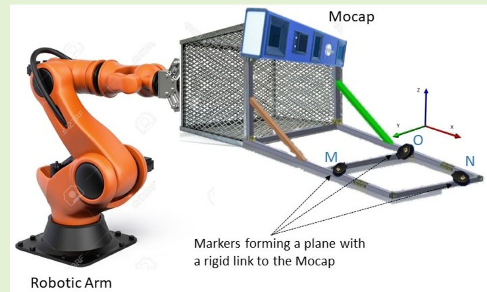
RobCap: A Mobile Motion Capture System Mounted on a Robotic Arm

Daniel Alshamaa, Andrea Cherubini, Robin Passama, Simon Pla, Loic Damm, Sofiane Ramdani

RobCap: A Mobile Motion Capture System Mounted on a Robotic Arm

Daniel Alshamaa, Andrea Cherubini, Robin Passama, Simon Pla, Loïc Damm and Sofiane Ramdani

***Abstract*— Fixed Motion Capture systems (*Mocap*) are well-established nowadays and are available in indoor and outdoor environments. Here, we introduce *RobCap*, a Moving Motion Capture System designed for acquiring accurate three-dimensional kinematics of human motion with large displacements. RobCap consists of an optoelectronic Mocap (Codamotion©) mounted on a robotic arm, to track the human motion in the local moving frame. This paper presents this device and expresses the local kinematics of the markers in a global coordinate system. We have acquired a set of 560 measurements, by controlling the Robotic arm to follow the subject's movement and to track the markers attached to his wrist. The performance of the RobCap is evaluated through comparison with state-of-the-art technologies. The 3D error is found to be 9.7 ± 4.1 mm and 3.9 ± 0.8 mm, respectively without and with calibration. RobCap increases the nominal capture volume of the fixed Mocap from 75 m^3 to 523 m^3 and the operational volume from 105 m^3 to 1436 m^3. Our findings show that the RobCap is accurate and precise enough to be used as a mobile Mocap system for large displacements.**



***Index Terms*— motion capture, robotic arm, error, rigid transformation, system calibration**

I. Introduction

Motion capture systems (Mocap) have received a huge interest in recent years. A wide range of applications in the robotics [1], health [2], [3] and sports [4]–[6] fields have motivated researchers and engineers to develop systems which accurately track human motion. Optoelectronic Mocap systems such as Vicon©, OptiTrack©, Qualisys©, Codamotion©, among others, are well-established nowadays and achieve millimetric accuracy [7], [8]. These systems differ by the type of markers, the number of cameras, the field of view and the reconstruction algorithm. One common feature is that they are all fixed, a characteristic which limits the work envelope to the field of view of the Mocap system. This technology uses passive or active markers. In the former case, it uses multiple fixed high speed cameras around the measurement area to triangulate a precise marker position. Infrared lighting allows the capture of high contrast images of the reflective markers up to 2kHz. At least two cameras at a time must capture a marker otherwise there are occlusion errors. Post-processing is required to differentiate the markers and restore the correct path [9]. In the latter case, active optical markers act as a light source instead of a reflector and are deployed as infrared emitting diodes. Here, the measurement frequency is reduced but less post-processing is required since since individual diodes can be identified.

This manuscript was submitted on June 2021. This work is part of the 'MOVCAP' project, funded by the European Regional Development Fund, region Occitanie in France.

D. Alshamaa, A. Cherubini, R. Passama and S. Ramdani are with LIRMM, Univ. Montpellier, CNRS, Montpellier, France (e-mail of corresponding author: alshamaa@opta-lp.com).

S. Pla and L. Damm are with EuroMov, Univ. Montpellier, Montpellier, France

A more flexible and less expensive solution is based on computer vision using an RGB-D camera and deep learning algorithms, e.g., OpenPose and DeepPose. As compared to optical technology, this is a more challenging and severely underconstrained problem. RGB-based pose estimation in 2D has been widely researched, but estimates only the 2D pose [10], [11]. Learning-based discriminative methods in general and deep learning methods in particular, represent the current state of the art in 2D pose estimation [12], [13]. Estimating the 3D pose using this technology is even a much harder challenge tackled by relatively fewer methods [14]–[17]. While this solution is promising and improves quickly with advancements in deep learning architectures, computation capacities and training databases, its current accuracy is still far behind that of optoelectronic technology. The average errors of such systems range from 5 to 15 centimeters, according to the complexity of the application and to the position of the joint being tracked, as compared to the millimetric accuracy of optoelectronic solutions.

Kinect is a low-cost sensor developed by Microsoft and uses an RGB camera and an infrared projector and sensor to track human joints in the three dimensional space [18]–[20]. However, the skeleton data obtained from this system exhibit a high level of jitter due to noise and estimation error. This jitter worsens when there is occlusion or a subject moves slightly out of the field of view of the sensor [21].

Other systems worth mentioning rely on inertial technology [22], [23]. Such systems use a set of inertial measurement units, containing a combination of gyroscopes, magnetometers and accelerometers to reconstruct the motion of the subject [24], [25]. For deducing joint kinematics, usually one sensor is positioned on each body segment to be captured. A calibration procedure is typically performed in order to obtain the relations between the IMUs and the segment coordinate systems. As compared to the optical technology, the placement procedure is less time-consuming and can possibly be performed by a non-expert. However, a major limitation of this technology is that positions are not measured directly. Rather, the trajectory of a person is approximated through the motion of a pre-defined biomechanical model that is driven by the IMU measurements [26]. These latter are not accurate as they suffer from noisiness and bias in the measured angular velocities, accelerations, and magnetic fields. As a result, although being a relatively cheaper technology, it is still far from millimetric accuracy achieved by the optical technology for instance.

Electromagnetic systems (EMSs) find the unknown positions of the measurement transponders by means of time-of-flight of the electromagnetic waves–radio waves–travelling from the transponder to the base stations. EMS provides large capture volumes, but are less accurate than optical systems. The advantage here is that such systems require to determine the positions of the transponders; moreover the human body is transparent for the field applied [27]. However, this technology suffers from several issues. One of the drawbacks of the system, at the level of the experimental setup, is the sensitivity for ferromagnetic material in the environment, leading to a decrease in the accuracy [28]. Moreover, such systems generally have low sampling frequencies, which are even lowered more when using multiple markers. For all the aforementioned reasons, we will focus in the rest of the paper on optoelectronic Mocap, since our approach relies on this technology, to obtain a high accuracy mobile motion capture system.

One major drawback of optoelectronic Mocap systems is that they are based on fixed cameras. This limits their work envelope to the sensor field of view. Yet, many applications (e.g. tracking the movement of runners, footballers, ski racers, among other sports, or monitoring patients in mobile environments) require a Mocap that can move with the subject to be tracked [29], [30]. As a consequence, the human motion can be studied only over a few number of cycles [31], [32]. The range of available Mocap systems is currently a couple of cubic meters. For instance, a simulation on the Vicon visualization website [33] shows that 8 infrared cameras achieve an excellent tracking quality only within 76% of a $5 \times 5 \times 3\ m^3$ volume. A larger working envelope or capture volume requires a higher number of cameras.

To solve the problem, the intuitive solution was to try to use stationary equipment such as treadmills and ergometers [34], [35]. In walking, for example, it is shown that treadmill and overground movements are similar, but can present significant differences in the extracted kinematic parameters, although reports argue that the order of these differences is relatively small [35]. Nonetheless, because of restrictions to the allowed movement excursion, such tools only partially reproduce the real motion and the treadmill walking and running mechanics cannot be generalized to the overground condition [36]. Another solution to address the problem of wider capture volumes has been proposed by Kanade et al. [37] who installed more than 50 cameras on a dome to cover larger displacements. In addition to the cost aspects, this solution introduces many technical difficulties, e.g, in defining a common volume for calibrating all cameras. Due to these limitations, participants are asked to perform only partial movements, limited to the capture volume of the Mocap system. This introduces a bias when evaluating the motion-related parameters of participants because the dynamics of partial and unconstrained movements can differ, and affect the performances.

To overcome these limits, researchers have proposed moving Mocap systems. These consist in fixing the cameras to a rigid rolling framework [38] or to a boat [39]. The Mocap can move according to the motion of the participant, to ensure a larger working envelope. A challenge for these systems is the error induced in the reconstructed 3D joint positions as the Mocap moves. Along with the noise and vibrations caused by the device carrying the Mocap, other issues (e.g., delays, variable angles of sight) can affect the reconstruction algorithm. Nonetheless, the implementation of these systems has shown encouraging results in [38] and [39]. Yet, their application was restricted to 2D motion capture in TV sports events and in the movie production industry and they were not considered for accurate 3D kinematics acquisition. The authors of [40] and [41] show that accurate 3D kinematic acquisition with a moving Mocap is possible. To this end, they install a set of Vicon cameras, capturing 3D motion, on a rolling framework. Yet, their device is complex, since it requires installing a heavy rolling framework in addition to a large number of cameras all around the participant's area to ensure proper 3D motion capture.

In this paper, we present a new device, called RobCap, featuring a compact optoelectronic Mocap system mounted on a robotic arm. The objective is to develop a high accuracy mobile Mocap for 3D kinematics acquisition in large areas. We mount Codamotion©, a Mocap made up of a compact bar with three 3 infrared cameras, on a robotic arm with six degrees of freedom. By controlling the robot arm, we can move the Mocap, so that it can follow the motion of an individual, equipped with infrared markers. As the Mocap moves, the participant's kinematics are acquired in the local frame of the cameras. Since the participant is moving with respect to the Mocap, which is itself moving with respect to the fixed global frame, it is not straightforward to estimate the global 3D kinematics of the individual. Hence, it also becomes very difficult to assess the RobCap estimation performance. To overcome this issue, we design a coordinate frame of 3 infrared markers rigidly attached to the mobile Mocap system, to form a local frame. These markers, i.e. the local frame, is visible in the global frame throughout the acquisition. This allows us to represent the motion of the local frame in the global frame and therefore transform the kinematics of the individual from the local to the global frame.

The rest of the paper is organized as follows. Section II describes the materials and methods used to develop the proposed system. Sections III and IV present and discuss the obtained results, respectively. Section V concludes the paper and lists some perspectives.

II. Materials and Methods

In this section, we present the materials and methods used to develop and evaluate RobCap. We first describe the RobCap setup, and then the experimental protocol. Finally, we detail the method used to acquire the global kinematics from the local motion, to calibrate RobCap and to evaluate its performance.

A. Experimental setup and protocol

The experimental setup is composed of 3 parts, shown in Fig. 1. The first part is dedicated to the RobCap. The second – for evaluating the RobCap performance – consists of a fixed Mocap system, with the corresponding global frame. The last part is the area where the participants' motion takes place.

The RobCap itself is composed of 3 elements: a robotic arm, a Codamotion Mocap system and a rigid structure. The robot used in our work is KUKA KR500-3 available at the EuroMov laboratory, Montpellier, France. It is a 6-axis robot arm that offers a 500 kg payload, 2825 mm reach and a repeatability of 0.08 mm. The role of the robotic arm is to displace the Mocap system according to the motion of the individual. The Codamotion is a classic Mocap device, which relies on active infrared optoelectronic technology. It is characterized by its compact size (a bar of 3 cameras of dimensions $800 \times 112 \times 80$ mm and 5 Kg weight). Its resolution is 0.3 mm along the depth axis and 0.05 mm along the two other axes, and it has a nominal capture volume (where maximum accuracy is maintained) of $5 \times 3 \times 5 = 75m^3$ and an operational working volume of 105 m^3. This Mocap device is the main element of the setup, as it must reconstruct the 3D trajectories of the markers being tracked. The third element is a rigid metallic chassis. It is rigidly attached to the Codamotion. The 2D dimensions of the plane of the chassis are 200×100 mm. At the far end of this structure lies a plate of dimensions 100×100 mm, where a set of three active infrared markers, designated O, A, and B are placed, to form a 2D Cartesian coordinate system. The objective of this structure will be made clear in the next paragraph.

To acquire the global kinematics of the individual and evaluate the performance of the RobCap, a second Codamotion CXS is fixed in the experimental area. The sampling frequency of both Mocap systems is set to 100Hz. The two systems are synchronized using a Sync cable. Data are sent to a hub then saved on a PC for offline processing. Data are processed using Matlab 2017 on an i7 Windows PC.

The movement of the participants must take place in the area visible to both the RobCap and this fixed Mocap. Three male participants (age = 30 ± 5 years, height = 1.80 ± 0.1 m, weight = 80 ± 5 Kg) are equipped with a set of 4 markers on their wrist. They are requested each at a time to move in the motion area and move their hand at their convenience, resulting in random patterns. We use a remote control, to make the robot follow the participants' motion. A total of 186 random movement patterns is acquired. These include random movement of RobCap, of the participants and their hands with the associated markers. During data acquisition, all 7 markers (4 on the participant and 3 on RobCap) are tracked by both Codamotion. Since processing markers gaps is out of the scope of this paper, we only keep the acquisitions where all the markers are detected for at least 95% of the total number of frames. This leaves only a set of 140 of the 186 original signals. We obtain the samples missing in these 140 signals, by linear interpolation. A short video for the system in action is attached to this paper as supplementary material.

B. Data processing

Suppose both the fixed and mobile Mocap are tracking a marker in the motion area. The coordinates of this marker as viewed by the RobCap, i.e. in the local frame, are denoted as

$$\boldsymbol{p}_M = [x_M, y_M, z_M]^\top .$$

The same marker has coordinates

$$\boldsymbol{p}_F = [x_F, y_F, z_F]^\top$$

in the global fixed frame. We also suppose that an acquisition $j \in \{1, \dots, N_t\}$ (N_t being the total number of acquisitions) is composed of N_j points. This means that the n marker's coordinates ($n \in \{1, \dots, N_j\}$) at acquisition j are

$$\boldsymbol{p}_M(n) = [x_M(n), y_M(n), z_M(n)]^\top$$

and

$$\boldsymbol{p}_F(n) = [x_F(n), y_F(n), z_F(n)]^\top ,$$

in the local and global frames respectively. In this work, a total of $N_t = 560$ measurements is acquired, so $j \in \{1, \dots, 560\}$.

Since RobCap moves to follow the motion of the participant, its local frame is not known anymore as it moves also with it. This prevents us from determining the global kinematics to interpret the human motion, and from evaluating the RobCap's performance in terms of 3D reconstruction accuracy. This is why we attach the rigid structure with the three markers O, A and B to RobCap. At initialization, we use these markers to define both the local frame of the RobCap and the global frame of the fixed Mocap. As the RobCap moves, these markers, visible by the fixed Mocap, allow us to know the exact position of the local frame in the global frame. The coordinates of the markers in the global frame are denoted as $\boldsymbol{O}_F(n)$, $\boldsymbol{A}_F(n)$ and $\boldsymbol{B}_F(n)$, respectively. Throughout the acquisition, the coordinates of these markers in the local frame are constant and given by:

$$\begin{cases} \boldsymbol{O}_M(n) = [0,0,0]^\top \\ \boldsymbol{A}_M(n) = [1,0,0]^\top \quad \quad \forall n \in \{1, \dots, N_j\} \ . \\ \boldsymbol{B}_M(n) = [0,1,0]^\top \end{cases}$$

We are interested in transforming the coordinates $\boldsymbol{p}_M(n)$ of the marker in the motion area from the local coordinate frame to the global coordinate frame, so that they can be compared with $\boldsymbol{p}_F(n)$. Suppose the transformed coordinates of $\boldsymbol{p}_M(n)$ are denoted $\boldsymbol{p}_{tr}(n) = [x_{tr}(n), y_{tr}(n), z_{tr}(n)]^\top$. The relation between $\boldsymbol{p}_{tr}(n)$ and $\boldsymbol{p}_M(n)$ is governed by the following equation [42]:

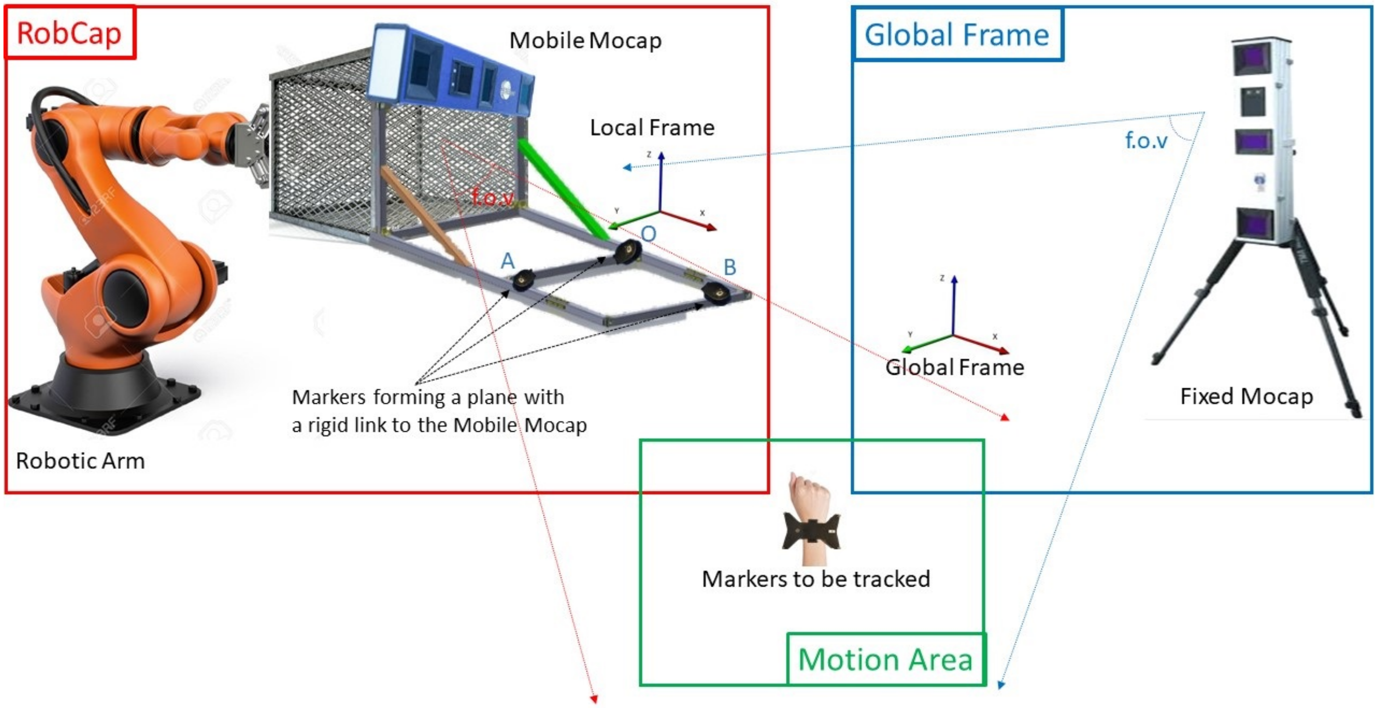


Fig. 1: An illustration of the experimental setup. RobCap is a mobile Mocap mounted on a robotic arm. Three markers are attached to the structure of the RobCap to form a plane, rigidly linked to the Mocap. All markers are detected by RobCap and by the fixed Mocap.

$$\boldsymbol{p}_{tr}(n) = \ {}^{F}\boldsymbol{T}_{M}(n) \ \cdot \ \boldsymbol{p}_{M}(n), \tag{1}$$

where ${}^{F}\boldsymbol{T}_{M}(n)$ is the transformation matrix between the local mobile frame M and the global fixed frame F for each point $n \in \{1, \ldots, N_j\}$ for any acquisition $j \in \{1, \ldots, N_t\}$.

Suppose that the bases of frames M and F at any point n are $\overrightarrow{\boldsymbol{m}}(n)$ and $\overrightarrow{\boldsymbol{f}}(n)$, respectively. The transformation matrix ${}^{F}\boldsymbol{T}_{M}(n)$ is a change of basis from $\overrightarrow{\boldsymbol{m}}(n)$ to $\overrightarrow{\boldsymbol{f}}(n)$, as shown in the following equation,

$$\overrightarrow{\boldsymbol{f}}\boldsymbol{(n)} = \ {}^{F}\boldsymbol{T}_{M}(n) \cdot \overrightarrow{\boldsymbol{m}}(n), \tag{2}$$

where the coordinates of $\overrightarrow{\boldsymbol{f}}(n)$ are expressed in $\overrightarrow{\boldsymbol{m}}(n)$. The transformation matrix ${}^{F}\boldsymbol{T}_{M}(n)$ can be thus deduced,

$${}^{F}\boldsymbol{T}_{M}(n) = \overrightarrow{\boldsymbol{f}}(n) \cdot \overrightarrow{\boldsymbol{m}}^{-1}(n). \tag{3}$$

To determine the components of the two bases $\overrightarrow{\boldsymbol{m}}(n)$ and $\overrightarrow{\boldsymbol{f}}(n)$, we use the coordinates of the three markers O, A and B in both bases at each point n. As we can notice from the coordinates of these markers in the local frame M, the basis $\overrightarrow{\boldsymbol{m}}(n)$ is the identity matrix and hence

$$\overrightarrow{\boldsymbol{m}}^{-1}(n) = \overrightarrow{\boldsymbol{m}}(n) = \boldsymbol{I}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{4}$$

To determine the basis $\overrightarrow{\boldsymbol{f}}(n)$, we have to express the coordinates of the fixed frame in the mobile frame. At first, we create a normalized vector $\boldsymbol{V}_F(n) = \left[\frac{\boldsymbol{u}(n)}{|\boldsymbol{u}(n)|}, \frac{\boldsymbol{v}(n)}{|\boldsymbol{v}(n)|}, \frac{\boldsymbol{w}(n)}{|\boldsymbol{w}(n)|}\right]^T$, such that $\boldsymbol{u}(n) = \overrightarrow{\boldsymbol{O}_F\boldsymbol{A}_F}(n), \boldsymbol{v}(n) = \overrightarrow{\boldsymbol{O}_F\boldsymbol{B}_F}(n)$ and $\boldsymbol{w}(n) = \boldsymbol{u}(n) \times \boldsymbol{v}(n)$. As a reminder, the coordinates $\boldsymbol{O}_F$, $\boldsymbol{A}_F$ and $\boldsymbol{B}_F$ of the three markers O, A and B respectively can be easily acquired by the fixed Mocap. The vectors $\boldsymbol{u}(n)$ and $\boldsymbol{v}(n)$ form the 2D basis of the coordinate system in the plane of (AOB) and $\boldsymbol{w}(n)$ is the cross product of vectors $\boldsymbol{u}(n)$ and $\boldsymbol{v}(n)$ forming the third axis of the coordinate system. Once the vector $\boldsymbol{V}_F(n)$ is defined, the basis $\overrightarrow{\boldsymbol{f}}(n)$ can be easily determined as follows,

$$\overrightarrow{\boldsymbol{f}}(n) = \begin{bmatrix} \boldsymbol{V}_F^T(n) & \boldsymbol{O}_F(n) \\ \boldsymbol{0}^T & 1 \end{bmatrix}. \tag{5}$$

Replacing (4) and (5) in (3) and then injecting the latter in equation (1), allows us to transform the coordinates of any marker in the motion area at any point of the acquisition from the local frame to the global frame at any point $\boldsymbol{P}_M$,

$$\boldsymbol{P}_{tr}(n) = \begin{bmatrix} \boldsymbol{V}_F^T(n) & \boldsymbol{O}_F(n) \\ \boldsymbol{0}^T & 1 \end{bmatrix} \cdot \boldsymbol{P}_M(n). \tag{6}$$

Algorithm 1 gives the pseudocode for transforming the 3D motion acquired by the RobCap in its local frame, to the global frame.

C. System calibration

In the previous paragraph, we have transformed the coordinates of any marker in the motion area from the local mobile frame of the RobCap to the global fixed frame to obtain $\boldsymbol{p}_{tr}(n), n \in \{1, \ldots, N_j\}, j \in \{1, \ldots, N_t\}$. However,

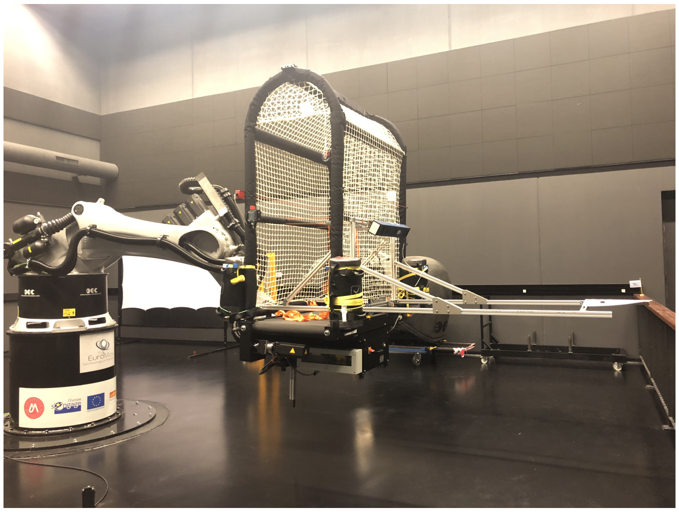


Fig. 2: A photo of the actual setup.

Algorithm 1: Acquiring global 3D kinematics from the local motion of the RobCap device.

Input : $\boldsymbol{P}_M(n), \boldsymbol{O}_F(n), \boldsymbol{A}_F(n), \boldsymbol{B}_F(n)$
Output: $\boldsymbol{P}_{Tr}(n)$

for $j \in \{1, \ldots, N_t\}$ **do**
  for $n \in \{1, \ldots, N_j\}$ **do**
    $\boldsymbol{u}(n) = \overrightarrow{\boldsymbol{O}_F\boldsymbol{A}_F}(n)$;
    $\boldsymbol{v}(n) = \overrightarrow{\boldsymbol{O}_F\boldsymbol{B}_F}(n)$;
    $\boldsymbol{w}(n) = \boldsymbol{u}(n) \times \boldsymbol{v}(n)$;
    $\boldsymbol{V}_F(n) = \left[\frac{\boldsymbol{u}(n)}{|\boldsymbol{u}(n)|}, \frac{\boldsymbol{v}(n)}{|\boldsymbol{v}(n)|}, \frac{\boldsymbol{w}(n)}{|\boldsymbol{w}(n)|}\right]^T$;
    $\boldsymbol{P}_{tr}(n) = \begin{bmatrix} \boldsymbol{V}_F^T(n) & \boldsymbol{O}_F(n) \\ \boldsymbol{0}^T & 1 \end{bmatrix} \cdot \boldsymbol{P}_M(n)$;
  end
end

a source of error that has the potential to negatively affect the expression of the 3D kinematics of the motion acquired by the RobCap in the global frame is the initial definition of both frames. When a set of markers is used to define the origin and the 2D axes of motion, both Mocap systems compute internal transformations so that the current coordinates of these markers form a Cartesian coordinate system. In addition, the error of the Mocap systems varies as a function of the angle of view of the cameras with respect to the position of the marker. Moreover, the number of cameras that detect each marker affects the reconstruction accuracy. Begon and Lacouture [43] have shown that changes in couples of cameras to reconstruct the 3D marker positions causes systematic error in position estimation. All these sources of error, although in the order of millimeters each, add up and may yield a relevant difference between the transformed coordinates and those measured directly in the global frame. To account for these errors, we propose to go through a calibration phase. Before going into the details of the calibration, it is important to note the following. At first, the objective of this phase is to suppress the sources of errors indicated above. However, it requires the presence of a fixed Mocap system as it will be explained in the next paragraph. This means that the absence of Mocap system other than the one used in RobCap hinders the application of this phase. In all cases, this phase is not mandatory for the functioning of the system.

We recall that $\boldsymbol{p}_F(n)$ and $\boldsymbol{p}_{tr}(n), n \in \{1, \ldots, N_j\}, j \in \{1, \ldots, N_t\}$ designate the coordinates of the marker as measured in the global fixed frame and those transformed from the mobile local frame of the RobCap into the global frame respectively. We suppose that all sources of error resulting in a difference between these coordinates can be accounted through a calibration matrix $\boldsymbol{C}$, such that,

$$\boldsymbol{p}_F(\cdot) = \boldsymbol{C} \cdot \boldsymbol{p}_{tr}(\cdot), \tag{7}$$

where the calibration matrix $\boldsymbol{C}$ is defined as,

$$\boldsymbol{C} = \begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} \\ c_{21} & c_{22} & c_{23} & c_{24} \\ c_{31} & c_{32} & c_{33} & c_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{8}$$

To calibrate our system, we need to determine the 12 unknown elements of $\boldsymbol{C}$. To this end, we do or use a single acquisition of N_C points. Since the calibration matrix has to be same for all the points of the acquisition, (7) can be written:

$$\begin{bmatrix} x_F(n) \\ y_F(n) \\ z_F(n) \\ 1 \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} \\ c_{21} & c_{22} & c_{23} & c_{24} \\ c_{31} & c_{32} & c_{33} & c_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_{tr}(n) \\ y_{tr}(n) \\ z_{tr}(n) \\ 1 \end{bmatrix}, \tag{9}$$

$\forall n \in \{1, \ldots, N_C\}$. The elements of the calibration matrix $\boldsymbol{C}$ are computed so as to minimize the mean square error between the observed coordinates $\boldsymbol{p}_F(n) = [x_F(n), y_F(n), z_F(n)]^\top$ and the transformed coordinates $\boldsymbol{p}_{tr}(n) = [x_{tr}(n), y_{tr}(n), z_{tr}(n)]^\top$ over all the points of the acquisition $n \in \{1, \ldots, N_C\}$. To determine these 12 unknown elements $c_{ij}, i \in \{1, 2, 3\}, j \in \{1, 2, 3, 4\}$, we transform equation (9) into another solvable form shown in equation (11) using all N_c points. Let us denote $\boldsymbol{\mathcal{K}}$ the $(4N_C \times 16)$ matrix, $\boldsymbol{C}_R$ the newly formatted $(16 \times 1)\,\boldsymbol{C}$ matrix and $\boldsymbol{\mathcal{L}}$ the $(4N_C \times 1)$ matrix in equation (11). As we can notice, equation (11) is in the form of $\boldsymbol{\mathcal{K}}\boldsymbol{C}_R = \boldsymbol{\mathcal{L}}$. The least square solution of this equation is $\hat{\boldsymbol{C}}_R = \left(\boldsymbol{\mathcal{K}}^T\boldsymbol{\mathcal{K}}\right)^{-1}\boldsymbol{\mathcal{K}}^T\boldsymbol{\mathcal{L}}$, which allows to determine all the elements $c_{ij}, i \in \{1, 2, 3\}, j \in \{1, 2, 3, 4\}$. Having computed the calibration matrix $\boldsymbol{C}$ using the N_C points, we use it to determine the final transformed and calibrated coordinates $\boldsymbol{p}_C(n) = [x_C, y_C, z_C]^T$ of any marker in the motion area from the local mobile frame of the RobCap to the global frame,

$$\boldsymbol{p}_C(n) = \boldsymbol{C} \cdot \boldsymbol{p}_{tr}(n), \tag{10}$$

$n \in \{1, \ldots, N_j\}, j \in \{1, \ldots, N_t\}$. Algorithm 2 describes the pseudocode for calibrating the proposed system.

$$\underbrace{\begin{bmatrix} x_{tr}(1) & y_{tr}(1) & z_{tr}(1) & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & x_{tr}(1) & y_{tr}(1) & z_{tr}(1) & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & x_{tr}(1) & y_{tr}(1) & z_{tr}(1) & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & x_{tr}(1) & y_{tr}(1) & z_{tr}(1) & 1 \\ x_{tr}(2) & y_{tr}(2) & z_{tr}(2) & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & x_{tr}(2) & y_{tr}(2) & z_{tr}(2) & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & x_{tr}(2) & y_{tr}(2) & z_{tr}(2) & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & x_{tr}(2) & y_{tr}(2) & z_{tr}(2) & 1 \\ \vdots & & & & & & & & & & & & & & & \vdots \\ x_{tr}(N_C) & y_{tr}(N_C) & z_{tr}(N_C) & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & x_{tr}(N_C) & y_{tr}(N_C) & z_{tr}(N_C) & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & x_{tr}(N_C) & y_{tr}(N_C) & z_{tr}(N_C) & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & x_{tr}(N_C) & y_{tr}(N_C) & z_{tr}(N_C) & 1 \end{bmatrix}}_{(4N_C \times 16)} \cdot \underbrace{\begin{bmatrix} c_{11} \\ c_{12} \\ c_{13} \\ c_{14} \\ c_{21} \\ c_{22} \\ c_{23} \\ c_{24} \\ c_{31} \\ c_{32} \\ c_{33} \\ c_{34} \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}}_{(16 \times 1)} = \underbrace{\begin{bmatrix} x_F(1) \\ y_F(1) \\ z_F(1) \\ 1 \\ x_F(2) \\ y_F(2) \\ z_F(2) \\ 1 \\ \vdots \\ x_F(N_C) \\ y_F(N_C) \\ z_F(N_C) \\ 1 \end{bmatrix}}_{(4N_C \times 1)} \tag{11}$$

Algorithm 2: Calibration of the RobCap device.

Input : $\boldsymbol{P}_{tr}(n), N_C$
Output: $\boldsymbol{P}_C(n)$

$j = 3$;
for $n \in \{1, \ldots, N_C\}$ **do**
  for $i \in \{0, \ldots, 3\}$ **do**
    $\boldsymbol{\mathcal{K}}[n+i, :] = \left[\underbrace{0 \ldots 0}_{4i \text{ times}} \quad \boldsymbol{P}_{tr}^T(n) \quad 1 \quad \underbrace{0 \ldots 0}_{12-4i \text{ times}} \right]$;
  end
  while $j < 4N_C$ **do**
    $\boldsymbol{\mathcal{L}}[:, n : n+j] = [\boldsymbol{P}_F(n) \quad 1]^T$;
    $j = j + 3$;
  end
end
$\hat{\boldsymbol{C}}_R = \left(\boldsymbol{\mathcal{K}}^T \boldsymbol{\mathcal{K}}\right)^{-1} \boldsymbol{\mathcal{K}}^T \boldsymbol{\mathcal{L}}$ %deduce elements c_{ij} thus $\boldsymbol{C}$;
for $j \in \{1, \ldots, N_t\}$ **do**
  for $n \in \{1, \ldots, N_j\}$ **do**
    $\boldsymbol{P}_C(n) = \boldsymbol{C} \cdot \boldsymbol{P}_{tr}(n)$;
  end
end

III. Results

As explained in Section II, we evaluate the proposed system by tracking 140 times the positions of 4 markers, attached to the participants' wrists, resulting in a total of $N_t = 560$ measurements. At first, we transform the coordinates of the markers acquired by RobCap to the global frame, via Algorithm 1. Then, the system is calibrated using one acquisition of N_C points following Algorithm 2.

To assess RobCap's performance, we calculate its 3D error as the Euclidean distance between the transformed coordinates and the ones measured by the fixed Codamotion (ground truth). Prior to calibration, we calculate this error at each point n as

$$\sqrt{(x_F(n) - x_{tr}(n))^2 + (y_F(n) - y_{tr}(n))^2 + (z_F(n) - z_{tr}(n))^2}.$$

After calibration, we calculate the error as

$$\sqrt{(x_F(n) - x_C(n))^2 + (y_F(n) - y_C(n))^2 + (z_F(n) - z_C(n))^2}.$$

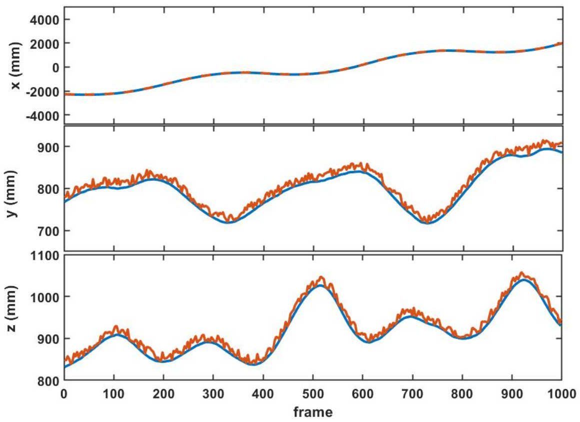
Figure 3 compares the trajectory of one of the markers attached to the participant's wrist acquired by the RobCap and transformed into the global frame, to that measured by the fixed Codamotion. Figures 3(a) and 3(c) show the x, y and z coordinates along with the 3D motion without system calibration, while Figures 3(b) and 3(d) show the same coordinates and motion after system calibration. The average, standard deviation and maximum error of x, y, z and the 3D coordinates of the system with and without system calibration over all points of all 560 acquisitions are shown in Table I.

Here, one single acquisition of $N_C = 100$ points is considered to calibrate the system. To study the influence of the length of the calibration acquisition, we vary N_C and evaluate the overall performance of RobCap transformed into the global frame. Figure 4 plots the errorbar of the system upon varying N_C from 0 to 250 with a step of 50. Finally, the proposed system is compared to state-of-the-art technologies. Table II compares the different approaches according to several criteria, such as accuracy, volume of capture, complexity and cost.

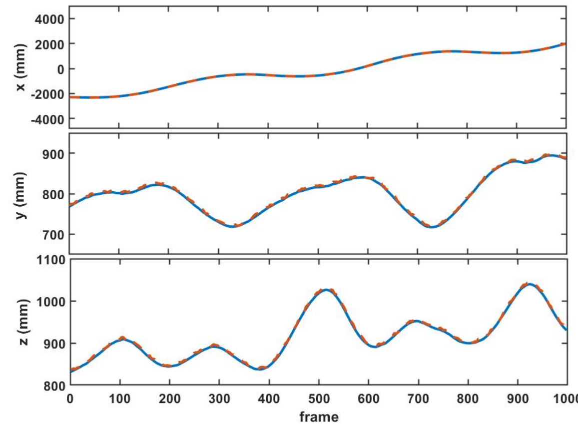
IV. Discussion

As Table I shows, the maximum error in reconstructing the 3D motion of RobCap transformed into the global frame without calibration is $22.8mm$ with an average of $9.7mm$ over all 560 measurements. The maximum error of reconstructing motion in the x- and y-directions does not exceed $10.8mm$ with an average not exceeding $5.4mm$ but could reach $16.1mm$ with an average of $9.7mm$ in the z-direction. The higher error along z is probably due to the fact that motion along this axis is computed when transforming it into the global frame through the vectorial product and not directly measured. Any errors in both x or y directions will add contribute to the z-direction error. On the other hand, the relatively small variation of error with a standard deviation of $4.1mm$ also shows that the system is quite precise. These results show that the RobCap has a high accuracy and precision in reconstructing the motion even in a situation when the whole environment is mobile, where the RobCap, the participant and the markers to be tracked are all moving.

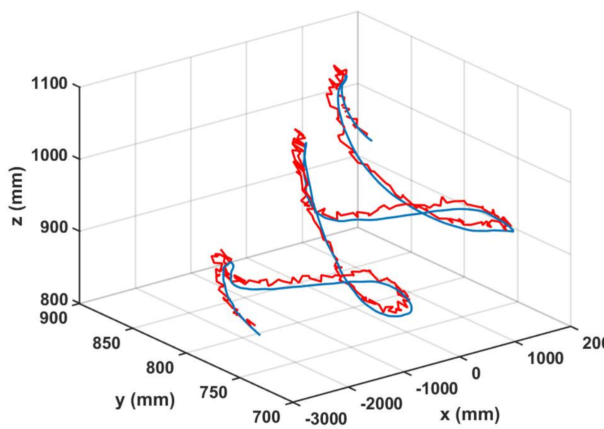
System calibration using one acquisition of $N_C = 100$ points reduces the average 3D error to $3.9mm$ with a maximum of $7.2mm$. This number of points is determined by varying N_C and noting the error. Fig. 4 shows that the error of the system reaches its minimum with $N_C = 100$ points. Longer calibration acquisitions do not result



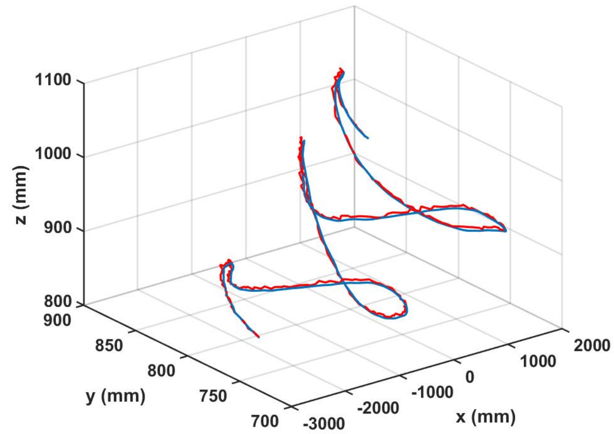
(a) Coordinates x, y and z without system calibration



(b) Coordinates x, y and z with system calibration



(c) 3D coordinates without system calibration



(d) 3D coordinates with system calibration

Fig. 3: Comparison between 3D motion coordinates of one marker acquired directly by a fixed Mocap system (blue) and those acquired by the RobCap and transformed to the global frame (red) without (left) and with (right) system calibration.

in better reconstruction accuracy. The achieved accuracy demonstrates the efficiency of the proposed calibration technique and equally the high accuracy of the RobCap after proper calibration. The calibration also has a significant positive impact on the precision of the system, enhancing the standard deviation of the error to around $0.8mm$. Nonetheless, one downside of the proposed calibration is the need for a fixed Mocap system in the experimental area, which is not always the case. Another downside is that the system has to be calibrated at each new definition of the initial and global frames. This means that each time new frames are defined, the previous calibration is not valid anymore and a new acquisition has to be considered. It is noteworthy recalling here that the calibration of the RobCap with the global frame serves only to express the motion acquired by the RobCap in the global frame. This might not be needed all the time, especially when the relative motion is sufficient.

In Table II, we compare the proposed system to other existing technologies and approaches. The criteria chosen here in our comparison are the error, the volume of capture, the complexity comprising the number of cameras and the difficulty to mount the system and the overall cost. Although other criteria could have also been chosen, we suppose that these four are the most relevant to evaluate a mobile Mocap system, especially that exact information regarding other criteria might not be available for all systems. We compare the proposed system RobCap to two fixed and three mobile Mocap systems. The first fixed Mocap system is of the same type as the one used to build our RobCap device, 'Codamotion'. We use this system as a reference to show the utility of making our proposed system mobile. The second fixed Mocap system is proposed by Kanade et al. [37]. Here, the authors build a system of 50 cameras on a hemispherical dome of radius $5m$ to enlarge the capture volume. We consider that the error of these fixed Mocap systems is equal to 0, as they form the global coordinate system and thus the reference that other systems compare with. The first mobile Mocap system is an RGB Camera-based solution using deep learning models for pose estimation [44]. Although this technology can be used for fixed and mobile Mocap systems, we consider here the broader range of applications for a fair comparison with our proposed system. The second mobile Mocap system is proposed by Kersting *et al.* [39]. It is composed of three Basler cameras sampled at 30Hz mounted on a large catamaran-type motor boat with cameras set at various heights and the most distant cameras about $14m$ apart. The third system is proposed by Begon *et al.* [40] where a rolling motion analysis system is built using eight cameras of type 'Vicon' placed at a $2.35m$ height and sampled at 100Hz. The system expresses the local kinematics of the participant walking on a pathway in a global system of coordinates. Table II provides approximate ranges of the performance of the system with respect to the four chosen criteria. This is because the performance of each system varies according to the experimental conditions. We thus consider an average value representing the performance of each system with respect to each

TABLE I: The 3D reconstruction accuracy of the RobCap after transformation to the global frame as compared to a fixed Mocap system, with and without system calibration.

Error (mm)	without calibration				with calibration $N_C = 100$			
	x	y	z	3D	x	y	z	3D
mean	5.2	5.4	8.8	9.7	1.8	1.7	2.9	3.9
std	2.6	2.8	3.1	4.1	0.3	0.4	0.7	0.8
max	10.6	10.8	16.1	22.8	4.2	3.3	4.9	7.2

TABLE II: Comparison between the proposed system and state-of-the-art technologies in terms of accuracy, volume of capture, complexity and cost.

System \ Criterion	error (mm)	volume (m^3)	complexity	cost
Fixed Mocap [45]	0	≈ 75	low	average
Kanade *et al.* [37]	0	≈ 60	high	high
Pose estimation [44]	$100 - 150$	> 500	low	low
Kersting *et al.* [39]	≈ 30	≈ 35	average	low
Begon *et al.* [40]	$20 - 24$	≈ 90	average	average
RobCap	$4 - 12$	> 500	average	high

criterion.

As the table shows, the proposed system outperforms the other mobile Mocap systems in terms of the overall accuracy, as shown by the lowest error. As mentioned earlier in this section, the fixed Mocap systems are the reference that we compare the performance of the mobile systems with and for that reason their error is said to be 0. The advantage of the mobile Mocap systems lies in enlarging the capture volume. As we can see, the proposed system along with the pose estimation method achieve the largest capture volume. The values shown here correspond to the nominal capture volume of all systems. In practice, operational working volumes are much higher than that. As compared to the fixed Mocap system of the same type, the device increases the nominal capture volume from $75m^3$ to $523m^3$ and the operational volume from $105m^3$ to $1436m^3$. We would argue that mounting the proposed system is relatively of low complexity, especially that the Mocap system is compact and ready to use and can be easily attached to a robotic arm. However, building the whole system in order to acquire the global kinematics, adds a non-negligible complexity. Another major challenge of such system is occlusion. In order to obtain measurements with a high detection ($> 95\%$ of the total frames), we had to realize the experiments in a very controlled environment in a way that all markers stay visible and detected by the Mocap systems. In practical environments, when the whole body of the person is equipped with markers, it would be difficult to guarantee such high visibility and detection rate. In such cases, methods to deal with occlusion and thus missing data have to be proposed. Another challenge that face the proposed system is the automatic control of the robotic arm. Here, it was displaced via a remote control, but a tracking system to follow the motion of the person and control the robotic arm is highly recommended. This might influence the quality of acquired signals and thus the high 2D reconstruction accuracy. Moreover, the proposed system comes at a relatively higher cost than the existing systems, with the cost of the robotic arm, the Codamotion Mocap system and miscellaneous costs of implementation. However, the major advantages of the proposed system as compared to the cheaper and less complex ones, is that it has at the same time a higher error and a larger capture volume, yet at the expense of higher cost and complexity.

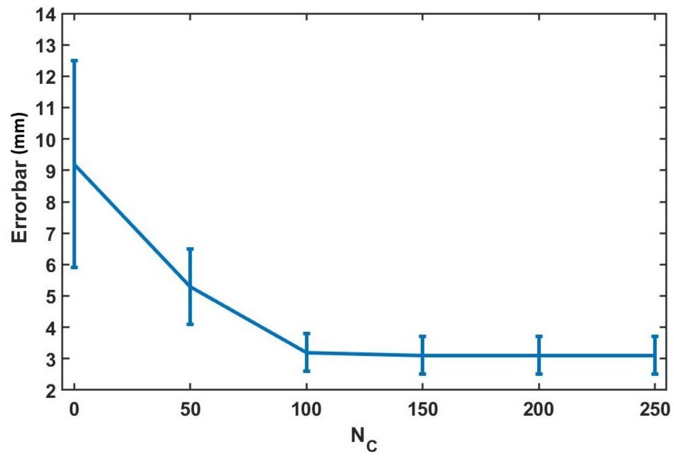


Fig. 4: Influence of the length of calibration acquisition N_C on the overall performance of the RobCap transformed into the global frame.

V. Conclusion and Perspectives

In this paper, we presented a new mobile Mocap system, called RobCap. The device is composed of a robotic arm and a compact Mocap system of type Codamotion. As compared to its fixed counterpart, the device increases the nominal capture volume from $75m^3$ to $523m^3$ and the operational volume from $105m^3$ to $1436m^3$. At first, the motion acquired by the RobCap was transformed to the global frames. After that, a calibration approach was proposed to account for the sources of error that affect the quality of the reconstruction when defining the local and the global frame. The system was evaluated on 560 measurements and the performance of the system was then compared to a fixed Mocap system of the same type, Codamotion. We noted an error of $9.7 \pm 4.1mm$ without calibration and $3.9 \pm 0.8mm$ with calibration using a single acquisition. The influence of the length of the calibration acquisition was studied and results showed that best performance was recorded for 100 points. System calibration using acquisitions beyond this number of points showed no additional improvement in the overall performance. Although of relatively higher cost and a bit more complex, the proposed system outperforms existing technologies in terms of accuracy and volume of capture. As a summary, the obtained results show that RobCap can act as a mobile Mocap device for large displacements.

Future work will investigate the usage of the real time processing function of the markers coordinates of Codamotion. This allows for real time motion tracking of the participant and thus automating the control of the robotic arm. Moreover, we aim to propose strategies to handle occlusion. As opposed to other optoelectronic Mocap systems where cameras generally surround the participant's motion, the compact Mocap system used here allows for only one side of view of motion, which implies having more markers occluded. At the experimental level, we aim to use and evaluate the proposed system in real applications with complex activities and 3D full-body reconstruction.

Acknowledgments

The authors thank Hugo Rodriguez for his help in the implementation of the experimental setup and the participants for their willingness to realize the experiments.

Conflicts of Interest

The authors declare no conflicts of interest.

Appendix

A video for the system in action is attached to the paper as supplementary material.

References

[1] O. Mazhar, S. Ramdani, B. Navarro, R. Passama, and A. Cherubini, "Towards real-time physical human-robot interaction using skeleton information and hand gestures," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1–6, IEEE, 2018.

[2] Y. Zheng, P. Zhong, K. Liu, K. Yang, and Q. Yue, "Human motion capture system based 3D reconstruction on rehabilitation assistance stability of lower limb exoskeleton robot climbing upstairs posture," *IEEE Sensors Journal*, vol. 20, no. 20, pp. 11778–11786, 2020.

[3] E. D'Antonio, J. Taborri, I. Mileti, S. Rossi, and F. Patané, "Validation of a 3D markerless system for gait analysis based on OpenPose and two RGB webcams," *IEEE Sensors Journal*, vol. 21, no. 15, pp. 17064–17075, 2021.

[4] S. Wang, Y. Xu, Y. Zheng, M. Zhu, H. Yao, and Z. Xiao, "Tracking a golf ball with high-speed stereo vision system," *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 8, pp. 2742–2754, 2019.

[5] M. A. Hidayat Yani, S. Bayu Aji, I. F. Ariyanti, S. Sukaridhoto, M. A. Zainuddin, and A. Basuki, "Implementation of motion capture system for swimmer athlete monitoring," in *2019 International Electronics Symposium (IES)*, pp. 400–405, 2019.

[6] M. P. Wilk, M. Walsh, and B. O'Flynn, "Multimodal sensor fusion for low-power wearable human motion tracking systems in sports applications," *IEEE Sensors Journal*, vol. 21, no. 4, pp. 5195–5212, 2021.

[7] E. van der Kruk and M. M. Reijne, "Accuracy of human motion capture systems for sport applications; state-of-the-art review," *European journal of sport science*, vol. 18, no. 6, pp. 806–819, 2018.

[8] M. Topley and J. G. Richards, "A comparison of currently available optoelectronic motion capture systems," *Journal of Biomechanics*, vol. 106, p. 109820, 2020.

[9] D. Alshamaa, R. Soubra, and A. Chkeir, "A radar sensor for automatic gait speed analysis in walking tests," *IEEE Sensors Journal*, vol. 21, no. 12, pp. 13886–13894, 2021.

[10] X. Chen and A. L. Yuille, "Articulated pose estimation by a graphical model with image dependent pairwise relations," in *NIPS*, 2014.

[11] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "Openpose: realtime multi-person 2d pose estimation using part affinity fields," *IEEE transactions on pattern analysis and machine intelligence*, vol. 43, no. 1, pp. 172–186, 2019.

[12] A. Newell, K. Yang, and J. Deng, "Stacked hourglass networks for human pose estimation," in *Computer Vision – ECCV 2016* (B. Leibe, J. Matas, N. Sebe, and M. Welling, eds.), (Cham), pp. 483–499, Springer International Publishing, 2016.

[13] A. Toshev and C. Szegedy, "Deeppose: Human pose estimation via deep neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1653–1660, 2014.

[14] H. Yu, Q. Fu, Z. Yang, L. Tan, W. Sun, and M. Sun, "Robust robot pose estimation for challenging scenes with an RGB-D camera," *IEEE Sensors Journal*, vol. 19, no. 6, pp. 2217–2229, 2019.

[15] X. Zhou, M. Zhu, S. Leonardos, K. G. Derpanis, and K. Daniilidis, "Sparseness meets deepness: 3d human pose estimation from monocular video," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4966–4975, 2016.

[16] D. Mehta, O. Sotnychenko, F. Mueller, W. Xu, S. Sridhar, G. Pons-Moll, and C. Theobalt, "Single-shot multi-person 3d pose estimation from monocular rgb," in *2018 International Conference on 3D Vision (3DV)*, pp. 120–130, IEEE, 2018.

[17] D. Mehta, S. Sridhar, O. Sotnychenko, H. Rhodin, M. Shafiei, H.-P. Seidel, W. Xu, D. Casas, and C. Theobalt, "Vnect: Real-time 3d human pose estimation with a single rgb camera," *ACM Transactions on Graphics (TOG)*, vol. 36, no. 4, pp. 1–14, 2017.

[18] M. Li, F. Wei, Y. Li, S. Zhang, and G. Xu, "Three-dimensional pose estimation of infants lying supine using data from a kinect sensor with low training cost," *IEEE Sensors Journal*, vol. 21, no. 5, pp. 6904–6913, 2021.

[19] J. Niu, X. Wang, D. Wang, and L. Ran, "A novel method of human joint prediction in an occlusion scene by using low-cost motion capture technique," *Sensors*, vol. 20, no. 4, p. 1119, 2020.

[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, 2019.

[21] J. Shu, F. Hamano, and J. Angus, "Application of extended Kalman filter for improving the accuracy and smoothness of Kinect skeleton-joint estimates," *Journal of Engineering Mathematics*, vol. 88, no. 1, pp. 161–175, 2014.

[22] Neuron. https://neuronmocap.com/.

[23] Xsens. https://xsens.com/.

[24] T. Lisini Baldi, F. Farina, A. Garulli, A. Giannitrapani, and D. Prattichizzo, "Upper body pose estimation using wearable inertial sensors and multiplicative kalman filter," *IEEE Sensors Journal*, vol. 20, no. 1, pp. 492–500, 2020.

[25] S. Zihajehzadeh, P. K. Yoon, B. Kang, and E. J. Park, "UWB-Aided Inertial Motion Capture for lower body 3-D dynamic activity and trajectory tracking," *IEEE Transactions on Instrumentation and Measurement*, vol. 64, no. 12, pp. 3577–3587, 2015.

[26] A. R. Anwary, H. Yu, A. Callaway, and M. Vassallo, "Validity and consistency of concurrent extraction of gait features using inertial measurement units and motion capture system," *IEEE Sensors Journal*, vol. 21, no. 2, pp. 1625–1634, 2021.

[27] H. M. Schepers and P. H. Veltink, "Stochastic magnetic measurement model for relative position and orientation estimation," *Measurement science and technology*, vol. 21, no. 6, p. 065801, 2010.

[28] N. Schuler, M. Bey, J. Shearn, and D. Butler, "Evaluation of an electromagnetic position tracking device for measuring in vivo, dynamic joint kinematics," *Journal of biomechanics*, vol. 38, no. 10, pp. 2113–2117, 2005.

[29] D. Alshamaa, F. Mourad-Chehade, and P. Honeine, "Decentralized kernel-based localization in wireless sensor networks using belief functions," *IEEE Sensors Journal*, vol. 19, no. 11, pp. 4149–4159, 2019.

[30] D. Alshamaa, F. Mourad-Chehade, P. Honeine, and A. Chkeir, "Fusion of multiple mobility and observation models for indoor zoning-based sensor tracking," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 56, no. 6, pp. 4315–4326, 2020.

[31] B. Pietraszewski, S. Winiarski, and S. Jaroszczuk, "Three-dimensional human gait pattern–reference data for normal men," *Acta of Bioengineering and Biomechanics*, vol. 14, no. 3, pp. 9–16, 2012.

[32] E. Mirek, J. L. Kubica, J. Szymura, S. Pasiut, M. Rudzińska, and W. Chwała, "Assessment of gait therapy effectiveness in patients with Parkinson's disease on the basis of three-dimensional movement analysis," *Frontiers in neurology*, vol. 7, p. 102, 2016.

[33] Vicon. https://www.vicon.com/visualization/.

[34] R. R. Bini and F. Diefenthaeler, "Kinetics and kinematics analysis of incremental cycling to exhaustion," *Sports Biomechanics*, vol. 9, no. 4, pp. 223–235, 2010.

[35] P. O. Riley, G. Paolini, U. Della 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, 2007.

[36] J. Sinclair, J. Richards, P. J. Taylor, C. J. Edmundson, D. Brooks, and S. J. Hobbs, "Three-dimensional kinematic comparison of treadmill and overground running," *Sports biomechanics*, vol. 12, no. 3, pp. 272–282, 2013.

[37] T. Kanade, P. Rander, and P. Narayanan, "Virtualized reality: Constructing virtual worlds from real scenes," *IEEE multimedia*, vol. 4, no. 1, pp. 34–47, 1997.

[38] F. Colloud, L. Chèze, N. André, and P. Bahuaud, "An innovative solution for 3d kinematics measurment for large volumes," in *16th Congress of the European Society of Biomechanics*, vol. 41, p. S57, 2008.

[39] U. G. Kersting, N. Kurpiers, B. J. Darlow, and V. W. Nolte, "Three-dimensional assessment of on water rowing technique: a methodological study," in *ISBS-Conference Proceedings Archive*, 2008.

[40] M. Begon, F. Colloud, V. Fohanno, P. Bahuaud, and T. Monnet, "Computation of the 3d kinematics in a global frame over a 40 m-long pathway using a rolling motion analysis system," *Journal of Biomechanics*, vol. 42, no. 16, pp. 2649–2653, 2009.

[41] G. R. Bernardina, T. Monnet, P. Cerveri, and A. P. Silvatti, "Moving system with action sport cameras: 3D kinematics of the walking and running in a large volume," *PloS one*, vol. 14, no. 11, p. e0224182, 2019.

[42] J. Chi, Z. Yang, G. Zhang, T. Liu, and Z. Wang, "A novel multi-camera global calibration method for gaze tracking system," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 5, pp. 2093–2104, 2020.

[43] M. Begon and P. Lacouture, "Accuracy of 3-d reconstruction with occlusions," *Journal of applied biomechanics*, vol. 26, no. 1, pp. 104–108, 2010.

[44] X. Zhou, S. Liu, G. Pavlakos, V. Kumar, and K. Daniilidis, "Human motion capture using a drone," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2027–2033, 2018.

[45] Codamtion. https://codamotion.com/.