB, in a subject. In black is represented the checkboard pattern while in red the two webcams. Checkboard and cameras.

was moved to the new reference frame CS_g, having the origin to the ground level and the X-Z axes parallel to the ground and directed as those of the IMUs system, aligning OP and IMUs reference systems. In this way, the rotation matrices ${}^{g}\boldsymbol{R}_{b_i}^{OP}$ and ${}^{g}\boldsymbol{R}_{b_j}^{OP}$ relative to the two body segments b_i and b_j can be computed in the ground reference system ${CS_g}^{OP}$. Then, the rototraslation matrices ${}^{g}\boldsymbol{H}_{b_i}^{OP}$ and ${}^{g}\boldsymbol{H}_{b_j}^{OP}$ were computed considering the transformation among the coordinate systems $CS_{b_i}^{OP}$ and $CS_{b_j}^{OP}$. In addition, the x, y and z axes of the reference frames $CS_{b_i}^{OP}$ relative to the body segment of pelvis, thigh and shank were defined as follows:

- ${}^{g}\boldsymbol{y}_{b_i}^{OP}$ axis parallel to the line joining the two joints that bounded the body segment, pointing upwards;
- the plane ${}^{g}\boldsymbol{y}z_{b_i}^{OP}$, parallel to the frontal plane, with ${}^{g}\boldsymbol{z}_{b_i}^{OP}$ axis pointing rightward.

Finally, the foot segment reference frame was defined with the ${}^{g}\boldsymbol{x}_{b_i}^{OP}$ axis parallel to the line joining the two points that bounded the foot segment and the ${}^{g}\boldsymbol{x}z_{b_i}^{OP}$ plane parallel to the transverse plane.

The rotation matrices ${}^{g}\boldsymbol{R}_{b_i}^{IMU}$ and ${}^{g}\boldsymbol{R}_{b_j}^{IMU}$ related to the IMUs system were computed by applying an already validated biomechanical model [45]. The biomechanical models obtained for OP and IMUs data, for a generic video frame n, for the camera configuration BB, in a single subject is shown in Fig. 4.

From the rotation matrices ${}^{b_j}\boldsymbol{R}_{b_i}^{IMU}$ and ${}^{b_j}\boldsymbol{R}_{b_i}^{OP}$ we computed the trajectory of the sagittal joint angle, which is the angle between two body segments projected into the walking plane ($\boldsymbol{z}^{g}\boldsymbol{y}^{g}$). Although the system is capable of providing joint angles along the three anatomical planes, the transverse and frontal ones were not examined since during walking angular rotations for lower limbs occur primarily in the sagittal plane [54]. From the raw data of OP and IMUs, sagittal angles were calculated for hip, knee and ankle joints of both lower limbs (Fig. 5). No kind of filter was applied at any computation step. Since a physical trigger for the synchronization of the cameras with the IMUs system was not available, the trajectories provided by the two systems were synchronized by software using a cross-correlation algorithm, which detected the lag between two signals. Each signal was further partitioned into strides, following the algorithm proposed by Salarian *et al.* [55], which consisted in the

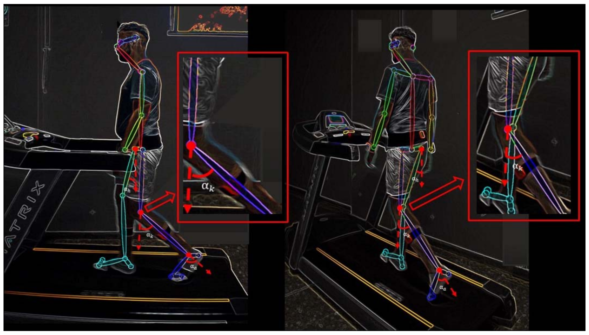


Fig. 5. Skeleton tracking provided by OP and computed angles in a single frame of the video recorded by two cameras in the configuration LL.

identification of the heel strike instant by analyzing the shank angular velocity detected by the IMU sensor. Then, a gait cycle was defined as the time interval between two consecutive heel strikes of the same foot. The first and the last strides were discarded from the successive analysis to guarantee the stabilization of subject walking speed, avoiding effects due to the acceleration and deceleration phase [56]. A total of 15 gait cycles were analyzed for the walking activity and of 20 for the running. Such values are in line with the findings proposed by previous authors who assessed 15-20 strides as robust enough to guarantee a reliable gait/running analysis [57], [58].

The kinematic accuracy of the system composed by OP, two webcams and a linear triangulation algorithm was evaluated considering the following gait parameters for each gait cycle: (i) absolute maximum (MAX) of angles, as indicator of the joint peak flexion; (ii) absolute minimum (MIN) of angles, as indicator of the joint peak extension; (iii) range of motion (ROM) of joint angles. The mentioned parameters were determined, using a custom Matlab script for OP and IMUs and then compared. OP performance in kinematic evaluation was quantified as the mean absolute error (ε), across the N strides, of the differences between joint angle parameters, related to each gait cycle as follows:

$$\varepsilon_{MAX} = \sum_{i=1}^{N} \frac{abs(MAX_{OP} - MAX_{IMU})}{N} \tag{3}$$

$$\varepsilon_{MIN} = \sum_{i=1}^{N} \frac{abs(MIN_{OP} - MIN_{IMU})}{N} \tag{4}$$

$$\varepsilon_{ROM} = \sum_{i=1}^{N} \frac{abs(MAX_{ROM} - MAX_{ROM})}{N} \tag{5}$$

Furthermore, we calculated the signed error as the mean over the gait cycles of the signed difference between the joint angles parameters computed by the two systems. It indicated the tendency of OP system to overestimate or underestimate the gait parameters, and it was defined as follows:

$$\varepsilon_{signed\_MAX} = \sum_{i=1}^{N} \frac{MAX_{OP} - MAX_{IMU})}{N} \tag{6}$$

$$\varepsilon_{signed\_MIN} = \sum_{i=1}^{N} \frac{MIN_{OP} - MIN_{IMU}}{N} \tag{7}$$

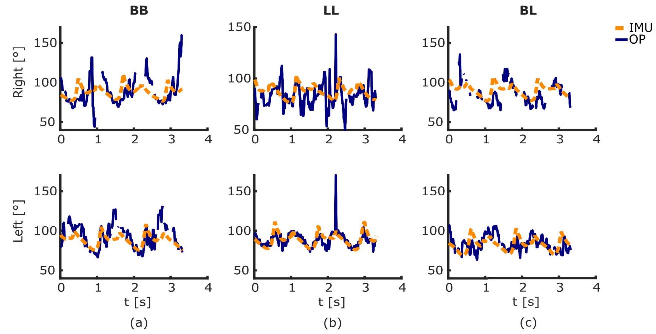


Fig. 6. Angular trajectories over time for the right and left ankle, in the three camera configurations and for the walking activity, in a single subject. OP displacements appear as solid blue lines and IMUs displacements appear as dotted orange lines. Ankle joint coordinates were lost in a wide number of frames.

$$\varepsilon_{signed\_ROM} = \sum_{i=1}^{N} \frac{MAX_{ROM} - MAX_{ROM}}{N} \tag{8}$$

Normality of the data was assessed using Shapiro-Wilk test, and sphericity condition for repeated measures analyses of variance (rANOVA) was assessed using the Mauchly test. The two-way rANOVA test was used to examine the effects, on the dependent variables (ε_{MAX}, ε_{MIN}, ε_{ROM}), averaged across the strides, of different webcam configurations and movement speeds, using as within-subjects factors: (i) 'camera position' (3 levels: BB, BL, LL) and 'activity' (2 levels: walking and running) and their interaction, independently for each examined joint. A post-hoc analysis was performed using the paired t-tests to evaluate the significant pairwise differences between each webcam position and activity. For all tests, the level of statistical significance was set at 0.05, except for post-hoc analysis, where the significance level was reduced at 0.008, for Bonferroni correction. Statistical analysis was conducted by using IBM SPSS Statistics 23.

III. Results

For all the camera configurations and for the two examined activities, OP was not able to record reliable ankle trajectories. In particular, for the BB and BL camera configuration, the system lost foot coordinates for most frames (Fig. 6 a–6c), while for the LL configuration the high variability precluded the detection of the flexo-extension peaks (Fig. 6b). A similar result was obtained also for the running activity.

Our results are limited to the analysis of the hip and knee joints, for which OP recorded angular displacement representative of normal gait, similar to the one computed by IMUs, but more variable than IMUs displacements. Fig. 7 shows, as example, the right hip and knee angles obtained in the camera configuration BB, for a single subject.

A. Over-and Under-Estimation of the Gait Parameters

Table I reports the mean differences (ε_{signed_MAX}, ε_{signed_MIN}, ε_{signed_ROM}) between the computed gait parameters, considering their sign, averaged among the subjects and the related standard deviations.

For the **right knee**, in all camera configuration and locomotion activities, we found the following outcomes: (i) the peak

TABLE I
SIGNED ERRORS FOR THE PEAK FLEXO-EXTENSION AND ROM

Signed error (Mean±SD) [°]	ε_{signed_MAX}				ε_{signed_MIN}				ε_{signed_ROM}			
	R_knee	L_knee	R_hip	L_hip	R_knee	L_knee	R_hip	L_hip	R_knee	L_knee	R_hip	L_hip
BB_{LOW}	**-4.2 ± 3.0**	**-4.5 ± 3.8**	**-1.0 ± 2.0**	**-1.9 ± 3.3**	**3.0 ± 4.9**	**1.9 ± 5.9**	**0.5 ± 3.7**	**0.3 ± 2.3**	**-7.2 ± 6.8**	**-6.3 ± 6.8**	**-1.5 ± 5.7**	**-2.2 ± 5.3**
BL_{LOW}	**-3.0 ± 3.3**	**-2.3 ± 3.4**	**0.9 ± 3.5**	**0.3 ± 7.3**	**4.0 ± 5.7**	**2.2 ± 2.0**	**0.8 ± 2.3**	**3.2 ± 3.1**	**-7.0 ± 8.7**	**-4.6 ± 4.3**	**0.1 ± 4.3**	**-2.8 ± 9.8**
LL_{LOW}	**0.0 ± 4.4**	**-1.2 ± 3.5**	**1.0 ± 6.0**	**-1.4 ± 4.2**	**2.1 ± 4.6**	**1.9 ± 3.1**	**0.7 ± 5.7**	**1.8 ± 2.4**	**-2.0 ± 8.9**	**-3.1 ± 5.4**	**0.3 ± 11.7**	**-3.2 ± 6.3**
BB_{HIGH}	**-1.3 ± 4.3**	**1.4 ± 2.9**	**-0.3 ± 6.4**	**-2.6 ± 5.7**	**1.9 ± 3.1**	**-1.9 ± 2.8**	**1.9 ± 6.2**	**3.4 ± 5.5**	**-3.3 ± 7.4**	**3.4 ± 5.5**	**-2.2 ± 12.3**	**-6.0 ± 11.2**
BL_{HIGH}	**-2.4 ± 5.2**	**2.3 ± 4.7**	**-3.8 ± 1.8**	**-2.0 ± 3.1**	**3.3 ± 2.9**	**-1.8 ± 4.5**	**2.0 ± 2.3**	**-0.1 ± 5.3**	**-5.7 ± 8.1**	**4.1 ± 8.6**	**-5.8 ± 3.9**	**-1.8 ± 7.6**
LL_{HIGH}	**-2.3 ± 3.9**	**4.4 ± 7.8**	**-2.3 ± 3.8**	**-0.8 ± 4.4**	**1.1 ± 3.7**	**-1.7 ± 4.5**	**-0.7 ± 5.8**	**-0.3 ± 3.6**	**-3.4 ± 7.5**	**6.1 ± 11.2**	**-1.6 ± 8.6**	**-0.5 ± 7.2**

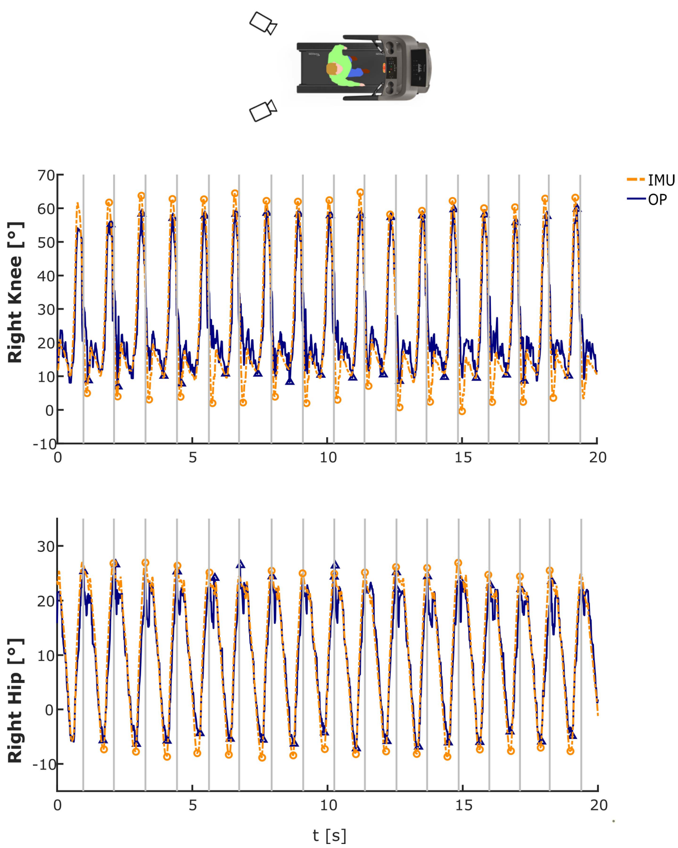


Fig. 7. Example of sagittal angles for the right hip and knee, for the BB camera configuration, during the walking activity, in a single subject. OP angle trajectory appears as solid blue lines while IMU trajectory appear as dotted orange lines. Small triangles highlight OP peak flexion and extension data, while small circles highlight IMU peak flexion and extension data. Vertical grey lines highlight gait cycles.

flexion values measured by OP resulted smaller than those measured by IMUs; (ii) the peak extension values, instead, measured by OP for the right knee, resulted larger than the one obtained by IMUs; (iii) the right knee ROM computed by OP resulted lower than the one evaluated by IMUs.

For the **left knee**, the peaks flexion and the ROM provided by OP were lower than the ones computed by the Xsens system, during the walking activity, while they were higher during the running. On the contrary, the peak extension measured by OP for the left knee, was lower than the one estimated by IMUs during the running activity, while it was larger during the walking.

A different result was found for the **right** and **left hip**: OP overestimated the peak extension in all configurations with the exception of LL in the running activity, for the right hip, and BL and LL in the run activity, for the left hip. Furthermore, the mean peak flexion and the ROM obtained for the **right hip** was always lower than the one measured by IMUs with the exception of the camera configuration LL and BL for the walking activity. The maxima computed by OP for the **left hip** were always smaller than the ones computed by IMUs with the exception of the camera configuration BL in the walking activity. Finally, OP underestimated the ROM, for the left hip, in all camera configurations and speeds.

B. Kinematic Accuracy

In order to investigate the effects of camera configurations and walking speeds on the kinematic accuracy (ε_{MAX}, ε_{MIN}, ε_{ROM}) a statistical analysis was performed.

As regards the **right knee** peak flexion, rANOVA showed a significant effect of the camera position ('camera position' effect: $F = 20.683$, $p<0.001$) and its interaction with the locomotion activity ('camera position∗activity' interaction effect: $F = 31.091$, $p<0.001$). In the case of the right knee peak extension, we found a significant effect of the camera position ('camera position' effect: $F = 43.323$, $p<0.001$), the locomotion activity ('activity' effect: $F = 23.353, p = 0.005$), and their interaction ('camera position∗activity' interaction effect: $F = 8.068, p = 0.008$). For the right knee ROM, rANOVA showed a significant difference of the camera position ('camera position' effect: $F = 23.306$, $p<0.001$), and its interaction with the locomotion activity ('camera position∗activity' interaction effect: $F = 37.660, p = 0.001$). A similar result has been found for the **left knee**, where rANOVA showed: (i) for the peak flexion, a significance difference of camera position ('camera position' effect: $F = 8.280$, $p = 0.008$), and its interaction with the locomotion activity ('camera position∗activity' interaction effect: $F = 94.720$, $p < 0.001$); (ii) for the peak extension, a significant effect of camera position ('camera position' effect: $F = 6.186$, $p = 0.018$) and activity ('activity' effect: $F = 34.695$, $p = 0.002$). In the case of left knee ROM we found an effect of the camera position ('camera position' effect: $F = 32.936$, $p<0.001$), and its interaction with the walking speed ($F = 17.253$, $p = 0.001$).

In the case of the **right hip** peak flexion, we found a significative effect of the camera position ('camera position' effect: $F = 17.611$, $p = 0.001$), the locomotion activity ('activity' effect: $F = 21.680$, $p = 0.006$), and their interaction ('camera position∗activity' interaction effect: $F = 22.014$, $p < 0.001$). For the hip extension only the interaction between camera position and activity showed a significant difference ('camera position∗activity' interaction effect: $F = 6.371$, $p = 0.016$). In the case of the right hip ROM we found a significant effect of camera position ('camera position' effect: $F = 6.120$, $p = 0.018$), and its interaction with the locomotion activity ('camera position∗activity' interaction effect: $F = 5.281$, $p = 0.027$). For the **left hip** peak flexion rANOVA showed a significant difference between the camera configurations ('camera position' effect: $F = 8.280$, $p = 0.008$), and the speed interaction with the camera position ('camera position∗activity' interaction effect: $F = 94.720$, $p < 0.001$). In addition, for the left hip extension peak we found significant differences between the camera configuration ('camera position' effect: $F = 6.186$, $p = 0.018$), and the movement speed ('activity' effect: $F = 34.695$, $p = 0.002$). Finally, for the left hip ROM rANOVA showed a significant difference between the camera configurations ('camera position' effect: $F = 13.329$, $p = 0.002$), the movement speeds ('activity' effect: $F = 44.686$, $p = 0.001$), and a significant effect of the interaction between camera and activity ('camera position∗activity' interaction effect: $F = 32.747$, $p < 0.001$).

Figures 8, 9 and 10 show the kinematic accuracy obtained for the peak flexion, peak extension and ROM, respectively, and the statistical differences obtained from the post-hoc analysis among the different camera configurations and locomotion activities, for each examined joint.

The paired t-tests results showed that the camera configurations influenced the kinematic accuracy in the measurement of flexion/extension peak and ROM of the **right knee**. In particular, the mean absolute error resulted lower in the LL configuration during the walking activity ($\varepsilon_{MAX} = 3.1 \pm 0.4°$, $\varepsilon_{MIN} = 2.5 \pm 0.5°$, $\varepsilon_{ROM} = 4.5 \pm 0.33°$), while it was larger in the BL configuration during the run activity ($\varepsilon_{MAX} = 6.1 \pm 0.7°$, $\varepsilon_{MIN} = 5.5 \pm 0.2°$, $\varepsilon_{ROM} = 10.3 \pm 1.7°$).

A similar result has been found for the **left knee** peak flexion, where the mean absolute error decreased in the LL configuration ($\varepsilon_{MAX} = 3.3 \pm 1.0°$) during the walking activity, while it increased in the BL configuration during the run activity ($\varepsilon_{MAX} = 5.6 \pm 0.7°$). Furthermore, in the BL and LL positions we found a significant difference between the two movement activities, with a larger error in the case of the run, for the BL camera configuration ($p = 0.005$), and during

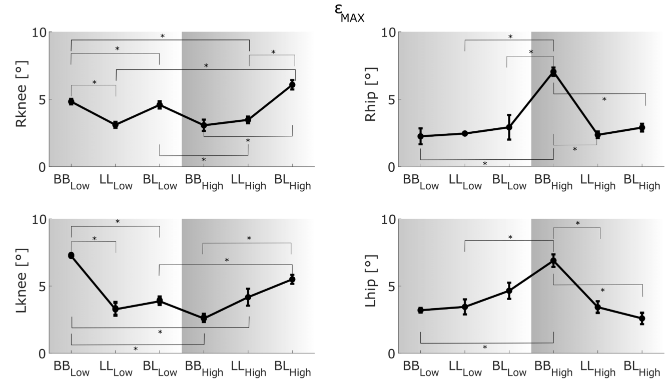


Fig. 8. Mean absolute error for the peak flexion values of the knee and hip, for both left and right side, obtained by OP system in the different camera configurations and movement speeds. Statistical differences between OP and IMUs pairs are starred.

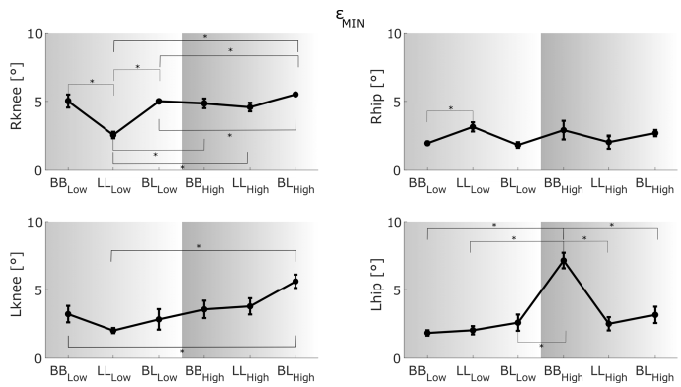


Fig. 9. Mean absolute error for the peak extension values of the knee and hip, for both left and right side, obtained by OP system in the different camera configurations and movement speeds. Statistical differences between OP and IMUs pairs are starred.

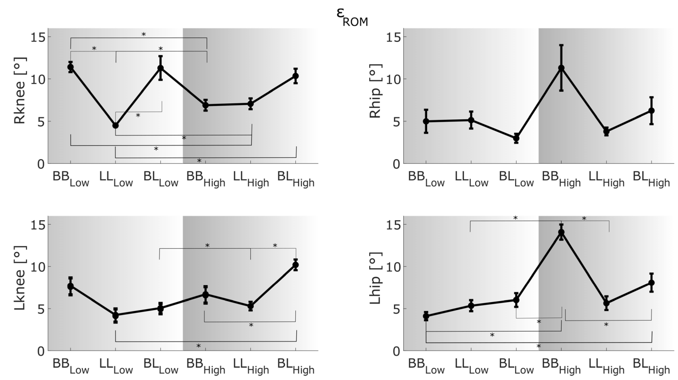


Fig. 10. Mean absolute error for the ROM values of the knee and hip, for both left and right side, obtained by OP system in the different camera configurations and movement speeds. Statistical differences between OP and IMUs pairs are starred.

the walking for the BB camera configuration ($p<0.001$). The left knee ε_{ROM} also increased in the BL configuration, during the running activity up to 10.1 ± 1.3 °.

A different result has been found for the **right hip** peak flexion, where the largest error has been found for the camera configuration BB, during the running activity ($\varepsilon_{MAX} = 7.04 \pm 0.76°$). The kinematic accuracy significantly improved respect

TABLE II
LOWER AND HIGHER ACCURACY VALUES FOR EACH GAIT PARAMETER

	Accuracy [°]		Joint	Camera	Locomotion Activity
MAX	Lower	7.3 ± 0.2	L_knee	BB	Walking
	Higher	2.2 ± 1.2	R_hip	BB	Walking
MIN	Lower	7.2 ± 1.2	L_hip	BB	Running
	Higher	1.6 ± 0.3	R_hip	BL	Walking
ROM	Lower	14.1 ± 1.8	L_hip	BB	Running
	Higher	2.8 ± 1.0	R_hip	BL	Walking

to the BB camera position, up to 2.9 ± 0.6° in the BL configuration ($p < 0.001$) and 2.3 ± 0.5° in the LL one ($p < 0.001$). On the contrary, during the walking activity, the right hip peak flexion error in the BB configuration decreased in respect to the running activity up to 2.2 ± 1.2° ($p < 0.001$). For the right hip peak extension in the LL configuration, the kinematic accuracy was found to increase than the BB configuration ($p = 0.003$), during the low-speed movement up to 3.1 ± 0.7°. Instead in the case of the right hip ROM measurement, no statistical differences were found.

The largest error has been found for the BB configuration, in the measurement of the **left hip** ROM, during the running. In this case, the kinematic accuracy significantly decreased up to 14.1 ± 1.8°. A similar trend has been found for the left hip peak extension and flexion, with an absolute mean error larger in the BB configuration, during the running activity.

Considering all camera configuration and locomotion activities the markerless system we developed showed an absolute error ranging from 1.6 ± 0.3° to 14 ± 1.8°. The lower and higher accuracy values for each gait parameter are shown in the Table II.

C. Accuracy in Body Joints Detection

The results related to OP accuracy in detecting body joints position on the two image planes and on the errors in the camera model showed that the RMSE for the top left vertex of the IMU on the mid pelvis, in the BB configuration was equal to 1.28 pixel, which are equivalent to 0.19 cm. The RMSE error for the top left vertex of the IMU placed on the right shank was equal to 11.0 pixels, which are equivalent to 1.65 cm. The RMSE error for the points provided by OP was equal to 21.7, 9.9, 15.9, 10.6 pixel, which are equivalent to 3.26, 1.49, 2.39, 1.59 cm, for right knee, right hip, left knee and left hip respectively. Fig. 11 shows the distances between the epipolar lines and the detected points of the IMUs on the pelvis and on the right shank, for a specific frame.

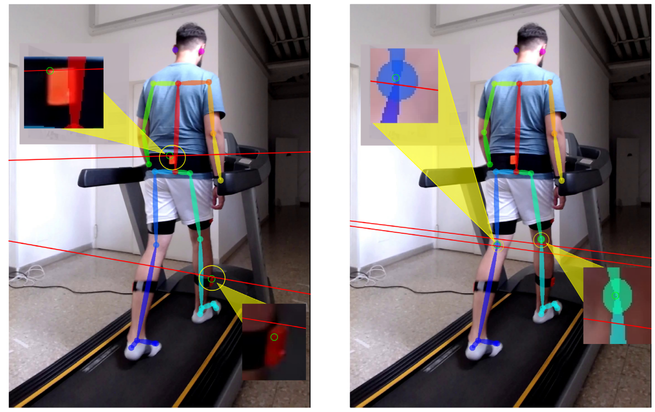


Fig. 11. Epipolar lines related to the pixel coordinates of the top left vertex of the IMU placed on the mid pelvis and on the right shank.

IV. DISCUSSION

In this study, a low-cost markerless system, consisting in two RGB cameras and a linear triangulation algorithm, has been developed and validated in terms of lower limb gait analysis by means of an IMU system, used as reference. In general, the angular traces provided by OP were similar to the ones computed by IMUs, in terms of waveform, but system inaccuracy affected mostly the computation of the gait parameters producing error up to 14 ± 1.8°. In addition, the system was not able to track foot coordinates for the most of frames and for all the camera positions.

Previous studies on markerless gait analysis resulted in errors generally larger in the ankle than the other joints [24], [59]. Since the ankle angular displacement resulted from the relative rotation between the foot and the shank, the inability to track this joint depended on the following factors: (i) the foot segment was hardly visible by the cameras, being smaller than the pelvis, thigh and tibia segments, used for the computation of the knee and hip angles; (ii) the angle between feet and camera direction did not allow a good triangulation. This problems could be solved by using additional cameras focusing exclusively on the subject's feet, as it was shown by Surer *et al.* [60], who highlighted that a single camera is sufficient to estimate the sagittal ankle angle of one leg, if the camera frames the sagittal view of the shank and the foot.

Our findings confirm the one reported by Zago *et al.* [33] who evidenced the limitations of OP when used to compute spatiotemporal gait parameters, using two RGB-cameras. They highlighted how the occlusion of body joints can lead to an erroneous reconstruction of 3D body joints, since the estimation of the hidden landmarks, provided by OP produce large errors in the 3D pose reconstruction. In our work we demonstrated that OP computed mismatched points on the two image planes, by means of a manual camera calibration. In the BB configuration, we found that the RMSE related to the known point on the pelvis was equal to 0.19 cm, since this point was properly framed by the cameras during the whole task execution. Considering also that all the paired pictures for which the reprojection error was greater than 0.23 cm were removed from the stereo camera calibration inputs, the RMSE error obtained for the right and left hip joints, depended mainly on the mismatch between the points provided by OP on the two image planes. The point on the right shank led to a lower triangulation accuracy since it was hidden to the cameras during several gait phases. In this case, the amount of error due to the stereoscopic reconstruction added to the one due to OP mismatching, led to a decrease of the accuracy in body joints detection up to 3.26 cm. Hence, the decrease of kinematic accuracy was primarily due the lack of a robust correspondence between points on two perspective images that caused the wrong reconstruction of 3D coordinates [61]. The error in the estimation of the body landmarks could be also related to the use of CNN for the classification of the joint center positions, which implies the absence of a model for the definition of anatomical points. In fact, OP detects the anatomical landmarks by predicting a set of 2D confidence maps of body part locations and a set of 2D vector fields that encode the location and orientation of limbs over the image domain [28]. The joint computation is based on the probability that a particular body part occurs at each pixel's location. The limitation due to the lack of the body modelling was highlighted in previous studies which showed how the wrong localization of body landmarks could lead to an error up to 13° in the orientation of the anatomical axes with an overestimation of joint angles [62]–[64].

The kinematic accuracy of our system depended on the camera configuration, which influenced the visibility of body joints during the task, and on the locomotion activity. The lower error with respect to the reference value has been found for the BL webcam position and the walking activity, in the measurement of the minimum amplitude of the right hip angle. Although the BL was the optimal arrangement for the right hip joint, it resulted in the worst configuration for several conditions, such as the right knee flexion peak. Overall, the LL configuration led to the highest performance for several conditions with an absolute mean error ranging from $2.0 \pm 0.4°$ to $7.1 \pm 1.3°$. In this configuration cameras pointed to the left side of the subject and consequently his right joints were not always visible to the cameras. It is clear that different camera positions cannot improve the overall kinematic accuracy of the system but act according to the individual considered joint. Considering that most of the markerless systems use more than 7 cameras [14], [24], increasing the number of cameras used simultaneously could improve the overall accuracy. To overcome the limits related to the occlusion of body joints, the use of 4 cameras (two pointing to the right side of the subject and two pointing to the left side of the subject) could allow a complete kinematic gait analysis. The system here tested is unsuitable for the overall measurement of joints flexion or extension, since for gait analysis, error values greater than 5° mean a significant alteration of kinematic information [65]. However, our results showed that the OP-based system is suitable for all those applications where the assessment of a specific lower limb joint angle is needed, such as knee rehabilitation after stroke or trauma injuries [66], [67]. To address this issue, further studies integrating triangulation matching algorithms and a larger number of cameras are needed to improve the overall accuracy, in a way that opens up the possibility to perform a complete quantitative gait/running analysis. The angular error obtained in our study is generally lower to the one obtained by Pfister *et al.* [18], who investigated the accuracy and repeatability of Microsoft Kinect in performing a body tracking, by a comparison with an optoelectronic system. They showed that Kinect performance decreased when the walking speed increased. In our study, the locomotion activity influenced the performance of the OP-based system. In the BB webcam positions the run activity led to a lower accuracy in the peak extension, flexion and ROM estimation of the left hip, and the peak flexion of the right hip, in particularly underestimating the full range of motion. This result, which could be related to the decrease of the number of frames recorded for each gait cycle ($\approx$ 80 frames, for the walking; $\approx$ 40 for the running), has been found also by Castelli *et al.* [42] who highlighted the importance of a sufficiently high frame rate in order to obtain reliable data in a bidimensional markerless gait analysis.

The simplicity of the setup presented here, which is mainly based on commercial webcams, offers the opportunity to use the system also outside a gait laboratory in free walking, running or exercise activities without the need of a treadmill. Since kinematic patterns during overground gait have been showed to be similar to the ones obtained on a treadmill, in both walking [68] and running [69] activities, we can assume that our results would not change during experimental sessions performed in free walking, within the field of view of the cameras.

V. Conclusion

This study aimed to perform a metrological characterization of a novel low-cost markerless system composed by OpenPose, two webcams and a linear triangulation algorithm, in the context of quantitative gait analysis.

Our outcomes showed that the observed differences between the OP-based system and IMUs depended on the ability of OP in measuring the joint center coordinates, the locomotion activity and the camera position that can lead to the occlusion of the body joints during the gait cycle. Two cameras in the BL or LL configuration can provide a sufficient accuracy for those applications, like videogames or virtual reality rehabilitative trials, in which visual information should change in real time according to the movements of user's body. The BB position, on the contrary, resulted the worst condition, with the highest underestimation of left range of motion, and it cannot be used for a complete real time pose estimation of all the lower limbs joints. Although the BB configuration complied with the convergence angle condition, so that the camera rays could properly intersect, the camera view point led to the occlusion of body joints in multiple stages of the gait cycles, resulting in a higher inaccuracy in the stereoscopic reconstruction.

The system here presented could be useful for assessing a specific lower limb joint angle or for human-body tracking in those applications where a real-time body tracking is required, i.e. video games, entertainment setups or virtual reality rehabilitative trials. Nevertheless, the current system accuracy is not adequate for a complete human body kinematic analysis.

Further studies with the integration of matching algorithms and a higher number of cameras are necessary to improve accuracy, which is intrinsically poor in markerless systems.

References

[1] U. D. 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, no. 2, pp. 226–237, 2005.

[2] R. Baker, "Gait analysis methods in rehabilitation," *J. Neuroeng. Rehabil.*, vol. 3, no. 1, p. 4, 2006.

[3] M. W. Whittle, "Clinical gait analysis: A review," *Hum. Movement Sci.*, vol. 15, no. 3, pp. 369–387, 1996.

[4] S. Armand, F. Moissenet, G. de Coulon, and A. Bonnefoy-Mazure, "Identifying and understanding gait deviations: Critical review and perspectives," *Movement Sport Sci.-Sci. Motricité*, vol. 98, no. 98, pp. 77–88, 2017.

[5] C. Schiefer *et al.*, "Optimization of inertial sensor-based motion capturing for magnetically distorted field applications," *J. Biomech. Eng.*, vol. 136, no. 12, Dec. 2014, Art. no. 121008.

[6] S. Kim and M. A. Nussbaum, "Performance evaluation of a wearable inertial motion capture system for capturing physical exposures during manual material handling tasks," *Ergonomics*, vol. 56, no. 2, pp. 314–326, Feb. 2013.

[7] X. Robert-Lachaine, H. Mecheri, C. Larue, and A. Plamondon, "Validation of inertial measurement units with an optoelectronic system for whole-body motion analysis," *Med. Biol. Eng. Comput.*, vol. 55, no. 4, pp. 609–619, Apr. 2017.

[8] S. J. Piazza and P. R. Cavanagh, "Measurement of the screw-home motion of the knee is sensitive to errors in axis alignment," *J. Biomech.*, vol. 33, no. 8, pp. 1029–1034, Aug. 2000.

[9] A. Cappozzo, U. D. Croce, A. Leardini, and L. Chiari, "Human movement analysis using stereophotogrammetry: Part 1: Theoretical background," *Gait Posture*, vol. 21, no. 2, pp. 186–196, 2005.

[10] P. Picerno, A. Cereatti, and A. Cappozzo, "Joint kinematics estimate using wearable inertial and magnetic sensing modules," *Gait Posture*, vol. 28, no. 4, pp. 588–595, Nov. 2008.

[11] M. Iosa, P. Picerno, S. Paolucci, and G. Morone, "Wearable inertial sensors for human movement analysis," *Expert Rev. Med. Devices*, vol. 13, no. 7, pp. 641–659, Jul. 2016.

[12] A. I. Cuesta-Vargas, A. Galán-Mercant, and J. M. Williams, "The use of inertial sensors system for human motion analysis," *Phys. Therapy Rev.*, vol. 15, no. 6, pp. 462–473, Dec. 2010.

[13] P. Picerno, A. Cereatti, and A. Cappozzo, "A spot check for assessing static orientation consistency of inertial and magnetic sensing units," *Gait Posture*, vol. 33, no. 3, pp. 373–378, Mar. 2011.

[14] M. Sandau, H. Koblauch, T. B. Moeslund, H. Aanæs, T. Alkjær, and E. B. Simonsen, "Markerless motion capture can provide reliable 3D gait kinematics in the sagittal and frontal plane," *Med. Eng. Phys.*, vol. 36, no. 9, pp. 1168–1175, Sep. 2014.

[15] S. Corazza, L. Mündermann, E. Gambaretto, G. Ferrigno, and T. P. Andriacchi, "Markerless motion capture through visual hull, articulated ICP and subject specific model generation," *Int. J. Comput. Vis.*, vol. 87, nos. 1–2, p. 156, 2010.

[16] E. Knippenberg, J. Verbrugghe, I. Lamers, S. Palmaers, A. Timmermans, and A. Spooren, "Markerless motion capture systems as training device in neurological rehabilitation: A systematic review of their use, application, target population and efficacy," *J. Neuroeng. Rehabil.*, vol. 14, no. 1, pp. 1–11, Dec. 2017.

[17] S. L. Colyer, M. Evans, D. P. Cosker, and A. I. T. Salo, "A review of the evolution of vision-based motion analysis and the integration of advanced computer vision methods towards developing a markerless system," *Sports Med.*, vol. 4, no. 1, p. 24, Dec. 2018.

[18] A. Pfister, A. M. West, S. Bronner, and J. A. Noah, "Comparative abilities of microsoft Kinect and vicon 3D motion capture for gait analysis," *J. Med. Eng. Technol.*, vol. 38, no. 5, pp. 274–280, Jul. 2014.

[19] Q. Wang, G. Kurillo, F. Ofli, and R. Bajcsy, "Evaluation of pose tracking accuracy in the first and second generations of microsoft Kinect," in *Proc. Int. Conf. Healthcare Informat.*, Oct. 2015, pp. 380–389.

[20] R. A. Clark, B. F. Mentiplay, E. Hough, and Y. H. Pua, "Three-dimensional cameras and skeleton pose tracking for physical function assessment: A review of uses, validity, current developments and Kinect alternatives," *Gait Posture*, vol. 68, pp. 193–200, Feb. 2019.

[21] A. Schmitz, M. Ye, G. Boggess, R. Shapiro, R. Yang, and B. Noehren, "The measurement of *in vivo* joint angles during a squat using a single camera markerless motion capture system as compared to a marker based system," *Gait Posture*, vol. 41, no. 2, pp. 694–698, Feb. 2015.

[22] S. Choppin and J. Wheat, "The potential of the Microsoft Kinect in sports analysis and biomechanics," *Sports Technol.*, vol. 6, no. 2, pp. 78–85, May 2013.

[23] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, "A comparison and evaluation of multi-view stereo reconstruction algorithms," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Jun. 2006, pp. 519–528.

[24] S. Corazza, L. Muendermann, 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," *Ann. Biomed. Eng.*, vol. 34, no. 6, pp. 1019–1029, 2006.

[25] K.-M. Cheung, S. Baker, and T. Kanade, "Shape-from-silhouette across time Part II: Applications to human modeling and markerless motion tracking," *Int. J. Comput. Vis.*, vol. 63, no. 3, pp. 225–245, Jul. 2005.

[26] L. Mündermann, S. Corazza, and T. P. Andriacchi, "The evolution of methods for the capture of human movement leading to markerless motion capture for biomechanical applications," *J. Neuroeng. Rehabil.*, vol. 3, p. 6, Mar. 2006.

[27] D. Mehta *et al.*, "VNect: Real-time 3D human pose estimation with a single RGB camera," *ACM Trans. Graph.*, vol. 36, no. 4, pp. 1–14, Jul. 2017.

[28] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime multi-person 2D pose estimation using part affinity fields," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1302–1310.

[29] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "OpenPose: Realtime multi-person 2D pose estimation using part affinity fields," 2018, *arXiv:1812.08008*. [Online]. Available: http://arxiv.org/abs/1812.08008

[30] A. Kendall, M. Grimes, and R. Cipolla, "PoseNet: A convolutional network for real-time 6-DOF camera relocalization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 2938–2946.

[31] A. Viswakumar, V. Rajagopalan, T. Ray, and C. Parimi, "Human gait analysis using OpenPose," in *Proc. 5th Int. Conf. Image Inf. Process. (ICIIP)*, Nov. 2019, pp. 310–314.

[32] X. Gu, F. Deligianni, B. Lo, W. Chen, and G. Z. Yang, "Markerless gait analysis based on a single RGB camera," in *Proc. IEEE 15th Int. Conf. Wearable Implant. Body Sensor Netw. (BSN)*, Mar. 2018, pp. 42–45.

[33] M. Zago, M. Luzzago, T. Marangoni, M. De Cecco, M. Tarabini, and M. Galli, "3D tracking of human motion using visual skeletonization and stereoscopic vision," *Frontiers Bioeng. Biotechnol.*, vol. 8, p. 181, Mar. 2020.

[34] E. D'Antonio, J. Taborri, E. Palermo, S. Rossi, and F. Patane, "A markerless system for gait analysis based on OpenPose library," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (IMTC)*, May 2020, pp. 1–6.

[35] D. Erika, T. Juri, P. Eduardo, R. Stefano, and P. Fabrizio, "Characterization of a low-cost markerless system for 3D gait analysis," in *Proc. IEEE Int. Symp. Med. Meas. Appl. (MeMeA)*, Jun. 2020, pp. 1–6.

[36] E. Scalona *et al.*, "Inter-laboratory and inter-operator reproducibility in gait analysis measurements in pediatric subjects," *Int. Biomech.*, vol. 6, no. 1, pp. 19–33, Jan. 2019.

[37] J. Perry and J. R. Davids, "Gait analysis: Normal and pathological function," *J. Pediatric Orthopaedics*, vol. 12, no. 6, p. 815, 1992.

[38] E. K. Antonsson and R. W. Mann, "The frequency content of gait," *J. Biomech.*, vol. 18, no. 1, pp. 39–47, Jan. 1985.

[39] T. N. Do and Y. S. Suh, "Gait analysis using floor markers and inertial sensors," *Sensors*, vol. 12, no. 2, pp. 1594–1611, Feb. 2012.

[40] M. Gabel, R. Gilad-Bachrach, E. Renshaw, and A. Schuster, "Full body gait analysis with Kinect," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2012, pp. 1964–1967.

[41] J. A. Albert, V. Owolabi, A. Gebel, C. M. Brahms, U. Granacher, and B. Arnrich, "Evaluation of the pose tracking performance of the azure Kinect and Kinect v2 for gait analysis in comparison with a gold standard: A pilot study," *Sensors*, vol. 20, no. 18, p. 5104, Sep. 2020.

[42] A. Castelli, G. Paolini, A. Cereatti, and U. D. Croce, "A 2D markerless gait analysis methodology: Validation on healthy subjects," *Comput. Math. Methods Med.*, vol. 2015, pp. 1–11, Apr. 2015.

[43] F. Petraglia, L. Scarcella, G. Pedrazzi, L. Brancato, R. Puers, and C. Costantino, "Inertial sensors versus standard systems in gait analysis: A systematic review and meta-analysis," *Eur. J. Phys. Rehabil. Med.*, vol. 55, no. 2, pp. 265–280, May 2019.

[44] J.-T. Zhang, A. C. Novak, B. Brouwer, and Q. Li, "Concurrent validation of Xsens MVN measurement of lower limb joint angular kinematics," *Physiol. Meas.*, vol. 34, no. 8, pp. N63–N69, Aug. 2013.

[45] E. Palermo, S. Rossi, F. Marini, F. Patanè, and P. Cappa, "Experimental evaluation of accuracy and repeatability of a novel body-to-sensor calibration procedure for inertial sensor-based gait analysis," *Measurement*, vol. 52, pp. 145–155, Jun. 2014.

[46] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[47] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, Nov. 2000.

[48] Y. LeCun and Y. Bengio, "Convolutional networks for images, speech, and time series," in *The Handbook of Brain Theory and Neural Networks*, vol. 3361, no. 10. Cambridge, MA, USA: MIT Press, 1995, p. 1995.

[49] T.-Y. Lin *et al.*, "Microsoft COCO: Common objects in context," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 740–755.

[50] R. I. Hartley and P. Sturm, "Triangulation," *Comput. Vis. Image Understand*, vol. 68, no. 2, pp. 146–157, 1997.

[51] B. Liu, M. Yu, D. Maier, and R. Männer, "An efficient and accurate method for 3D-point reconstruction from multiple views," *Int. J. Comput. Vis.*, vol. 65, no. 3, pp. 175–188, Dec. 2005.

[52] G. Olague and R. Mohr, "Optimal camera placement for accurate reconstruction," *Pattern Recognit.*, vol. 35, no. 4, pp. 927–944, Apr. 2002.

[53] P. Rahimian and J. K. Kearney, "Optimal camera placement for motion capture systems," *IEEE Trans. Vis. Comput. Graphics*, vol. 23, no. 3, pp. 1209–1221, Mar. 2017.

[54] D. A. Neumann, *Kinesiology of the Musculoskeletal System: Foundations for Rehabilitation*. Amsterdam, The Netherlands: Elsevier Health Sciences, 2013.

[55] A. Salarian *et al.*, "Gait assessment in Parkinson's disease: Toward an ambulatory system for long-term monitoring," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 8, pp. 1434–1443, Aug. 2004.

[56] J. Taborri *et al.*, "Measuring changes in gait kinematics due to walking-related fatigue in patients with multiple sclerosis," in *Proc. IEEE Int. Symp. Med. Meas. Appl. (MeMeA)*, Jun. 2019, pp. 1–6.

[57] D. Kroneberg *et al.*, "Less is more—Estimation of the number of strides required to assess gait variability in spatially confined settings," *Frontiers Aging Neurosci.*, vol. 10, p. 435, Jan. 2019.

[58] A. Tura, L. Rocchi, M. Raggi, A. G. Cutti, and L. Chiari, "Recommended number of strides for automatic assessment of gait symmetry and regularity in above-knee amputees by means of accelerometry and autocorrelation analysis," *J. Neuroeng. Rehabil.*, vol. 9, no. 1, pp. 1–8, 2012.

[59] A. Ong, I. S. Harris, and J. Hamill, "The efficacy of a video-based marker-less tracking system for gait analysis," *Comput. Methods Biomech. Biomed. Eng.*, vol. 20, no. 10, pp. 1089–1095, Jul. 2017.

[60] E. Surer, A. Cereatti, E. Grosso, and U. D. Croce, "A markerless estimation of the ankle–foot complex 2D kinematics during stance," *Gait Posture*, vol. 33, no. 4, pp. 532–537, Apr. 2011.

[61] Z. Zhang, R. Deriche, O. Faugeras, and Q.-T. Luong, "A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry," *Artif. Intell.*, vol. 78, nos. 1–2, pp. 87–119, Oct. 1995.

[62] U. D. Croce, A. Cappozzo, and D. C. Kerrigan, "Pelvis and lower limb anatomical landmark calibration precision and its propagation to bone geometry and joint angles," *Med. Biol. Eng. Comput.*, vol. 37, no. 2, pp. 155–161, Mar. 1999.

[63] F. Marin, H. Mannel, L. Claes, and L. Dürselen, "Correction of axis misalignment in the analysis of knee rotations," *Hum. Movement Sci.*, vol. 22, no. 3, pp. 285–296, Aug. 2003.

[64] R. Baker, L. Finney, and J. Orr, "A new approach to determine the hip rotation profile from clinical gait analysis data," *Hum. Movement Sci.*, vol. 18, no. 5, pp. 655–667, Oct. 1999.

[65] 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, no. 3, pp. 360–369, Apr. 2009.

[66] S. Sridar, Z. Qiao, N. Muthukrishnan, W. Zhang, and P. Polygerinos, "A soft-inflatable exosuit for knee rehabilitation: Assisting swing phase during walking," *Frontiers Robot. AI*, vol. 5, p. 44, May 2018.

[67] G. McGinty, J. J. Irrgang, and D. Pezzullo, "Biomechanical considerations for rehabilitation of the knee," *Clin. Biomech.*, vol. 15, no. 3, pp. 160–166, Mar. 2000.

[68] P. O. Riley, G. Paolini, U. D. Croce, K. W. Paylo, and D. C. Kerrigan, "A kinematic and kinetic comparison of overground and treadmill walking in healthy subjects," *Gait Posture*, vol. 26, no. 1, pp. 17–24, Jun. 2007.

[69] R. E. Fellin, K. Manal, and I. S. Davis, "Comparison of lower extremity kinematic curves during overground and treadmill running," *J. Appl. Biomech.*, vol. 26, no. 4, pp. 407–414, Nov. 2010.

Erika D'Antonio (Graduate Student Member, IEEE) received the bachelor's *(cum laude)* degree in biomedical engineering from the Sapienza University of Rome, in January 2014, the master's *(cum laude)* degree in biomedical engineering, in January 2016, and the Ph.D. (Hons.) degree in industrial and civil engineering from the Mech. Meas. and Microelectr. Laboratory (M3lab), Department of Engineering, University Niccolò Cusano, under the supervision of Prof. F. Patané, in 2021. She is currently a Postdoctoral Researcher with the Institute of Computer Engineering (ZITI), Heidelberg University, Germany, under the supervision of Prof. Lorenzo Masia. Her main research activities focus on the development of motion-analysis systems, robotic mechanisms, haptic interfaces, and exoskeletons and virtual/augmented reality setups for neurorehabilitation.

Juri Taborri received the master's *(cum laude)* degree in biomedical engineering from the Sapienza University of Rome, Italy, in 2014, and the Ph.D. degree in industrial and management engineering in 2017. He is currently an Assistant Professor in mechanical and thermal measurements with the University of Tuscia, Viterbo, Italy. His research interests include biomechanical measurements, validation and metrological characterization of innovative sensor systems, electromyography and muscle synergies, and machine-learning algorithms for human motion identification and control of robotic device for rehabilitation.

Ilaria Mileti received the master's degree in biomedical engineering from the Sapienza University of Rome, Italy, in 2015, and the Ph.D. degree in industrial and management engineering in 2019. She is currently an Assistant Professor in mechanical and thermal measurements with the Department of Engineering, University Niccolò Cusano, Rome, Italy. Her research interests include measurements for experimental biomechanics in clinical and work environments, wearable systems, machine learning algorithms, and robotic devices.

Stefano Rossi received the Mechanical Engineering (Hons.) degree from the Sapienza University of Rome in 2004, and the Ph.D. degree from Padua University in 2008. He is currently an Associate Professor and the Ph.D. Advisor with the Department of Economics, Engineering, Society, and Business Organization, Tuscia University. He is also the Coordinator of the master's degree in mechanical engineering and the Head of the Industrial Measurements and Experimental Biomechanics Laboratory. His research are focused on the design of new robotic systems for the movement recovery and the individuation of indices to objectively measure human motor performance by means of inertial systems, optoelectronic systems, and electromyography.

Fabrizio Patané received the Mechanical Engineering degree from the Sapienza University of Rome in 2000, and the Ph.D. degree from University of Padua in 2004. He is currently an Associate Professor of Thermal and Mechanical Measurements with the Faculty of Engineering, University Niccolò Cusano, Rome, Italy. He is the Coordinator of the Teaching and the Research Commissions and a member of the Doctorate School Teaching Staff in industrial-civil engineering with the University Niccolò Cusano. His scientific skills regard mechanical and thermal measurements, environmental and inertial measurements, and motion-analysis systems for biomedical applications and robotic mechanisms for neurorehabilitation.