CoMo: A Novel Comoving 3D Camera System

Andrea Cavagna, Xiao Feng, Stefania Melillo, Leonardo Parisi, Lorena Postiglione, and Pablo Villegas

***Abstract*— Motivated by the theoretical interest in reconstructing long 3D trajectories of individual birds in large flocks, we developed CoMo, a comoving camera system of two synchronized cameras coupled with rotational stages, which allows us to dynamically follow the motion of a target flock. With the rotation of the cameras, we overcome the limitations of standard static systems that restrict the duration of the collected data to the short interval of time in which targets are in the cameras common field of view, but, at the same time, we change, in time, the external parameters of the system, which have then to be calibrated frame by frame. We address the calibration of the external parameters measuring the position of the cameras and their three angles of yaw, pitch, and roll in the system *home* configuration (rotational stage at an angle equal to 0°) and combining this static information with the time-dependent rotation due to the stages. We evaluate the robustness and accuracy of the system by comparing reconstructed and measured 3D distances in what we call 3D tests that show a relative error of the order of 1%. The novelty of the work presented in this article is not only on the system itself but also on the approach that we use in the tests, which we show to be a very powerful tool in detecting and fixing calibration inaccuracies, and it, for this reason, may be relevant for a broad audience.**

***Index Terms*— 3D, 3D reconstruction, camera calibration, camera system, field data collection, panning system, wide field of view.**

I. Introduction

IN RECENT years, technological advances in the field of imaging and computer vision, together with the growing demand for 3D contents, contributed to make digital camera stereo systems accurate in the 3D reconstruction and, at the same time, accessible to a wide audience. This led to the proliferation of stereo vision applications in fields as diverse as entertainment [1]–[3], surveillance [4]–[7], navigation [8]–[11], robotics [12]–[15], medicine [16]–[19], and biology [20]–[23].

The experimental design of a 3D system is delicate because of the several factors that contribute to its reliability and feasibility, which strictly depends on the specific data to be gathered and on the environmental and logistic constraints of the data-acquisition location. Standard stereo systems are designed in a static fashion with the position and the orientation of the cameras fixed in time, thus, with a fixed field of view. This setup is suitable for most of the laboratory experiments, where the phenomena to be reconstructed happen in a confined volume, but it represents a severe limitation for nonconfined field experiments.

Ideally, when dealing with nonconfined phenomena, one would like to have a wide field of view and a high resolution of the system. However, in fact, this is not possible since both factors depend on the cameras' focal length, which needs to be short to have a large field of view, and it needs to be long to have a high resolution. Therefore, one has to lower the data-taking expectations, finding a compromise between the two factors, which, most of the time, ends up reducing both the field of view and the resolution of the system. A smarter, though more complicated, strategy is to replace the static setup with a dynamic one, effectively widening the field of view with a controlled rotation of the cameras aimed at following the targets [24]–[30]. The dynamic setup overcomes the limitation of the static one by actually breaking the link between the size of the field of view and the resolution of the system, which is now the only factor depending on the focal length. Hence, the resolution of the system can be set as high as needed without reducing the 3D volume covered by the system.

The rotation of the cameras makes the external parameters (orientation and position of the cameras in the world reference frame) time-dependent quantities that have then to be carefully calibrated frame by frame to guarantee high accuracy in the reconstruction of the scene. The literature suggests two different calibration approaches [31]: 1) 3D methods that reconstruct key points of calibrated 3D targets and estimate the external parameters as the ones that minimize the 3D reconstruction error [32]–[36] and 2) 2D methods that match features across the cameras, reconstruct the correspondent 3D points that are then projected back on the cameras, and estimate the external parameters as the ones that minimize the reprojection error [37]–[45].

In [46], we show that both approaches are essential to achieve high accuracy in the 3D reconstruction. The 2D methods give the best performance in the first step of the 3D reconstruction process, which consists of matching the images across the cameras. The 3D methods give the best performance in the second step of the 3D reconstruction process, where the point-to-point correspondences identified in the first step are used to triangulate and actually determine the 3D position of the targets.

In their standard implementations, both the methods start from a set of correspondences that should cover the entire field of view to guarantee the reliability of the two methods [33]. This is not problematic for 2D methods, where point-to-point

Manuscript received January 22, 2021; revised March 14, 2021; accepted April 3, 2021. Date of publication April 20, 2021; date of current version May 19, 2021. This work was supported by the ERC grant RG.BIO under Grant 785932. The Associate Editor coordinating the review process was Weiwen Liu. *(Corresponding author: Stefania Melillo.)*

The authors are with the Collective Behaviour in Biological Systems Lab (CoBBS Lab), National Research Council—Institute for Complex Systems (CNR–ISC), Sapienza University of Rome, 00185 Rome, Italy (e-mail: stefania.melillo@cnr.it).

This article has supplementary downloadable material available at https://doi.org/10.1109/TIM.2021.3074388, provided by the authors.

Digital Object Identifier 10.1109/TIM.2021.3074388

correspondences may be found all over the acquired images, but it represents a severe limitation for 3D methods in wide-field setup, where it is not always possible to cover the entire field of view with calibrated targets. The third approach, which is robust with respect to the 3D reconstruction accuracy regardless of the size of the field of view, consists of calibrating the external parameters of the system by directly measuring the orientation and position of all the cameras in a common reference frame. This latter approach represents a valid alternative to the 3D methods described above, but it is generally not used because it requires particular care in the system setup that has to be specifically designed to guarantee a precise measurement of the external parameters.

In this article, we present a novel comoving 3D system, CoMo, inspired by the human ability to follow the trajectory of a target with a coordinate movement of the eyes: cameras are coupled with rotational stages that drive a controlled rotation of all the cameras in the same direction and at the same rotational speed, in this way, dynamically adapting the field of view to the motion of the targets.

We developed and tested CoMo in the context of 3D data-taking of flocks of birds with cameras pointing at a wide region of the sky. This makes 3D standard methods for the calibration of the external parameters not appropriate. Therefore, for the calibration of the set of external parameters to be used in the 3D reconstruction process, we adopt the direct measure approach, measuring the position and the three angles of yaw, pitch, and roll of all the cameras in a common reference frame with the technique described in Section III-C2 and Appendix A in the Supplementary Material, while we use the standard 2D method described in [20] for the calibration of the parameters used for the identification of point-to-point correspondences across the cameras. We also propose a new procedure to improve the standard calibration of the camera focal length [47] that we found to be not sufficiently accurate for our purposes. We discuss this new procedure in Section VII-B, where we show how we could detect and fix the inaccuracy on the focal length by performing 3D reconstruction tests on calibrated targets.

We extensively tested CoMo to evaluate its performance in terms of the 3D reconstruction, see Section VII-C where we show that the comparison between reconstructed and measured 3D quantities on calibrated targets gives excellent results with a 3D reconstruction error of the order of 1%.

With a full-fledged experimental data-taking campaign in the field, we could also check the feasibility of the experiment with the CoMo setup, which proved to be easy to mount and easy to calibrate in the field. The data collected in the field confirmed that, with the comoving strategy, we can actually track the flocks significantly longer than with a standard static system, as shown in Video1 in the Supplementary Material.

This article is organized as follows. In Section II, we state the requirements for the system in terms of 3D reconstruction accuracy. In Section III, we describe the design of the system, our field setup, and the calibration procedure for both the internal and external parameters. In Section IV, we address the mathematical formalism of the 3D reconstruction for our dynamic system. In Section V, we show the tests that we performed on the equipment to measure the temporal offset between cameras and rotational stages and assess the camera synchronization and the camera frame consistency. In Section VI, we show the tests we performed on the rotational stages home repeatability and angle accuracy. In Section VII, we show in detail the tests on the 3D reconstruction accuracy of the system, and we introduce a novel approach to fix inaccuracies in the calibration of the cameras' focal length. Finally, in Section VIII, we show an example of the 3D reconstructed trajectories of a flock of starlings collected with our dynamic system.

II. 3D Reconstruction Accuracy Requirements

The requirements on the 3D reconstruction accuracy are strictly dependent on the application for which the data are collected. We collect field data of bird flocks with the aim of understanding the mechanisms behind the emergence of collective behavior, and in particular, we investigate the correlation properties of these systems [48], [49]. We mainly use the data to measure, how far (in space) and for how long (in time), the change in the direction of flight of a bird[1] influences the change in the direction of flight of the other birds in the flock.

In this framework, the absolute positions of the birds are not very useful, while the relevant quantities are the birds' directions of flight and the bird-to-bird distances. Therefore, we need CoMo to be particularly accurate in the 3D reconstruction of the distances between targets. More precisely, we require the relative error on the reconstructed 3D target-to-target distances to be: 1) not dependent on the position of the targets to avoid a spatial bias on the quantities that we compute; 2) not dependent on the instants of time where the targets live, to avoid a temporal bias on the quantities that we compute; and 3) smaller than 0.01, which we define to be the threshold of the accuracy acceptability.

We evaluated CoMo 3D reconstruction accuracy with the tests described in detail in Section VII, showing that the system fulfills all the requirements above.

III. CoMo System

In this section, we describe the hardware design of CoMo, its field setup, and the calibration procedure that we developed to fulfill the 3D reconstruction accuracy requirements listed in Section II.

A. Design

The design of CoMo is shown in Fig. 1: each of the two IDT OS10-4K cameras (resolution: 3840 px × 2400 px, sensor size: 17.9 mm × 11.2 mm, and frame rate: 155 fps), equipped with Schneider Xenoplan 28 mm f/2.0 optics,

[1] The 3D velocity vector, v, of a bird is given by $\Delta X/\Delta t$, where ΔX is the 3D displacement vector of the bird in the interval of time Δt. The bird direction of flight, $\hat{v}$, is defined as the velocity versor $\hat{v} = v/|v|$. The change in the direction of the bird is instead given by $\hat{v} - \hat{V}$, where $\hat{V}$ is the direction of flight of the group that is computed averaging the direction of flight of all the birds in the flock. The change in the direction of flight is, therefore, computed from the distance between the 3D position of the bird at times t and $t + \Delta t$.

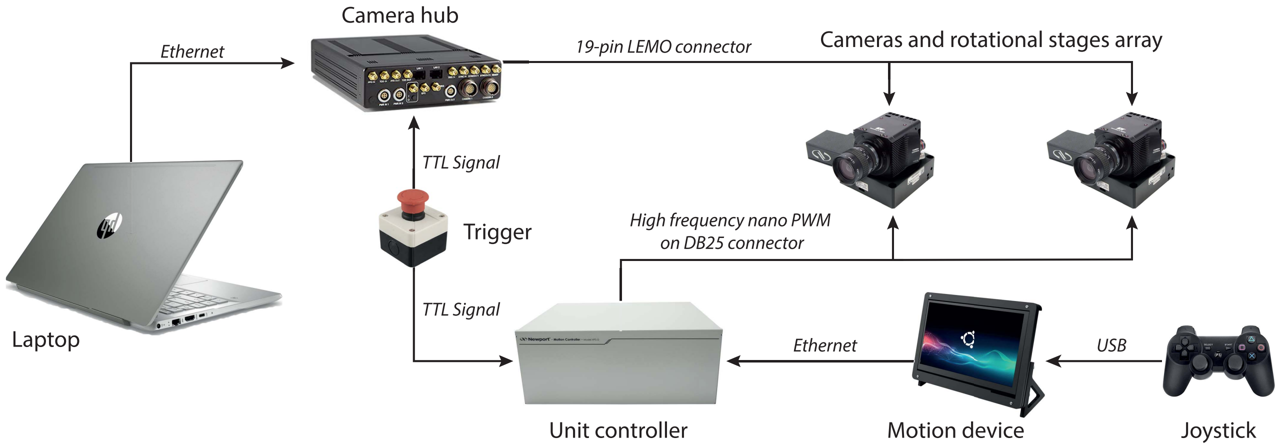


Fig. 1. Scheme of the system. The two IDT OS10-4K cameras (resolution: 3840 px × 2400 px, sensor size: 17.9 mm × 11.2 mm, and frame rate: 155 fps), equipped with Schneider Xenoplan 28 mm f/2.0 optics, are coupled with the two high-speed one-axis rotational stages (Newport RVS80CC; nominal accuracy: 10^{-4} rad and nominal home repeatability: $4 \cdot 10^{-3}$ rad). Each camera is connected with a 19-pin Lemo cable to the IDT TC-19 hub, which is also connected to a control laptop. The camera parameters, such as the exposure time, the sensitivity, and the frame rate, are manually set on the IDT proprietary software Motion Studio running on the laptop. They are sent via an Ethernet connection from the laptop to the hub, which redirects them to the cameras through the 19-pin Lemo cable using an IDT proprietary protocol. The hub sends to the cameras also the synch signal, which is generated by the hub itself. The direction and the speed of rotation of the stages are manually controlled via a Joypad Logitech F310 connected, via a UBS cable, to a motion device, namely, a Raspberry Pi 3 Model B+ connected to a 7” touchscreen. The motion device communicates, via an Ethernet connection, to the unit controller, which redirects the signals to the rotational stages on a DB25 cable, in the form of a high-frequency nano PWNM. The data acquisitions start with both the cameras and the stages in the *waiting from trigger mode* until they simultaneously receive the trigger signal, a 5 V TTL signal. The trigger signal is generated with a standard trigger button, and it is sent at the same time to the hub, which redirects the signal to the cameras, and to the unit controller, which redirects the signal to the stages.

is mounted on a high-speed one-axis rotational stage (Newport RVS80CC; nominal accuracy: 10^{-4} rad and nominal home repeatability: $4 \cdot 10^{-3}$ rad). The cameras are connected to the hub IDT TC-19 that has the double tasks of redirecting the signals from a laptop controller to the cameras and of synchronizing the cameras via a trigger and a synch signal.

1) Motion Control: The rotation of the stages is manually controlled by an operator via a motion device connected to a unit controller (XPS-RL4), which is also connected to the stages.

The data acquisition procedure starts with cameras and stages in *waiting for the trigger* mode, until they simultaneously receive a signal from a hardware trigger connected to the camera hub and the stage unit controller. We developed two different motion modes for CoMo.

a) Off-line motion mode: The speed and the direction of rotation are set before the acquisition starts independently for each stage.

b) Online motion mode: The speed and the direction of rotation may be chosen online from an operator via a joypad, but they are set to be equal for all three stages (see Appendix B in the Supplementary Material).

The two different motion modes have different applications: we use the off-line mode when performing tests on the system, where we need to be versatile on the cameras’ rotation, while we use the online mode when we collect data in the field, and it is of great importance to change the orientations of the cameras in real time in order to track the moving target.

B. Field Setup

We perform experiments on bird flocks in the urban environment of Rome, Italy, setting up CoMo on the roof of Palazzo Massimo alle Terme, Rome, Italy, in front of one of the bigger and more stable birds roosting site in Rome.

In this location, our working distance is about 150 m with a system baseline, i.e., the distance between the cameras, of about 25 m. The coupling between the cameras (with a sensor size of 17.9 mm × 11.2 mm) and the optics (with a focal length of 28 mm) produces a wide field of view of 35.5° in width and 22.6° in height.

C. CoMo Calibration

Our field setup with a working distance of 150 m and a wide field of view of 35.5° × 22.6° makes the calibration of both the internal and external parameters particularly tough.

We calibrated the internal parameters focal length, the position of the image center, distortion coefficients), and the external parameters (orientation and position of all the cameras with respect to a common reference frame in the 3D space) with two different procedures.

1) Calibration of the Internal Parameters: For the calibration of the internal parameters, we adopt a two-step procedure. In the first step, we use a standard calibration approach. We calibrate each camera separately in the lab using a standard calibration method based on [47]: we collect 50 images of a 13 × 19 checkerboard in different positions, we randomly pick 20 of these pictures, and we estimate the focal length, the position of the image center, and the first-order radial distortion coefficient. We iterate this process 50 times, and we choose each parameter as the median value obtained in the iterations.

For our dynamic setup, this standard calibration approach proved to be not accurate enough, producing a time-dependent 3D reconstruction error due to a slight miscalibration of the focal length, which we estimated to be of the order of 0.5%. Therefore, we designed a second step of the calibration to adjust the focal length, using the dynamic approach described in detail in Section VII-B1.

Note that, because of the large working distance and of the large field of view, we cannot perform the standard calibration of the internal parameters with a calibration target kept at the working distance while filling the entire field of view, as this would require a planar target of 98 m × 60 m. Therefore, we chose to reduce the distance of the calibration target in favor of filling the field of view.

This might be the reason for the miscalibration of the cameras' focal length obtained with the standard method in the first step of our calibration procedure. However, our results are also compatible with a different scenario, which may be the scope of the interesting future investigation: the standard calibration approach is less sensitive than the dynamic one to small variations of the estimated focal length; hence, these variations are boosted and, therefore more detectable, using dynamic information. This latter scenario suggests that the dynamic approach to the calibration may be an efficient and relatively simple strategy to improve the calibration performance both for static and dynamic system configurations.

2) Calibration of the External Parameters: In [46], we point out the need for two different sets of external parameters: the first set to be used to match points across the cameras and the second set to be used in the 3D reconstruction process. For the calibration of the first set of parameters, we use a standard 2D calibration procedure, and we refer the interested reader to [20], while here we focus on the calibration of the second set of parameters, i.e., the one used for the 3D reconstruction.

Our experimental setup, with a working distance of 150 m and with a large field of view, is not suitable to calibrate the external parameters with standard procedures. Our field of view is essentially a wide area of the sky, where we cannot locate any calibration 3D target; hence, we cannot use a 3D calibration method. We prefer to not use a feature-based calibration routine because of their low accuracy in the 3D reconstruction, which, in [46], we show to be higher than 1%. Therefore, we address the calibration with a different strategy.

CoMo external parameters are actually given by the combination of a static term, which does not change in time and describes the initial position/orientation of the cameras, and a dynamic term, which is time-dependent and describes the rotation of the cameras due to the stages. We directly measure these two terms separately.

We initially set the rotational stages in their *home* position, i.e., angle of rotation equal to 0 rad, and we set the pitch and roll angles of both cameras, respectively, to 0 and 0.22 rad using a clinometer (RS Pro Digital level 667-3916; accuracy: $3 \cdot 10^{-3}$ rad). We set the yaw angle of the left camera, α_L, to 0.11 rad and the yaw angle of the right camera, α_R, to -0.11 rad with a simple but effective technique, with which we achieve an accuracy of 10^{-3} rad and we extensively tested on static camera systems [50], [51] (see Fig. 2 and Appendix A in the Supplementary Material). With this procedure, we measure the orientation of both cameras in a common reference frame. To define the positions of the two cameras in the *real* 3D world, we still need to fix a metric scale factor that we calibrate by measuring the system baseline, i.e., the distance between the cameras, with a high precision range finder (Hilti Laser the probability distribution (PD)-E; accuracy: 1 mm).

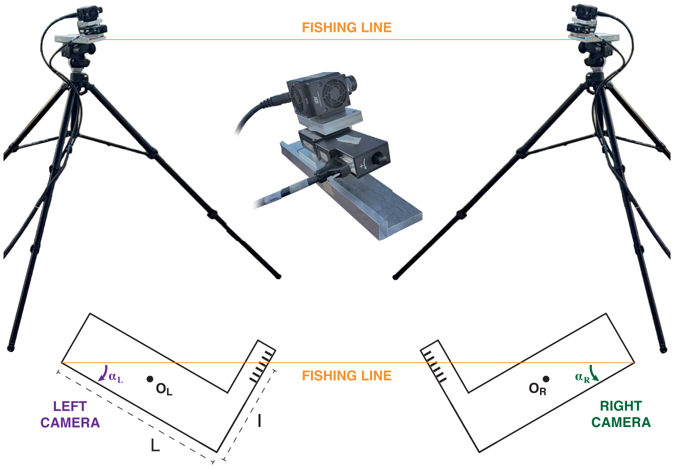


Fig. 2. Experimental setup. Each camera is mounted on a rotational stage that is locked on an L-shape bar and then on a tripod. The L-bars have a gauge on their small edge (on the left-hand side for the right camera and on the right-hand side for the left camera). We set the yaw angles of the cameras by tightening a fishing line, i.e., a thin nylon line, between the two external edges of the bars, so that the fishing line crosses the gauge, and it can be used as a pointer on the gauge. Denoting the long side of the L-bar with L and the distance from the point where the line crosses the gauge and the side of the bar with l, we can measure the yaw angles as $\mathrm{atan}(l/L)$, with the negative sign for the right camera and with the positive sign for the left camera. The accuracy of the measured angle is 10^{-3} rad, obtained as $\delta l/L$, with δl being the thickness of the wire.

We start the data acquisition by moving the cameras from this *home* calibrated configuration and recording the time-dependent angles of rotation of the stages. With a postprocessing procedure, we can then associate to each camera frame the correspondent external parameters combining the ones measured in the *home* configuration and the time-dependent rotation of the stages recorded during the data acquisition (see Sections IV-D and IV-E).

IV. Dynamic 3D Reconstruction

There is a vast literature about 3D reconstruction for static camera systems, i.e., system with fixed cameras orientation, [33]–[36], [38], [39], [41], [42], [44], [45]. Here, we move a step forward to generalize the 3D reconstruction theory to our dynamic system.

A. Camera Reference Frame

The camera reference frame $O_C xyz$ has the origin, O_C, in the camera optical point, the z-axis directed as the optical axis, and the xy plane parallel to the sensor with the x-axis pointing right and the y-axis pointing down (see Fig. 3). In our dynamic setup, this reference frame is not fixed in time, but it rotates on the xz plane around the camera optical center.

B. Pinhole Model

The pinhole camera model describes the mapping between the 3D real world and the 2D camera world as a central projection (see Fig. 3): the 2D image, q, of the 3D point

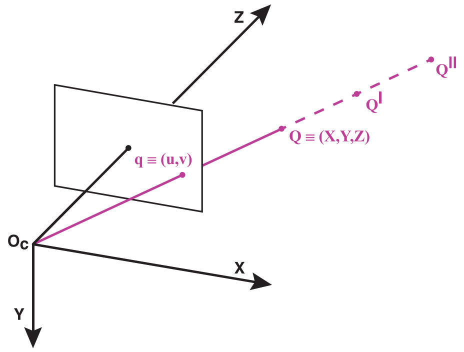


Fig. 3. Single camera. The camera reference frame has the origin in the camera optical point, O_C, the z-axis directed as the optical axis, and the xy plane parallel to the sensor with the x-axis pointing right and the y-axis pointing down. The pinhole model describes the relationship between the 3D world and the 2D camera sensor as a central projection: the image q of a 3D point lies at the intersection between the sensor and the line passing through Q and the camera optical center. This correspondence is not one-to-one because q is not only the image of the point Q but also of all the other 3D points belonging to the optical line, $O_C Q$. This ambiguity makes a single camera not sufficient for the 3D reconstruction.

Q lies at the intersection between the camera sensor and the line between Q and the camera optical center, O_C. Its natural mathematical framework is then projective geometry, where the correspondence between a 3D point $Q \equiv (X, Y, Z)$ and its 2D image $q \equiv (u, v)$[2] is expressed in a very simple formalism

$$\mathbf{q} = P \cdot \mathbf{Q} \tag{1}$$

where $\mathbf{q} = (\bar{u}, \bar{v}, \bar{w})$ is the 2D projective point corresponding to q, namely, $u = \bar{u}/\bar{w}$ and $v = \bar{v}/\bar{w}$, and $\mathbf{Q} = (X, Y, Z, 1)$ represents the homogeneous projective coordinates of Q [52]. P is the 3×4 matrix of the form $P = K \cdot [R|T]$, where K is the 3×3 matrix of the camera internal parameters, R and T are, respectively, the 3×3 rotation matrix and the three components translation vector that bring the camera reference frame in the world reference frame where Q lives, and they both depend on the external parameters of the system.

This definition of P can be further simplified by noting that

$$T = -R \cdot C \tag{2}$$

where C is the vector that connects the origin of the world reference frame to the origin of the camera reference frame; hence, $P = KR \cdot [I| - C]$, where I denotes the 3×3 identity matrix.

In a static camera, both R and C are fixed in time, but, in our dynamic system, the camera reference frame rotates about the camera optical center. Hence, the vector C is constant in time, while $R \equiv R(t)$. The time-dependent generalization of the projective matrix is then straightforward

$$P(t) = KR(t) \cdot [I| - C]. \tag{3}$$

C. World Reference Frame

We denote the two cameras' reference frames by $O_L x_L y_L z_L$ (for the left camera) and $O_R x_R y_R z_R$ (for the right camera).

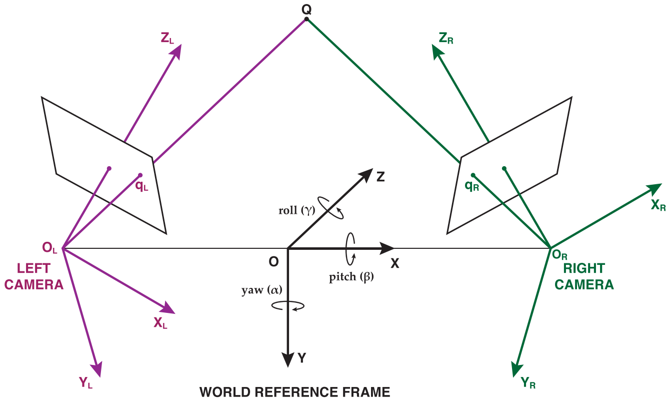


Fig. 4. Camera system. $O_L x_L y_L z_L$ and $O_R x_R y_R z_R$ represent the left and right camera reference frames. $Oxyz$ is, instead, the world reference frame, with the origin on the middle point of the camera baseline, $O_L O_R$, the x-axis pointing toward O_R, the y-axis pointing down along the world gravity axis, and the z-axis pointing outward following the right-hand rule. In this reference frame, the coordinates of the two camera centers are $C_L = (-d/2, 0, 0)$ and $C_R = (d/2, 0, 0)$. The circle arrows specify the positive direction and the axis of rotation for the yaw, pitch, and roll angles.

We define also a third reference frame, $Oxyz$, with the origin on the middle point of the camera baseline, $O_L O_R$ (see Fig. 4), the x-axis pointing toward O_R, the y-axis pointing down along the world gravity axis, and the z-axis pointing outward following the right-hand rule. This reference frame is fixed in time, and this is the reference frame within which we will reconstruct the scene. It is then in this reference frame that we need to express the projective matrices of the two cameras.

D. External Parameters

As we already stated in Section III-C2, with our setup, the external parameters are the combination of a static term, which describes the *home* configuration, and a dynamic term, which describes the rotation due to the stages; thus, the camera rotational matrices are of the form

$$R_C = R_{y_C}(-\varphi_C(t)) \cdot R_S(\alpha_C, \beta_C, \gamma_C) \tag{4}$$

where the subscript C indicates a generic camera (left or right), $R_{y_C}(-\varphi_C(t))$ is the time-dependent rotation, about the y-axis of the camera reference frame, which takes into account the rotation of the stage of an angle $\varphi_C(t)$, $R_S(\alpha_C, \beta_C, \gamma_C)$ is the static rotation matrix that takes into account the *home* orientation of the camera, and α_C, β_C, and γ_C are the angle of yaw, pitch, and roll, respectively.[3] In particular, for our system

$$R_S = R_{z_C}(-\gamma_C) \cdot R_{x_C}(-\beta_C) \cdot R_{y_C}(-\alpha_C). \tag{5}$$

Note that the order of the rotations in (5) is crucial, and it explicitly depends on the tripod model used in the experimental setup (see Appendix A in the Supplementary Material).

Note also that our choice of the world reference frame, with the origin at the center of the camera baseline and the x-axis

[2] For the sake of simplicity, in this article, we will refer to the 2D coordinate of an image point as defined in the image reference frame with the origin in the image center instead of the standard reference with the origin in the top left.

[3] In the world reference frame, the x-axis is parallel to the fishing line that we use to measure the two yaw angles, α_L and α_R (see Fig. 2), which are then automatically measured with respect to the world reference frame. A similar argument holds also for the pitch and roll angles that we measure using a clinometer because the y-axis of the world reference frame is parallel to the gravity direction.

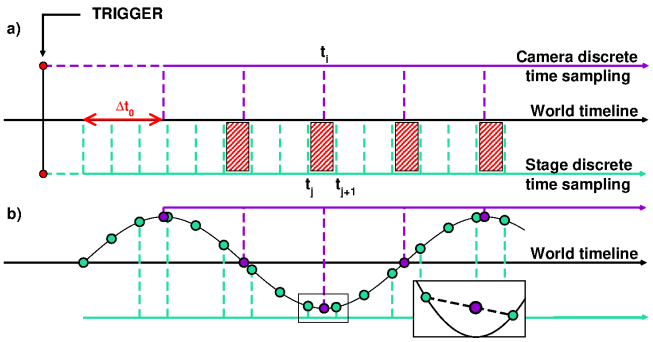


Fig. 5. Time discretization. (a) Two-time discretization of the camera (purple dashed line) and the stages (green dashed line) have to be matched to associate the position of the stage at each camera sample, i.e., frame. In the continuous world timeline, cameras and stages receive the trigger signal simultaneously, but, due to hardware lag time, they do not start to record immediately and, in general, not at the same time. We do not need to know the recording starting time of cameras and stages in the world timeline, but we need to measure the camera-stage offset (red double arrow on the world timeline axis). Once we know this offset, we can match each camera time sample, t_i, with its two closest time samples of the stage, t_j and t_{j+1}: $t_i \in [t_j, t_{j+1}]$. These intervals are highlighted with white and red striped boxes. (b) Black sinusoidal line represents the angle of rotation of the stage. The green circles correspond to the time samples of the stage, where the angle is actually measured, while the purple circles correspond to the camera time samples where we need to know the stage position. We associate to each camera time sample the angle obtained with a linear interpolation at time t_i between the two points $(t_j, \varphi(t_j))$ and $(t_{j+1}, \varphi(t_{j+1}))$.

pointing toward the right camera, makes the expression for the two camera centers, C_L and C_R, extremely convenient: $C_L = (-d/2, 0, 0)$ and $C_R = (d/2, 0, 0)$ with d being the length of the baseline.

E. Time Discretization

In Section IV-D, we derived the expression of the cameras' rotational matrices implicitly considering the time as a continuous variable, while, in the actual experimental setup, time is, in fact, measured in discrete steps, what we normally call *frames*.

In a standard static system, the only relevant time rate is one of the cameras. Time discretization can then be efficiently addressed by expressing all the dynamic quantities in the camera frame unit of time. In our dynamic system, we have instead two time rates: one of the cameras defined by the cameras' frame rate and one of the rotational stages defined by their sampling rate. Cameras and stages discretize time with two different rates (the cameras shoot at 155 fps, and the stages gather the data at 1000 Hz); see Fig. 5 where the camera and the stage sampling times are highlighted with purple and light green dashed lines, respectively. We reconstruct the position of the targets from the images; hence, our primary time rate is one of the cameras. In order to perform an accurate calibration of the external parameters, we need to match this primary time line with the secondary time line of the stages and associate the correct stage position at each camera frame.

In addition to these two discretizations of time, we have also the continuous world timeline. In the world timeline, cameras and stages receive the trigger signal simultaneously, but, due to hardware time lags, which are different for the cameras and the stages, they do not start to record immediately and, in general, not at the same time. We do not need to know the recording starting times with respect to the world reference, but it is crucial to know the time delay between cameras and stages, Δt_0, highlighted with a red arrow on the world timeline in Fig. 5. We measured Δt_0 with the procedure described in Section V-A, and we estimated a delay of 3 ms of the cameras with respect to the stages.

Once this time offset is measured, we can express the time corresponding to the camera frame and the time corresponding to the stage samples in the same reference, defining the ith camera time as $t_i = \Delta t_0 + i\,\Delta t_C$ and the jth stage time as $t_j = j\,\Delta t_S$, where $\Delta t_C = 1/155$ s and $\Delta t_S = 1/1000$ s denote the time steps of the cameras and the stages, respectively.

Finally, we associate to the ith camera frame, t_i, its two closest stage samples, t_j and t_{j+1}, such that $t_i \in [t_j, t_{j+1}]$ (see Fig. 5) where these last intervals are highlighted with white and red striped boxes, and we define the angle $\varphi_C(t_i)$ with a linear interpolation of the two angles $\varphi_C(t_j)$ and $\varphi_C(t_{j+1})$ measured by the stages.

F. 3D Reconstruction

The ambiguity of the camera projection, which associates to the same 2D image all the 3D points lying on the same optical line shown in Fig. 3, can be solved with two cameras (see Fig. 4): if q_L and q_R are the images of the same point, Q, in the left and the right camera, Q must lay on the two optical lines, one for each camera, passing through the two images, and it is then the point at the intercept between the two lines. In a mathematical formalism, this consists of solving the following system in the unknown $\mathbf{Q}$:

$$\begin{cases} \mathbf{q_L} = P_L(t) \cdot \mathbf{Q} \\ \mathbf{q_R} = P_R(t) \cdot \mathbf{Q} \end{cases} \tag{6}$$

where $\mathbf{q_L}$ and $\mathbf{q_R}$ are the 2D projective points corresponding to q_L and q_R and $\mathbf{Q} = (X, Y, Z, 1)$ is the 3D homogeneous projective point corresponding to Q. $P_L(t)$ and $P_R(t)$ are the projective matrices of the left and the right cameras defined as in (1), each with its own calibration matrix, K_L and K_R, its own rotation matrix defined as in (4), R_L and R_R, and its own center in the world reference frame, C_L and C_R.

In deriving (6), we assumed that, at each instant of time, we detect the exact position of the targets on the images, without considering any kind of noise. The direct effect of noise is that the two lines defined by the system (6) do not intersect anymore. Therefore, the 3D reconstructed coordinates cannot be found as the exact solution of the system but as its approximation, which we obtain using the standard DLT (direct linear triangulation) method in [52].

Note that, in (6), we identify the camera positions with the optical centers even though we do not know their exact position. We assume that the optical centers are located in the same position on the camera body (except for small fluctuations) because the factory design is the same for both cameras. We assume also that they are both located at the center of the camera body, which may be not completely correct. With this choice, we may then produce a misposition

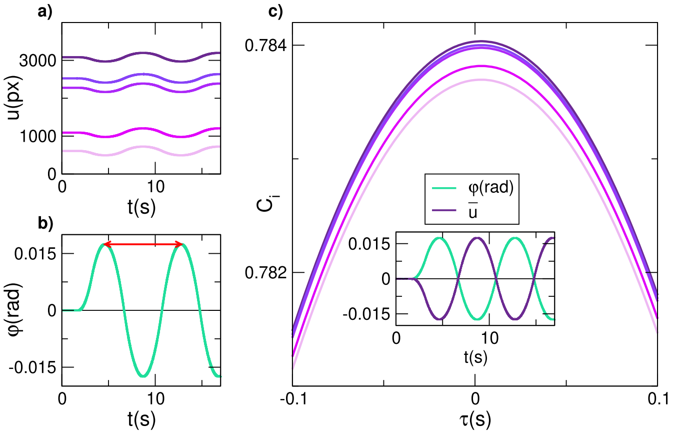


Fig. 6. Time offset. In order to measure the camera-stage time offset, we acquired images of five different targets while rotating the cameras with a periodic movement between 1° and $-1°$. The targets are still; hence, the rotation of the cameras due to the stages produces an apparent rotation of the 2D coordinates of the targets at the same speed but in the opposite direction. We estimate the offset from the cross correlation between the signal recorded by the stage and the position of the targets. (a) Evolution in time of the position of the five targets used in the test, each highlighted in a different color. (b) Signal is recorded by the rotational stage. (c) Correlation function $C_i(\tau)$ for each of the five targets. The maximum of all the cross correlation functions occurs at the same time, which is the offset Δt_0. Inset: the angle recorder by the stage, green line, and the position of one of the targets, purple line, normalized to be represented on the same scale in the plot. The comparison between the two signals shows the apparent movement of the target at the same speed of the stage but in the opposite direction.

of the two cameras, which, in principle, may affect the 3D reconstruction accuracy of the system. However, the error that we are introducing is a systematic error, i.e., equal for both cameras; hence, we may induce a systematic error on all the 3D reconstructed points, namely, a *solid translation* of the 3D world. This may be relevant for the accuracy of the absolute position of the targets, but it does not affect the accuracy of the mutual distance between pairs of targets, which is what we are interested in, as we stated in Section II.

V. Time Discretization: Tests

We extensively tested the equipment to measure the time offset Δt_0 between the cameras and the stages (see Section IV-E) and highlighted with a red arrow in Fig. 5. We also checked the consistency of the cameras' frame rate and the synchronization between the cameras.

A. Time Offset

We measured the time offset between the cameras and the stages recording images of five targets (2×2 cardboard checkerboards) while rotating the stages with a periodic movement between 1° and $-1°$, starting with the stages in their *home* position.

The targets are still; hence, the rotation of the cameras (due to the stages) produces an apparent rotation of the 2D coordinates of the targets: if a camera rotates in the clockwise direction at a certain speed, we will detect a rotation of the u-coordinate of the targets with the same speed but in the counterclockwise direction, and vice versa, a counterclockwise rotation of the camera corresponds to a clockwise rotation of the targets. Therefore, we can estimate the time offset comparing the signal gathered by the stages with the position of the targets (see Fig. 6(a) where we plot the u-coordinates of the five targets and Fig. 6(b) where we plot the angle recorded by the stage as a function of time). We compute the cross correlation of the two signals, taking care of the following three factors: 1) the two signals are recorded with different time discretization; 2) the duration of the signals is finite in time; and 3) targets positions are not centered in 0.

We oversampled the signal from the cameras, i.e., the u-coordinate of the targets, with linear interpolation. In this way, we resampled the camera signal at 1000 Hz (the gathering frequency of the stages) so that the time resolution of the cross correlation is defined by the time discretization of the rotational stage. We took care of the finite duration of the two signals restricting the signal of the stage at one period (from the first to the second maximum), highlighted with a red arrow in Fig. 6(b). Finally, we normalized the target coordinates by subtracting their *home* position (see the inset of Fig. 6(c) where we plot the signal from the stage in light green and the position of one of the targets in purple, normalized to be on the same y-scale of the stage).

We define the cross correlation between the signal of the stage and the coordinate of the ith target as

$$C_i(\tau) = \frac{1}{T-\tau}\sum_{t=0}^{T}\varphi(t)\bar{u}_i(t+\tau) \tag{7}$$

where $\bar{u}_i$ is the normalized position of the ith target. For each target, we can define τ_i as the point where $C_i(\tau)$ reaches its maximum. We found that all the targets have the maximum of $C_i(\tau)$ at the same point [see Fig. 6(c)], which is the time offset Δt_0 between the cameras and the stages, and which we estimated to be equal to 3 ms.

B. Frame Rate Consistency and Cameras Synchronization

We checked the frame rate consistency and the synchronization between the cameras using a chronometer that we built specifically for these tests: a needle spins at a constant rotational velocity (20 rps) over a protractor. Knowing the rotational speed of the needle, the frame rate and synchronization accuracy are then directly measured from the angle between the positions of the needle in two different images.

We tested the frame rate consistency for each camera separately, by measuring the angle span by the needle between two subsequent images. We found a negligible error, i.e., the error is below our resolution of $7 \cdot 10^{-5}$ s corresponding to 0.5° at a rotational speed of 20 rps. We also tested the synchronization between the cameras, comparing the position of the needle on the images acquired at the same time frame from different cameras, and again, we found a negligible error.

VI. Yaw Angles Accuracy in Time

The accuracy of the time-dependent yaw angles, $\varphi_L(t)$ and $\varphi_R(t)$, depends on two factors: the rotational stage home repeatability and the accuracy on the interpolation we use to compute $\varphi_L(t)$ and $\varphi_R(t)$, as described in Section IV-E. We evaluated the accuracy both on the home repeatability and the interpolation on each pair camera/stage separately, by performing the tests shown in this section.

A. Stage Home Repeatability

In the rotational stage *home* procedure we include the initialization of the stage, namely we first initialize the stage, and then, we move it to the home position. The unit controller offers also a direct procedure to home the stage from a generic position, but we chose the indirect procedure because of its higher consistency (see Appendix C in the Supplementary Material for more details).

We denote by φ_0 the stage home position. We cannot have an absolute measure of φ_0; hence, we measure its fluctuation, $\Delta\varphi_0$. With the camera mounted on the stage, we collect a set of 100 images of seven targets (2×2 checkerboard) acquired after the initialization and homing procedure of the stage. Between two consecutive acquisitions, we initialize the stage and send it to the home position. We detect the targets on the images with the subpixel routine in [53] that associates to each target the position of its central corner, and we measure the angular fluctuation of the home position within each pair of consecutive images as the displacement of the targets, normalized by the camera focal length $\boldsymbol{\Omega}$.

To evaluate the natural fluctuations in the targets positions due to the detection routine, we perform a first test that we will use as a reference acquiring a set of 100 images with the stage still in the home position. We compute the PD of the angular fluctuations, as highlighted in black in Fig. 7. Then, we perform the actual home repeatability test acquiring a set of 100 images with the stage in the home position, after performing the initialization and homing procedures, and we compute the PD of this homing procedure fluctuations, as highlighted in purple in Fig. 7, which shows a zero median and a zero mean (median equal to $4.7 \cdot 10^{-7}$ rad and mean equal to $6.2 \cdot 10^{-7}$ rad) and values smaller than $6 \cdot 10^{-5}$ rad.

The plot shows the high compatibility of the two PDs; hence, we conclude that the error on the home position is negligible and smaller than the one guaranteed by the factory equal to $4 \cdot 10^{-3}$ rad.

B. Angle Interpolation

In order to measure the error on the interpolation of the angles recorded by the stages, we perform the following test: we separate cameras and stages, and we stuck on each stage a 13×19 checkerboard printed on a foam board. We put the stage in rotation, and we acquired images of the rotating checkerboard keeping the camera still, starting with the rotational stage in the *home* position.

We use the evolution in time of the position of the checkerboard corners to estimate the angle of rotation of the stage, in this way, computing the rotation angle with a method that does not depend on the angular position gathered from the stage.

For each image, we detect the corners of the checkerboard with the subpixel routine in [53]. We use the first part of the acquisition, when the stage is still, to define a reference position for each corner, namely, we associate to each corner the average of its coordinates over all the images with the stage in its home position. Then, we compute the angle of rotation of the stage corresponding to a given camera frame, t, using the Kabsch algorithm [54]. More in detail, we associate to each frame t the rotational matrix that minimizes the root mean squared deviation (RMSD) computed with the Kabsch algorithm between the positions of the corners detected at time t and the reference positions. Finally, we compared the angle found with the Kabsch algorithm and the angle that we would associate with the same camera frame interpolating the angular positions gathered by the stages.

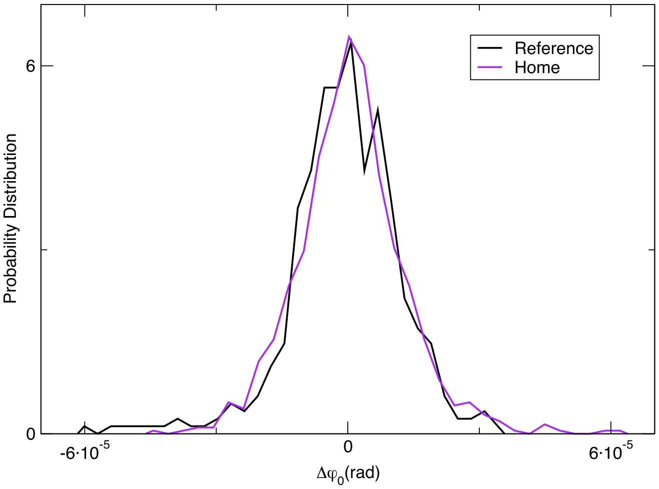


Fig. 7. Home repeatability. The PD obtained while acquiring images with the stage still represents the reference distribution for our test. The reference distribution, highlighted in black, gives the measure of the fluctuations due to the target's detection routine. The PD of the fluctuations on the home position of the stage, highlighted in purple, obtained homing the stage after its initialization procedure. The PD is compatible with the reference distribution, with the fluctuations smaller than $6 \cdot 10^{-5}$ rad and with a zero median and a zero mean (median equal to $4.7 \cdot 10^{-7}$ rad and mean equal to $6.2 \cdot 10^{-7}$ rad).

We carried out this test with the stages performing periodic rotation in three different configurations, corresponding to different choices of the parameters.

1) *Slow:* $\varphi_{\max} = 2°$, $\upsilon_{\max} = 1°/\text{s}$, and $a_{\max} = 0.5°/\text{s}^2$.
2) *Moderate:* $\varphi_{\max} = 10°$, $\upsilon_{\max} = 10°/\text{s}$, and $a_{\max} = 10°/\text{s}^2$.
3) *Fast:* $\varphi_{\max} = 18°$, $\upsilon_{\max} = 36°/\text{s}$, and $a_{\max} = 72°/\text{s}^2$.

Here, $\upsilon_{\max}$ and $a_{\max}$ are the maximum speed and the maximum acceleration reached by the stages, and $\varphi_{\max}$ denotes the amplitude of the periodic rotation, i.e., the stage performs a periodic rotation between $\varphi_{\max}$ and $-\varphi_{\max}$.

The results of these tests are shown in Fig. 8 where, in the first column, we plot the angle gathered by the stage in the three different tests, and in the second column, we show the PD of the error on the angle, $\Delta\varphi$, defined as the difference between the interpolation of the angle measured by the stage and the angle measured via the Kabsch algorithm.

As expected, we found that the error grows with the speed of the rotation because of a decreasing accuracy in the interpolation, but, in all cases, we found an error smaller than

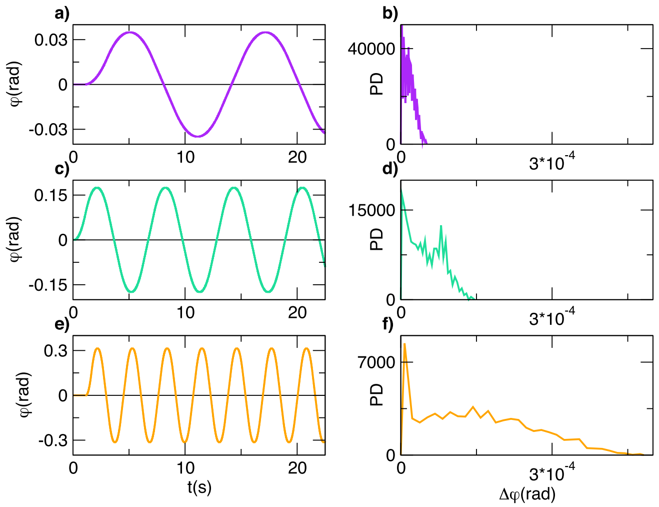


Fig. 8. Angle accuracy. (a), (c), and (e) Angle gathered by the stage in the three different tests (slow, moderate, and fast). (b), (d), and (f) PDs of the error on the angle, $\Delta\varphi$, defined as the difference between the interpolation of the angle measured by the stage and the angle measured with the Kabsch algorithm. The first row refers to the slow configuration, the second row refers to the moderate configuration, and the third row to the fast configuration. As expected, the error grows with the speed due to the decrease of the interpolation accuracy with the speed, and in all three cases, the error is below $5 \cdot 10^{-4}$ rad, being smaller than $5 \cdot 10^{-5}$ rad for the slow configuration.

$5 \cdot 10^{-4}$ rad, being smaller than $5 \cdot 10^{-5}$ rad for the slowest test, which can be considered negligible for all our practical purposes.

VII. System Accuracy Evaluation: 3D Tests

The question at the very core of all 3D reconstruction systems is: how accurate is the system in reconstructing the position of an object Q at a specific time t? Answering this question is not straightforward, especially if, as in our case, experiments are performed in the field where the system cannot be mounted and calibrated once and for all. We believe that a fair answer can only be given by checking reconstructed quantities against reality.

This is what we actually do for our system in what we call 3D tests: with a laser range finder (Hilti Laser PD-E; accuracy: 1 mm), we measure the distance between pair of targets in the common field of view of the cameras, we reconstruct the position of the targets in our world reference frame, and from these positions, we compute reconstructed target-to-target distances. Finally, we compare reconstructed and measured distances, and we compute the percentage error on the measured distances.

We perform the 3D tests in two different fashions.

1) *Static 3D Test:* Cameras are set up in their *home* configuration, and they do not move during the data acquisitions.
2) *Dynamic 3D Test:* Cameras rotate during the data acquisition.

A. 3D Test Setup

We evaluate the 3D reconstruction accuracy of the system, checking that the requirements described in Section II are fulfilled, performing the tests described in detail in Section VII-B. In principle, we should perform the tests exactly in the experimental configuration: camera baseline at 25 m, targets at a distance from the cameras in the range between 100 and 150 m, and pitch angles of both cameras set to 0.22 rad.

However, due to logistic constraints, we are forced to perform the tests in a slightly different configuration: 1) we set the camera baseline at about 10 m with targets at a distance from the cameras in the range between 20 m and 40 m and 2) we do not manage to have targets in the common field of view of the cameras for a pitch value of 0.22 rad, but we can achieve the maximum pitch of 0.15 rad. We take care of these two logistic limitations in the design of the test and in the data analysis (see Section VII-C).

B. Accuracy on the Calibration of the Internal and External Parameters

To evaluate the accuracy of the calibration procedures, we perform 3D tests in a special configuration where we can write the explicit coordinates of the reconstructed points. To this aim, we set pitch and roll angles of both cameras equal to 0, and we obtain the following explicit form of the Z-coordinate of a 3D point (see Appendix D in the Supplementary Material):

$$Z(t) = \frac{\Omega d}{s(t) - (\alpha + \varphi(t))\Omega} \tag{8}$$

where d is the system baseline, i.e., the distance between the cameras that we measure with the laser range finder, Ω is the cameras focal length,[4] $s(t) = u_L(t) - u_R(t)$ is the disparity, $\alpha = \alpha_R - \alpha_L$ and $\varphi(t) = \varphi_R(t) - \varphi_L(t)$ are the mutual orientation of the cameras due to the system *home* configuration and the rotation of the stages, respectively.

From (8), we obtain the explicit expression of the relative error on Z, $\delta Z/Z$

$$\frac{\delta Z(t)}{Z} = \frac{\delta d}{d} + \frac{\delta\Omega}{\Omega} + \frac{Z}{\Omega d}[\psi(t)\delta\Omega + \Omega\delta\psi(t)] \tag{9}$$

where $\psi(t) = \alpha + \varphi(t)$, and δd, $\delta\Omega$, and $\delta\psi$ denote the error on d, Ω, and ψ. Note that, here, we are not considering the contribute due to the error on $s(t)$, i.e., error in the position of the targets on the images, because it is not relevant in the 3D test setup (see Appendix A in the Supplementary Material).

Denoting the distance between two targets as ΔZ, we also obtain that

$$\frac{\delta(\Delta Z)}{\Delta Z} = \frac{\delta d}{d} + \frac{\delta\Omega}{\Omega} + 2\frac{\bar{Z}}{\Omega d}[\psi(t)\delta\Omega + \Omega\delta\psi(t)] \tag{10}$$

where $\bar{Z}$ is the mean Z-coordinate of the two targets.

We do not measure the absolute positions of the targets but their mutual distances, ΔR; hence, in the 3D test, we can only estimate the error on these distances, $\delta(\Delta R)$. In Appendix A in the Supplementary Material, we show that ΔR is proportional to ΔZ, which means that $\delta(\Delta R)$ is proportional to $\delta(\Delta Z)$. Therefore, we can write the explicit expression of the relative

[4] For the sake of simplicity, we are assuming the same value of the focal length for both cameras.

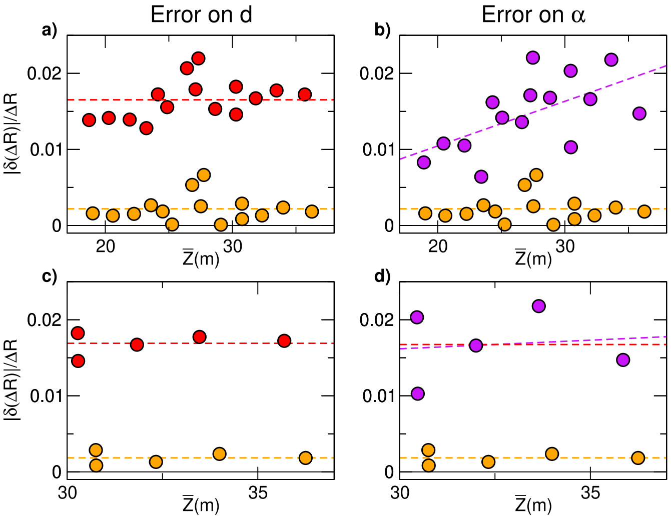


Fig. 9. Static 3D test. The plots show $|\delta(\Delta R)|/\Delta R$ for each pair of targets as a function of their mean distance from the cameras, $\bar{Z}$. The orange circles represent the result of the 3D test obtained with the original calibration parameters. The orange dashed line is the mean value of $|\delta(\Delta R)|/\Delta R$. Red circles (left column) represent the results obtained by manually introducing an error of 0.1605 m in the baseline length ($\delta d/d = 0.015$), while purple circles (right column) represent the result obtained by manually introducing an error of 0.003 rad in the angle α. (a) Error on d produces an increment of the error, constant for all the targets. A constant fit of the data (red dashed line) gives an estimate of $\delta d/d$ equal to 0.016, compatible with the experimental $\delta d/d = 0.0015$. (b) Error on α produces an increment of the error linear in Z. A linear fit of the data (purple dashed line) gives a slope equal to 0.00059 m^{-1}, which corresponds to $\delta\alpha = 0.0031$ rad in perfect agreement with the experiment. (c) When reducing the span of Z, the error due to d is still well-estimated, with a constant fit that predicts an error on $\delta d/d$ equal to 0.016. (d) When reducing the span of Z, the error due to α cannot be detected and estimated properly. The constant fit (red dashed line) and the linear fit (purple dashed line) are both compatible with the data.

error on target-to-target distances by substituting $\delta(\Delta Z)/\Delta Z$ with $\delta(\Delta R)/\Delta R$ in (10), which gives

$$\frac{\delta(\Delta R)}{\Delta R} = \frac{\delta d}{d} + \frac{\delta\Omega}{\Omega} + 2\frac{\bar{Z}}{\Omega d}[\psi(t)\delta\Omega + \Omega\delta\psi(t)]. \quad (11)$$

Equation (11) shows that $\delta(\Delta R)/\Delta R$ is made of a constant term, which depends on the error on d and on Ω, and a linear term in $\bar{Z}$, which depends on the error on ψ and Ω.

The idea now is to use the information of (11) to detect potential sources of error in the system.

From the trend of $\delta(\Delta R)/\Delta R$ in $\bar{Z}$, we can make the first discrimination between errors due to an incorrect measure of the baseline versus errors due to inaccuracies in Ω and ψ, as shown in Fig. 9, where we present the effect on a static 3D test of an error on d or of an error in α. The difference between the two results is evident: an error on d produces an increase of the errors constant with Z, while the error on α produces errors with a trend in Z. We stress here that it is possible to discriminate between the two situations only if the span in Z of the targets is large enough (see Fig. 9) where, in the bottom figure, we show how the results would have looked like with a short span in Z. To further discriminate between an error in Ω and an error in $\psi(t)$, we need dynamic information. To this aim, we derive Z with respect to time, and we obtain the following expression:

$$\partial_t(\delta Z) = \frac{Z^2}{\Omega d}[\partial_t\varphi(t) \cdot \delta\Omega] \quad (12)$$

which tells us that the evolution in time of the error on Z is quadratic in Z with a coefficient that depends on the rotational speed, $\partial_t\varphi$, and the error on the focal length, $\delta\Omega$.

In Section III-C1, we mentioned that we need a two-step procedure for the calibration of the internal parameters because of the low accuracy on the estimation of Ω with the standard calibration approach. We will use this last equation to show how to detect and how to quantify the error $\delta\Omega$. Once we corrected the error on Ω, we can go back to (11) and check for a potential error on the cameras' orientation, with the tests described in Section VII-B.

1) Improving the Focal Length Calibration: We check the accuracy of the standard calibration of Ω with the following 3D test: we put in rotation one camera per time at a constant rotational speed ($\upsilon = 6°/s$). We check Ω of the left camera rotating only the left camera in the clockwise direction

$$\partial_t\varphi_L(t) = \upsilon \text{ and } \partial_t\varphi_R(t) = 0 \quad (13)$$

while we check Ω of the right camera rotating only the right camera in the counterclockwise direction

$$\partial_t\varphi_L(t) = 0 \text{ and } \partial_t\varphi_R(t) = -\upsilon. \quad (14)$$

Therefore, in both tests, $\partial_t\varphi(t) = \partial_t\varphi_R(t) - \partial_t\varphi_L(t) = -\upsilon$, and (12) reads

$$\partial_t(\delta Z(t)) = -\upsilon\frac{Z^2}{\Omega d}\delta\Omega. \quad (15)$$

Note that δZ is the reconstruction error; hence, $\delta Z(t) = Z_{3D}(t) - Z$, where Z_{3D} is the reconstructed Z. This implies that $\partial_t(\delta Z(t)) = \partial_t Z_{3D}(t) - \partial_t Z$, but the targets are still; hence, their position is constant in time and $\partial_t Z = 0$. Equation (15) can then be written as

$$\partial_t(Z_{3D}(t)) = -\upsilon\frac{Z^2}{\Omega d}\delta\Omega \quad (16)$$

which tells us that the derivative of Z_{3D} with respect to time is constant for each target, and it linearly depends on the speed of rotation and the error in Ω. We checked the evolution in time of $Z_{3D}(t)$, and we found the linear trend shown Fig. 10(a) and (f), which is also the reason for the large error bars of $\delta(\Delta R)/\Delta R$ in Fig. 10(b) and (g).

Equation (16) tells us more because it shows that $\partial_t Z_{3D}(t)$ is quadratic in Z, which means that targets at different distances from the cameras will have a linear trend in time with different slopes: the further apart the target the higher the slope. With a linear fit of $Z_{3D}(t)$, we computed $\partial_t Z_{3D}(t)$ for each target, and we plot these quantities versus $\langle Z_{3D}^2\rangle_t$[5] [see insets in Fig. 10(c) and (h)]. From these last plots, we estimated $\delta\Omega$ with a linear fit, and we found an error of 41.61 px for the left camera ($\Omega_L = 6356.41$ px with the standard calibration and $\Omega_L = 6314.8$ px with the dynamic calibration) and an

[5] $\langle Z_{3D}^2\rangle_t$ is the average in time of $Z_{3D}^2(t)$, and it is the most accurate estimate of Z that we can give since we do not measure the absolute position of the targets but targets mutual distances.

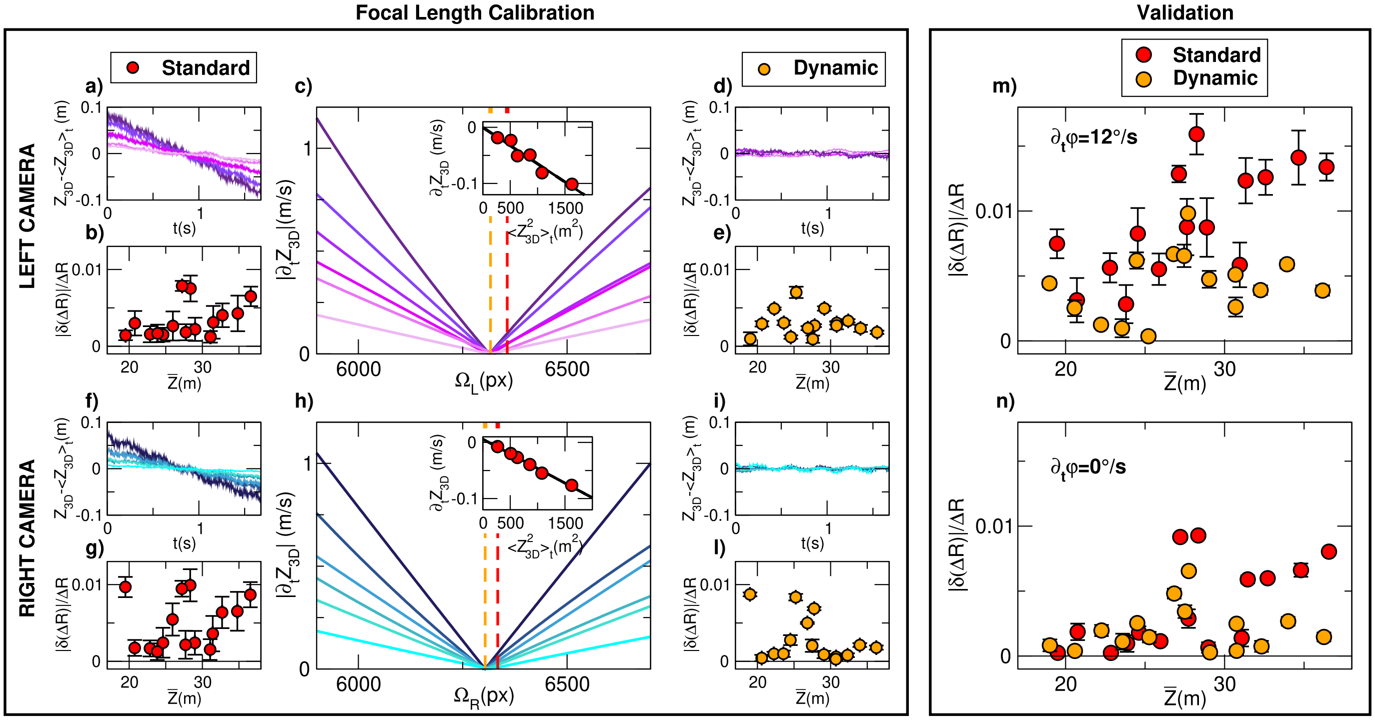


Fig. 10. Improving the focal length calibration. In the left box, data refer to the calibration improvement procedure while on the right box to its validation. Left box: the top part refers to the calibration of the left camera and the bottom part to the calibration of the right camera. Data are collected with a dynamic 3D test with one camera per time in rotation at a constant speed $v = 6°/s$. In the first column, we show the results of the 3D test obtained with the standard Ω, i.e., Ω calibrated with the standard method. In the right column, we show the same quantities but obtained with the dynamic Ω, i.e., Ω calibrated with the dynamic procedure. (a), (d), (f), and (i) Reconstructed Z, $Z_{3D}(t)$, for all the targets, each highlighted with a different color. $Z_{3D}(t)$ is normalized by its mean in time to have all the targets in the same range. (a) and (f) Standard Ω: $Z(t)$ shows a linear trend in t. (d) and (i) Dynamic Ω: $Z(t)$ does not show any trend in t. (b), (e), (g), and (l) Mean in time of $\delta(\Delta R)/\Delta R$ for each pair of targets as a function of the pairs mean distance from the cameras, $\bar{Z}$. Error bars are computed as standard deviation. (b) and (g) Standard Ω: large error bars reflect the high variability of the targets $Z(t)$ due to their linear trend in t. (d) and (l) Dynamic Ω: error bars are in most of the cases smaller than the symbols, and they reflect the absence of the trend in time of the targets $Z(t)$. (c) and (h) Absolute value of the slope of the reconstructed Z, $|\partial_t Z_{3D}(t)|$, as a function of Ω. At a fixed value of Ω, the slope increases with the target distance from the camera, which is embedded in the color code, going from light purple and light blue for the closest target to dark purple and dark blue for the furthest. All the targets present a well-defined minimum of the slope for the same value of Ω, highlighted with orange dashed line, which corresponds to the dynamic Ω, while the standard Ω is highlighted with the red dashed line. In the inset, we show the linear trend of $\partial_t Z_{3D}$ with the average of Z_{3D} in time, $\langle Z^2_{3D}\rangle_t$ for the standard Ω. Right box: we validate the dynamic calibration comparing the absolute value of the relative error in the target-to-target distances using the focal length obtained with the standard (red circles) and dynamic (orange circles) calibration. (m) We tested the dynamic calibration with a dynamic 3D test rotating both cameras simultaneously at a speed of 6°/s in the two opposite directions. The plot shows that, with the dynamic calibration, we obtain smaller relative errors and much smaller error bars than with the standard calibration. Moreover, we see that the trend in Z that is quite evident for the standard calibration becomes negligible with the dynamic calibration. (n) We validate the dynamic calibration on a 3D test reproducing our experimental procedure, with both cameras rotating simultaneously and in the same direction. Here, we do not appreciate a decrease of the error bars because the effect of $\delta\Omega$ is negligible due to the effective speed $v = 0°/s$, but we still see that the overall errors get smaller.

error of 33.52 px for the right camera ($\Omega_R = 6333.81$ px with the standard calibration and $\Omega_R = 6300.29$ px with the dynamic calibration).

But we can estimate $\delta\Omega$ more precisely with a different strategy: we run again the analysis of the 3D test moving the value of Ω in the interval [5900 px, 6700 px], and for each value of Ω, we compute $|\partial_t Z_{3D}(t)|$ of each target. We found that all the targets have a well-defined minimum of $|\partial_t Z_{3D}(t)|$ that occurs at the same value of Ω [see Fig. 10(c) and (h)]. We choose then Ω corresponding to this minimum as our new calibrated focal length, i.e., the dynamic Ω highlighted with a dashed orange line in Fig. 10(c) and (h). With this procedure, we found an error on Ω for the left camera equal to 40 px and for the right camera equal to 30 px, compatible with the estimate obtained from the linear fit of $\partial_t Z_{3D}$ versus Z^2_{3D}. We checked that, using this dynamic Ω, Z_{3D} does not show anymore a trend in t, and we also found a reduction of the error bars for $\delta(\Delta R)/\Delta R$, as shown in Fig. 10(d), (e), (i), and (l).

We validated the dynamic calibration by performing other two 3D tests in different conditions. We perform the first test rotating both the cameras simultaneously at a constant speed of 6°/s but in opposite directions, in this way, amplifying a potential error on Ω: we rotate the left camera in the clockwise direction, $\partial_t\varphi_L(t) = -v$, and the right camera in the counterclockwise direction, $\partial_t\varphi_R(t) = v$; hence, the effective rotational speed $\partial_t\varphi$ is equal to 12°/s. We also performed a second test to simulate the experimental setup, thus rotating the cameras in the same direction at the same speed, $\partial_t\varphi_L(t) = \partial_t\varphi_R(t) = v$, with an effective rotational speed, $\partial_t\varphi(t) = 0°/s$.

The results of these two tests are shown in Fig. 10(m) and (n), where the red circles refer to the standard calibration and the orange circles to the dynamic calibration. As expected the effect of the standard Ω is

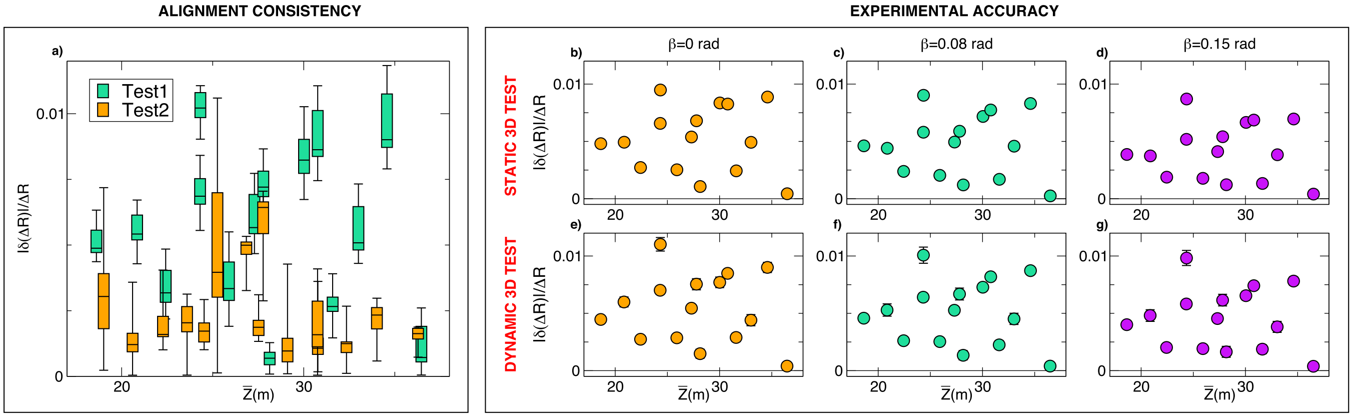


Fig. 11. System accuracy. Left box (panel a): orange and green boxplots represent the relative error in the 3D reconstruction as a function of $\bar{Z}$ for two different sets of static 3D tests. Data from the two sets are collected mounting and unmounting the entire system, which justifies differences in the $\bar{Z}$ values for the two tests. Data within the same set are collected by repeating the alignment procedure with the fishing line (eight times for the orange set and ten times for the green one). The line inside the box corresponds to the median of the relative error in the 3D reconstruction for a single target-to-target distance, the two edges of the box correspond to the first and the third quartiles, and the two whiskers correspond to the minimum and the maximum value. The plot shows no trend in $\bar{Z}$, hence showing a not appreciable error on the angle α. The data show variability within the same test (quite large error bars), due to the alignment, and also variability within the two different tests, due to the setup procedure, but this does not affect the accuracy and the consistency of the 3D reconstruction that gives always relative errors smaller than 0.012. Right box (panels b–g): data presented in the first and second rows are collected performing, respectively, static (panels b–d) and dynamic (panels e–g) 3D tests for different values of β. The dynamic tests are performed in the field configuration with the cameras rotating simultaneously at the same speed and in the same direction. The plots show $|\delta(\Delta R)|/\Delta R$ for each pair of targets as a function of their mean distance from the cameras, $\bar{Z}$. Static tests are performed shooting one single image; hence, we do not have error bars. For the dynamic tests instead, we plot $|\delta(\Delta R)|/\Delta R$ averaged in time, and error bars, which are most of the times smaller than the symbols, represent standard deviation. We do not see any trend of the error with β nor in the static tests nor in the dynamic tests. The comparison between static and dynamic tests at a fixed value of β shows relative errors of the same order and always smaller than 0.01.

more evident in the test at $\partial_t \varphi(t) = 12°/\text{s}$ where we see large error bars of $|\delta(\Delta R)|/\Delta R$ and also a trend with $\bar{Z}$, while, for $\upsilon = 0°/\text{s}$, error bars are quite small. In both tests, the dynamic Ω reduces $|\delta(\Delta R)|/\Delta R$, and it makes the error bars for the test at $\upsilon = 12°/s$ comparable with the ones of the test at $\partial_t \varphi(t) = 0°/\text{s}$. These two factors, lower $|\delta(\Delta R)|/\Delta R$ and smaller error bars, confirm that the dynamic Ω is more correct than the one obtained with the standard calibration.

From these tests, we learn that, for accurate calibration of the internal parameters, we need first to perform the standard calibration procedure described in Section III-C1, and then, we need to perform two dynamic 3D tests, each with only one camera per time in rotation at a constant speed. From the linear fit of $\partial_t Z_{3D}(t)$ versus $\langle Z_{3D}^2 \rangle_t$, we estimate the error on the focal length of the two cameras, which we use to correct the results obtained with the standard calibration approach. With this two-step calibration procedure, we fulfill the requirement on the time independence of the reconstruction error at the relatively low cost of performing two dynamic 3D tests, namely, few hours of work.

2) Set-up and Alignment Consistency: Field experiments are often performed in locations where the apparatus cannot be mounted once and for all as it happens for our experiment, which is carried out on the roof of a building where we are forced to mount and unmount the entire system on a daily basis. It is then important to design an easy-to-mount system and a consistent calibration procedure. We tested CoMo to evaluate our consistency in the mounting procedure and in the alignment of the cameras with the fishing line, as described in Fig. 2.

To this aim, we performed two sets of static 3D test mounting and unmounting the entire system between the two. In each set, we repeat several times the alignment procedure taking at every alignment a static picture of the targets. We then reconstructed the position of the targets; we computed target-to-target distances and $\delta(\Delta R)/\Delta R$. Finally, we evaluate the variability of the reconstruction error within each set of data and between the two sets.

The results of this test are shown in Fig. 11(a), where we show the relative error, $\partial(\Delta R)/\Delta r$, of each pair of targets as a function of $\bar{Z}$. The plot shows variability within the same test, which is due to the alignment procedure, and variability within different tests, but with relative errors always below 0.012. The absence in both tests of a trend in $\bar{Z}$ shows that inaccuracies in the calibration of α are negligible and the alignment technique is consistent, while the upper limit of 0.0012 of the reconstruction error shows the consistency of our mounting procedure.

C. 3D Reconstruction Accuracy in Field Setup

We evaluate the 3D reconstruction accuracy of the system performing again 3D tests but this time with a setup as similar as possible to the experimental one. In principle, we should perform this 3D test exactly in the experimental configuration: camera baseline at 25 m, targets at a distance from the cameras in the range between 100 and 150 m, and pitch angles of both cameras set to 0.22 rad.

However, due to logistic constraints, we are forced to perform the tests in a slightly different configuration: 1) we set the camera baseline at about 10 m with targets at a distance from the cameras in the range between 20 and 40 m and 2) we do not manage to have targets in the common field of view of the cameras for a pitch value of 0.22 rad, but we can achieve the maximum pitch of 0.15 rad.

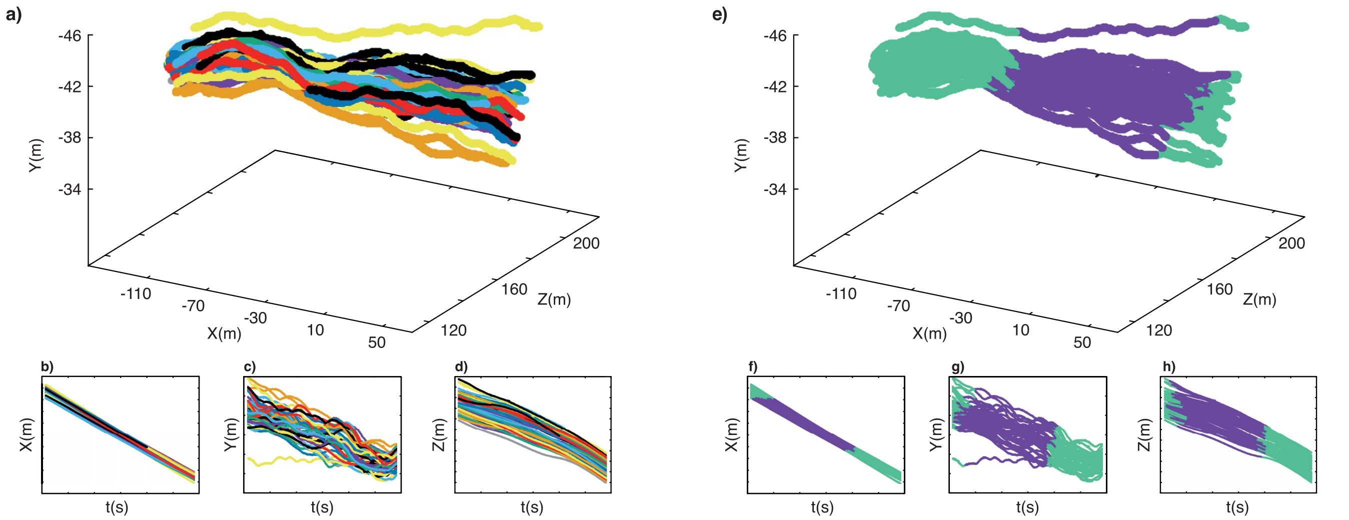


Fig. 12. 3D reconstructed trajectories of a starling flock. (a) 3D reconstructed trajectories of a flock of 50 birds, where each individual trajectory is highlighted in a different color. (b)–(d) X, Y, and Z coordinates of the trajectories as a function of time. (e) 3D reconstructed coordinates of the trajectories of the same flock represented in (a). The purple part of the trajectories represents the part of the acquisition that we would have acquired with our cameras in static mode, while the green part of the trajectories represents the extra data that we obtained with the dynamic system. (f)–(h) X, Y, and Z coordinates of the trajectories as a function of time, with the same color code of (e). In terms of the time length of the acquisition, we followed the flock for 11.2 s against 7.9 s that we would have taken with the static system, with a time increment of 3.3 s.

We take care of these two logistic limitations in the design of the test and the data analysis. In particular: 1) in the 3D test, the ratio Z/d is between 2 and 4, while this ratio in the field is between 4 and 6. The factor Z/d is relevant when we find a trend with $\bar{Z}$ in $\delta(\Delta R)/\Delta R$ in which case, to estimate the experimental error, we have to renormalize the 3D reconstruction error found in the tests by a factor 2. 2) we perform three series of tests at different pitch angles: $\beta = 0$ rad, $\beta = 0.08$ rad, and $\beta = 0.15$ rad to detect a potential trend of the error with β and, in this case, to predict the range of the reconstruction error in the field conditions, i.e., $\beta = 0.22$ rad.

We perform the 3D test in the following way: for each of the three pitch values, we perform first a static 3D test in the *home* configuration, and then, we put both cameras in rotation as in the field, namely both cameras rotate in the same direction and with the same rotational speed. We perform the test rotating the cameras at a constant speed of 6°/s (0.1 rad/s), which is the maximum speed that we use in the field.

The results are shown in the right box of Fig. 11, where, in the first row, we plot the relative error for the three static tests and, in the second row, the results of the dynamic 3D tests. In both cases, we have excellent results with relative errors smaller than 0.01, without any trend in Z, and we did not find any trend of the error with β. We do not have here to renormalize the relative error to take care of the different value of Z/d in the test and the field because we do not see any trend of the relative error in Z, nor we need to make any prediction of the error at $\beta = 0.22$ rad because there are nonappreciable differences of the errors for different βs.

The results of the static 3D test at $\beta = 0$ essentially reflect the accuracy of the cameras' alignment procedure. The comparison between static tests at different values of β shows that the introduction of a nonzero pitch angle produces a negligible error because, with different values of β, we obtain errors of the same order. With a similar argument, the comparison between static and dynamic tests shows that the introduction of the rotation due to the stages does not add affect the accuracy of the external parameters calibration. Therefore, the 3D tests show that the dominant source of error on the external parameters calibration is the alignment technique and, in particular, on the measurement of the cameras' yaw angles. From the results of the 3D tests, we estimated this angular error to be smaller than 0.001 rad, hence confirming the high precision of our alignment procedure.

VIII. Field Results

We tested the feasibility of the data collection with CoMo with the first experimental campaign on starling flocks, setting up the apparatus on the roof of Palazzo Massimo alle Terme in front of one of the bigger and more stable starlings roosting place in Rome.

With this first experimental campaign, we could prove the feasibility of the data-taking with CoMo. We proved that the system is easy to mount and easy to calibrate, and we also checked that the parameters we chose for the rotational controller are suitable to chase the flocks. We collected a quite large amount of data, ∼ 50 flocks of different sizes going from few individuals to large flocks of ∼ 1000 birds. With this first experimental campaign, we could also prove that, with the dynamic approach used by CoMo, we considerably expand the time length of the acquired data.

In Fig. 12(a), we show the 3D reconstructed trajectories of a flock of 50 birds, where each trajectory is highlighted with a different color, while, in Fig. 12(b)–(d), we show the X, Y, and Z coordinates as a function of t. In Fig. 12(e), we show the trajectories of the same flock but with a different color code to emphasize the temporal expansion that we obtain with the dynamic approach. In purple, we plot the part of the acquisition that we would have acquired with our cameras in static mode, while, in green, we plot the extra data that we were able to take due to the dynamic system. In terms of the time length of the acquisition, we followed the flock for 11.2 s against 7.9 s that we would have taken with the static system, with a time increment of 3.3 s.

IX. Conclusion

We presented a novel comoving camera stereo system, CoMo, developed in the context of 3D tracking of large groups of targets moving in a wide and nonconfined space. To overcome the limitation of standard static setup, where the size of the field of view is defined by the fixed position of the cameras and, in most of the cases, narrowed to achieve a sufficient resolution of the system, we designed CoMo to follow the motion of the targets with a controlled and synchronized rotation of the cameras driven by rotational stages (one for each camera).

The 3D reconstruction for a dynamic and wide-field system is rather demanding because the external parameters of the system have to be calibrated frame by frame, and they cannot be calibrated with standard methods, which are not accurate enough on wide-field data. We propose a novel technique for the calibration of the external parameters that separate their static component, corresponding to the system in the *home* configuration (rotational stages at the 0° position), from their dynamic component, corresponding to the rotation due to the stages. We calibrate the static component of the external parameters by measuring the position and the three angles of yaw, pitch, and roll of the cameras in a common reference frame, and we combine this information with the frame-by-frame rotation gathered from the stages.

We validated this calibration approach performing what we call *3D tests*: we set up the system, we acquire images of a set of still targets, and we accurately measure with a laser distometer the distance between each pair of the target. From the collected images, we reconstruct the position of the targets, and we compute their mutual distances that we compare with the measured ones. The results of the 3D tests show the consistency of the calibration method for the external parameters and the high accuracy of the system (3D reconstruction error below 1%).

The 3D tests represent a fair and objective method to evaluate the accuracy of a 3D system, but the very relevance of the 3D tests is in the designing phase of a 3D system because, as we showed in this article, 3D tests are a powerful tool to detect potential sources of errors, also providing a well-defined procedure to discriminate errors due to an incorrect measurement of the cameras' position versus errors due to an incorrect measurement of the cameras' orientation. Finally, 3D tests are on the basis of the new method that we proposed to improve the standard calibration of the focal length, which we could found to be inaccurate by performing dynamic 3D tests and noting an unexpected trend of the reconstructed position with time.

We carried the first experimental campaign using CoMo to collect data on starling flocks that are an emblematic example of targets moving in large groups in a nonconfined space. To this aim, we set up the apparatus on the roof of Palazzo Massimo alle Terme, where we are forced to mount and unmount the system every day. With this first experimental campaign, we proved that the system is easy to mount and easy to calibrate and confirmed that the design of CoMo considerably expands the time length of the acquired data. The simplicity of the system makes CoMo suitable for all those applications in the field of science, surveillance, entertainment, and robotics, where the experimental objective is to follow a target whose motion cannot be predicted in advance.

The limitations of the system in its current setup are essentially two. First, the manual control of the cameras' rotation may result in a suboptimal chasing of the target. The operator checks the flock position on the cameras by online watching the images on the laptop screen that has a low refresh rate, which may induce a delay reaction of the operator to sudden changes in the direction of the flock. Second, the rotation of both the cameras in the same direction may result in the target loss on one of the cameras when the flock moves in between the cameras along the direction orthogonal to the system baseline. In this special situation, we would need to rotate the two cameras in opposite directions, as it happens to human eyes when trying to follow the tip of a finger moving toward the nose. To overcome both these limitations, we are planning to upgrade CoMo to automatically follow the target estimating the optimal rotation needed to keep the target at the center of the field of view for each camera separately.

Acknowledgment

The authors thank Zachary Stamler for the fruitful and stimulating discussion about the calibration of the external parameters of the system.

References

[1] T. Bebie and H. Bieri, "A video-based 3D-reconstruction of soccer games," *Comput. Graph. Forum*, vol. 19, no. 3, pp. 391–400, Sep. 2000.

[2] O. Grau, G. A. Thomas, A. Hilton, J. Kilner, and J. Starck, "A robust free-viewpoint video system for sport scenes," in *Proc. 3DTV Conf.*, May 2007, pp. 1–4.

[3] A. Kulshreshth, J. J. LaViola, Jr., and J. Schild, "Evaluating user performance in 3D stereo and motion enabled video games," in *Proc. Int. Conf. Found. Digit. Games (FDG)*, 2012, pp. 33–40.

[4] S. Fleck, F. Busch, P. Biber, and W. Straber, "3D surveillance a distributed network of smart cameras for real-time tracking and its visualization in 3D," in *Proc. Conf. Comput. Vis. Pattern Recognit. Workshop (CVPRW)*, Jun. 2006, p. 118.

[5] S.-I. Yu, Y. Yang, X. Li, and A. G. Hauptmann, "Long-term identity-aware multi-person tracking for surveillance video summarization," 2016, *arXiv:1604.07468*. [Online]. Available: http://arxiv.org/abs/1604.07468

[6] P. Michel, J. Chestnutt, S. Kagami, K. Nishiwaki, J. Kuffner, and T. Kanade, "GPU-accelerated real-time 3D tracking for humanoid locomotion and stair climbing," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 2007, pp. 463–469.

[7] L. Wen, Z. Lei, M.-C. Chang, H. Qi, and S. Lyu, "Multi-camera multi-target tracking with space-time-view hyper-graph," *Int. J. Comput. Vis.*, vol. 122, no. 2, pp. 313–333, Apr. 2017.

[8] J. Haeling, M. Necker, and A. Schilling, "Dense urban scene reconstruction using stereo depth image triangulation," *Proc. SPIE*, vol. 11433, Jan. 2020, Jan. 114331R.

[9] D. Murray and J. J. Little, "Using real-time stereo vision for mobile robot navigation," *Auto. Robots*, vol. 8, no. 2, pp. 161–171, 2000.

[10] A. Broggi, C. Caraffi, R. I. Fedriga, and P. Grisleri, "Obstacle detection with stereo vision for off-road vehicle navigation," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Sep. 2005, p. 65.

[11] K. Konolige, M. Agrawal, R. C. Bolles, C. Cowan, M. Fischler, and B. Gerkey, "Outdoor mapping and navigation using stereo vision," in *Experimental Robotics*. Berlin, Germany: Springer, 2008, pp. 179–190.

[12] M. Bitzidou, D. Chrysostomou, and A. Gasteratos, "Multi-camera 3D object reconstruction for industrial automation," in *Advances in Production Management Systems. Competitive Manufacturing for Innovative Products and Services* (IFIP Advances in Information and Communication Technology), vol. 397. Sep. 2012, pp. 1–8.

[13] S. Ghosh and J. Biswas, "Joint perception and planning for efficient obstacle avoidance using stereo vision," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2017, pp. 1026–1031.

[14] R. Usamentiaga and D. F. Garcia, "Multi-camera calibration for accurate geometric measurements in industrial environments," *Measurement*, vol. 134, pp. 345–358, Feb. 2019.

[15] K. Schmid, T. Tomic, F. Ruess, H. Hirschmuller, and M. Suppa, "Stereo vision based indoor/outdoor navigation for flying robots," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Nov. 2013, pp. 3955–3962.

[16] F. M. M. Marreiros *et al.*, "Superficial vessel reconstruction with a multiview camera system," *J. Med. Imag.*, vol. 3, no. 1, Feb. 2016, Art. no. 019801.

[17] H. Fernandes, P. Costa, V. Filipe, L. Hadjileontiadis, and J. Barroso, "Stereo vision in blind navigation assistance," in *Proc. World Automat. Congr.*, Sep. 2010, pp. 1–6.

[18] C. Bert, K. G. Metheany, K. Doppke, and G. T. Y. Chen, "A phantom evaluation of a stereo-vision surface imaging system for radiotherapy patient setup," *Med. Phys.*, vol. 32, no. 9, pp. 2753–2762, Aug. 2005.

[19] T. Probst, K.-K. Maninis, A. Chhatkuli, M. Ourak, E. V. Poorten, and L. Van Gool, "Automatic tool landmark detection for stereo vision in robot-assisted retinal surgery," *IEEE Robot. Autom. Lett.*, vol. 3, no. 1, pp. 612–619, Jan. 2018.

[20] A. Attanasi *et al.*, "GReTA—A novel global and recursive tracking algorithm in three dimensions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 12, pp. 2451–2463, Dec. 2015.

[21] D. H. Theriault *et al.*, "A protocol and calibration method for accurate multi-camera field videography," *J. Experim. Biol.*, vol. 217, no. 11, pp. 1843–1848, Jun. 2014.

[22] I. Watts, M. Nagy, R. I. Holbrook, D. Biro, and T. B. de Perera, "Validating two-dimensional leadership models on three-dimensionally structured fish schools," *Roy. Soc. Open Sci.*, vol. 4, no. 1, Jan. 2017, Art. no. 160804.

[23] X. E. Cheng, Z.-M. Qian, S. H. Wang, N. Jiang, A. Guo, and Y. Q. Chen, "A novel method for tracking individuals of fruit fly swarms flying in a laboratory flight arena," *PLoS ONE*, vol. 10, no. 6, Jun. 2015, Art. no. e0129657.

[24] H. Zou, Z. Gong, S. Xie, and W. Ding, "A pan-tilt camera control system of UAV visual tracking based on biomimetic eye," in *Proc. IEEE Int. Conf. Robot. Biomimetics*, Dec. 2006, pp. 1477–1482.

[25] K. Fujimura, Y. Hyodo, and S. Kamijo, "Pedestrian tracking across panning camera network," in *Proc. 12th Int. IEEE Conf. Intell. Transp. Syst.*, Oct. 2009, pp. 1–6.

[26] H. Chen, X. Zhao, and M. Tan, "A novel pan-tilt camera control approach for visual tracking," in *Proc. 11th World Congr. Intell. Control Autom.*, Jun. 2014, pp. 2860–2865.

[27] K. Zhao, U. Iurgel, M. Meuter, and J. Pauli, "An automatic online camera calibration system for vehicular applications," in *Proc. 17th Int. IEEE Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2014, pp. 1490–1492.

[28] M. S. Al-Hadrusi, N. J. Sarhan, and S. G. Davani, "A clustering approach for controlling PTZ cameras in automated video surveillance," in *Proc. IEEE Int. Symp. Multimedia (ISM)*, Dec. 2016, pp. 333–336.

[29] D. D. Doyle, A. L. Jennings, and J. T. Black, "Optical flow background estimation for real-time pan/tilt camera object tracking," *Measurement*, vol. 48, pp. 195–207, Feb. 2014.

[30] R. Stolkin, A. Greig, and J. Gilby, "A calibration system for measuring 3D ground truth for validation and error analysis of robot vision algorithms," *Meas. Sci. Technol.*, vol. 17, no. 10, pp. 2721–2730, Aug. 2006.

[31] J. Salvi, X. Armangué, and J. Batlle, "A comparative review of camera calibrating methods with accuracy evaluation," *Pattern Recognit.*, vol. 35, no. 7, pp. 1617–1635, Jul. 2002.

[32] S. N. Fry, M. Bichsel, P. Müller, and D. Robert, "Tracking of flying insects using pan-tilt cameras," *J. Neurosci. Methods*, vol. 101, no. 1, pp. 59–67, Aug. 2000.

[33] Z. Liu, F. Li, X. Li, and G. Zhang, "A novel and accurate calibration method for cameras with large field of view using combined small targets," *Measurement*, vol. 64, pp. 1–16, Mar. 2015.

[34] F. Gu, H. Zhao, Y. Ma, P. Bu, and Z. Zhao, "Calibration of stereo rigs based on the backward projection process," *Meas. Sci. Technol.*, vol. 27, no. 8, Jul. 2016, Art. no. 085007.

[35] B. Shan, W. Yuan, and Z. Xue, "A calibration method for stereovision system based on solid circle target," *Measurement*, vol. 132, pp. 213–223, Jan. 2019.

[36] J. Chaochuan, Y. Ting, W. Chuanjiang, F. Binghui, and H. Fugui, "An extrinsic calibration method for multiple RGB-D cameras in a limited field of view," *Meas. Sci. Technol.*, vol. 31, no. 4, Jan. 2020, Art. no. 045901.

[37] J. Davis and X. Chen, "Calibrating pan-tilt cameras in wide-area surveillance networks," in *Proc. 9th IEEE Int. Conf. Comput. Vis.*, vol. 1, Oct. 2003, pp. 144–149.

[38] M. Machacek, M. Sauter, and T. Rösgen, "Two-step calibration of a stereo camera system for measurements in large volumes," *Meas. Sci. Technol.*, vol. 14, no. 9, pp. 1631–1639, Jul. 2003.

[39] Z. Song and R. Chung, "Use of LCD panel for calibrating structured-light-based range sensing system," *IEEE Trans. Instrum. Meas.*, vol. 57, no. 11, pp. 2623–2630, Nov. 2008.

[40] Z. Wu and R. J. Radke, "Using scene features to improve wide-area video surveillance," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2012, pp. 50–57.

[41] Z. Wang, Z. Wu, X. Zhen, R. Yang, J. Xi, and X. Chen, "A two-step calibration method of a large FOV binocular stereovision sensor for onsite measurement," *Measurement*, vol. 62, pp. 15–24, Feb. 2015.

[42] P. Cornic *et al.*, "Another look at volume self-calibration: Calibration and self-calibration within a pinhole model of scheimpflug cameras," *Meas. Sci. Technol.*, vol. 27, no. 9, Aug. 2016, Art. no. 094004.

[43] Y. Wang, X. Wang, Z. Wan, and J. Zhang, "A method for extrinsic parameter calibration of rotating binocular stereo vision using a single feature point," *Sensors*, vol. 18, no. 11, p. 3666, Oct. 2018.

[44] J. Zhang, H. Yu, H. Deng, Z. Chai, M. Ma, and X. Zhong, "A robust and rapid camera calibration method by one captured image," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 10, pp. 4112–4121, Oct. 2019.

[45] N. Machicoane, A. Aliseda, R. Volk, and M. Bourgoin, "A simplified and versatile calibration method for multi-camera optical systems in 3D particle imaging," *Rev. Sci. Instrum.*, vol. 90, no. 3, Mar. 2019, Art. no. 035112.

[46] R. Beschi, X. Feng, S. Melillo, L. Parisi, and L. Postiglione, "Stereo camera system calibration: The need of two sets of parameters," 2021, *arXiv:2101.05725*. [Online]. Available: http://arxiv.org/abs/2101.05725

[47] K. H. Strobl and G. Hirzinger, "More accurate pinhole camera calibration with imperfect planar target," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCV Workshops)*, Nov. 2011, pp. 1068–1075.

[48] A. Attanasi *et al.*, "Information transfer and behavioural inertia in starling flocks," *Nature Phys.*, vol. 10, no. 9, pp. 691–696, Sep. 2014.

[49] A. Cavagna *et al.*, "Short-range interactions versus long-range correlations in bird flocks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 92, no. 1, Jul. 2015, Art. no. 012705.

[50] A. Cavagna *et al.*, "The STARFLAG handbook on collective animal behaviour: 1. Empirical methods," *Animal Behaviour*, vol. 76, no. 1, pp. 217–236, Jul. 2008.

[51] A. Cavagna *et al.*, "Error control in the set-up of stereo camera systems for 3D animal tracking," *Eur. Phys. J. Special Topics*, vol. 224, nos. 17–18, pp. 3211–3232, Dec. 2015.

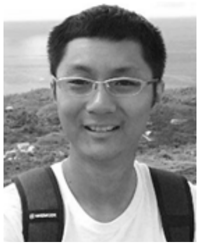
[52] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[53] W. Förstner and E. Gülch, "A fast operator for detection and precise location of distinct points, corners and centres of circular features," in *Proc. ISPRS Intercommission Conf. Process. Photogramm. Data*, Interlaken, Switzerland, 1987, pp. 281–305.

[54] W. Kabsch, "A solution for the best rotation to relate two sets of vectors," *Acta Crystallographica Sect. A*, vol. 32, no. 5, pp. 922–923, Sep. 1976.

Andrea Cavagna received the Ph.D. degree in theoretical physics from the Sapienza University of Rome, Rome, Italy, in 1998.

From 1999 to 2001, he was a Post-Doctoral Fellow with the University of Oxford, Oxford, U.K., and The University of Manchester, Manchester, U.K. He is currently a Researcher with the National Research Council (CNR)—Institute for Complex Systems (ISC), Rome, where, together with Irene Giardina, he leads the Collective Behaviour in Biological Systems Group (CoBBS Group). His research interests include statistical physics, collective animal behavior, optimization, and computer vision.



Xiao Feng is currently a post-doctoral researcher in mechanical engineering. His research is focused on the nonlinear dynamics with control. His theoretical work is mostly related to a motion control system, which is designed for moving the cameras to follow the fast-moving flocks precisely. As a member of the experimental team, he has conducted experiments and collected a lot of data about starling flocks in the past few months.

Stefania Melillo received the Ph.D. degree in mathematics from the Sapienza University of Rome, Rome, Italy, in 2010.

She is currently a Researcher with the Collective Behaviour in Biological Systems Group (CoBBS Group), National Research Council (CNR)—Institute for Complex Systems (ISC), Rome, where she leads the experimental and tracking activities. Her scientific interests span from the dynamics of biological systems—experiments, 3D tracking, processing, and analysis of experimental data—to different fields of applied mathematics.

Leonardo Parisi is currently a computer scientist. His work is mainly focused on computer vision, 3D tracking, and GPU programming. He joined the Collective Behaviour in Biological Systems Group (CoBBS Group), Institute for Complex Systems (ISC), Rome, Italy, where he is a Researcher. He also works on the reconstruction of the 3D trajectories of the space debris, in collaboration with the 5SLab, Department of Mechanical and Aerospace Engineering (DIMA), Sapienza University of Rome, Rome.

Lorena Postiglione is currently a post-doctoral researcher in biomedical engineering. In April 2019, she joined the Experimental Team, CoBBS Lab, Sapienza University of Rome, Rome, Italy. She is involved in conducting experimental campaigns on flocks (starlings) and cell colonies (mesenchymal human stem cells). She also works on the development of segmentation and tracking algorithms for the lineaging of proliferating stem cells.

Pablo Villegas is a researcher focused on exploring the idea of criticality in living matter and its consequences for the emergence of collective phases and cooperative phenomena. In December 2019, he was part of the Experimental Team, CoBBS Lab, Sapienza University of Rome, Rome, Italy, taking and processing all the images from the field.