



Machine Learning for 3D Particle Tracking in Granular Gases

Dmitry Puzyrev¹ · Kirsten Harth¹ · Torsten Trittel¹ · Ralf Stannarius¹

Received: 25 March 2020 / Accepted: 3 May 2020 / Published online: 18 July 2020
© The Author(s) 2020

Abstract

Dilute ensembles of granular matter (so-called granular gases) are nonlinear systems which exhibit fascinating dynamical behavior far from equilibrium, including non-Gaussian distributions of velocities and rotational velocities, clustering, and violation of energy equipartition. In order to understand their dynamic properties, microgravity experiments were performed in suborbital flights and drop tower experiments. Up to now, the experimental images were evaluated mostly manually. Here, we introduce an approach for automatic 3D tracking of positions and orientations of rod-like particles in a dilute ensemble, based on two-view video data analysis. A two-dimensional (2D) localization of particles is performed using a Mask R-CNN neural network trained on a custom data set. The problem of 3D matching of the particles is solved by minimization of the total reprojection error, and finally, particle trajectories are tracked so that ensemble statistics are extracted. Depending on the required accuracy, the software can work fully self-sustainingly or serve as a base for subsequent manual corrections. The approach can be extended to other 3D and 2D particle tracking problems.

Keywords Machine learning · Granular gas · Particle tracking · Object detection · Mask-CNN

Introduction

Granular gases are dilute ensembles of macroscopic grains, which in the simplest case interact only upon contact during collisions, without any long-range interactions. Studies of granular gases are relevant, for example, for gaining deeper insights into fundamental physics of non-equilibrium systems (Pöschel and Luding 2001; Pöschel and Brilliantov 2003, 2004), as a basis for modeling collisional dynamics in planetary rings or other astrophysical assemblies of solid objects, even in some stages of planet formation (Hestroffer et al. 2019). A quantitative macroscopic description of their ensemble properties will also aid simulations of fluidized granular materials.

In contrast to molecular gases, all collisions among particles or between particles and the container walls are dissipative, i.e. part of the kinetic energy is lost. Consequently, without external energy supply, the ensemble will gradually lose its kinetic energy in a process called granular cooling. This cooling process itself is non-trivial, and most of its properties are insufficiently proven experimentally. It manifests, e.g., in the formation and kinetics of particle clusters (Hopkins and Louge 1991; Goldhirsch and Zanetti 1993), and in the evolution of velocity and energy distributions of the ensemble (Maaß et al. 2008; Brilliantov et al. 2018). Steady energy supply is required to maintain a granular gas in a steady state. Even then, the dissipative character of the collisions causes numerous distinct properties, such as non-Gaussian velocity distributions in the individual degrees of freedom (Yanpei et al. 2011; Harth et al. 2013a, b), or non-equipartition of kinetic energies between different degrees of freedom (Harth et al. 2013a, b). The literature contains numerous analytical and numerical predictions of structural and dynamical properties of granular gases, relying on more or less realistic assumptions of the particle interactions. However, experimental data, in particular regarding 3D granular gases, are still scarce.

Experiments with two-dimensional systems usually suffer from the influence of container walls (van Zon et al.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s12217-020-09800-4>) contains supplementary material, which is available to authorized users.

This article belongs to the Topical Collection: The Effect of Gravity on Physical and Biological Phenomena
Guest Editor: Valentina Shevtsova

✉ Dmitry Puzyrev
dmitry.puzyrev@ovgu.de

¹ Institute of Physics, Otto von Guericke University, Universitätsplatz 2, D-39106 Magdeburg, Germany

2004) and of the energy input mechanisms. Levitation by external electro-magnetic fields (Aranson and Olafsen 2002; Maaß et al. 2008), gas layers (Nichol and Daniels 2012; Burton et al. 2013; Daniels et al. 2009) or reduction of container wall friction in micro-gravity (Hou et al. 2008; Yanpei et al. 2011; Grasselli et al. 2015) represent an alternative. Even then, the out-of-plane contributions to the collision dynamics, e. g. mediated by particle rotations, are artificially constrained. Many of these experiments were restricted to small particle numbers. Studies of granular cooling are hardly possible in a 2D geometry: An early experiment (Maaß et al. 2008) dealt with approximately 30 magnetically levitated grains. The decay of the mean absolute velocity followed Haff's prediction (Haff 1983). For more realistic experiments, it is desirable to study particle dynamics in three-dimensional (3D) granular gases in detail, with sufficiently large ensembles to obtain reliable statistics. Both free cooling and externally excited states are of interest. Complete 3D information on the particle trajectories is desired. Realizations of 3D granular gases require microgravity. One needs a sufficient understanding of the particle properties and the interactions of grains with container walls.

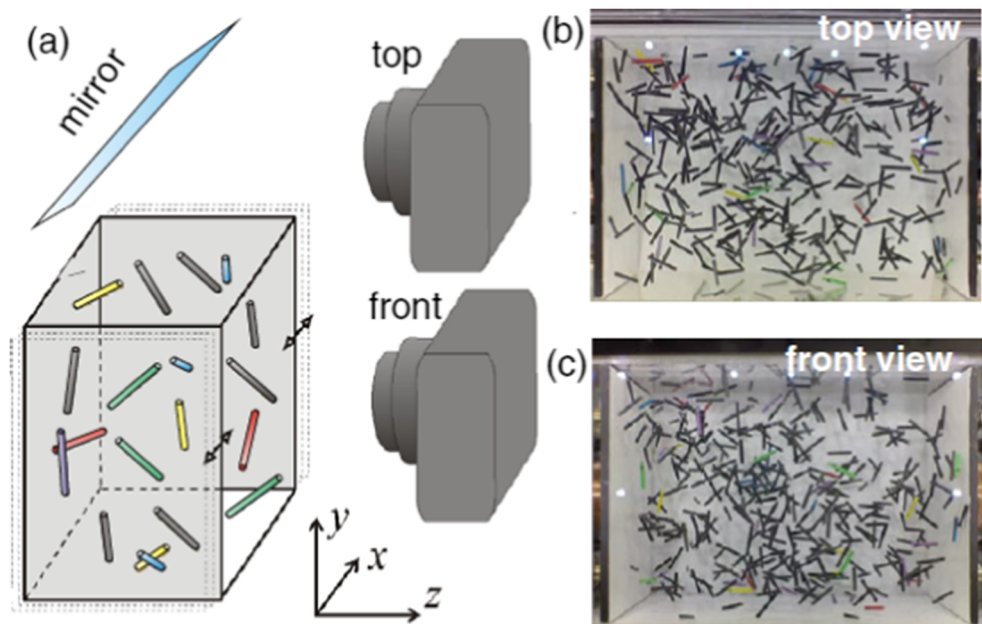
First microgravity experiments were performed by Falcon et al. (1999, 2006) using spherical grains in strongly vibrated containers. In these experiments, dynamic clustering (Falcon et al. 1999) due to strong external excitation was identified, a phenomenon studied in more detail in subsequent works, e.g. Refs. Opsomer et al. (2011), Sack et al. (2013), Noirhomme et al. (2018), and Aumaitre et al. (2018). In the regime where particle-particle collisions dominate over particle-wall collisions (beyond the Knudsen

regime), i. e. at sufficiently high particle number densities, the projected images of spherical grains heavily overlap. Consequently, tracking the particles is challenging. This problem cannot be resolved with improved optical recording techniques.

We study ensembles of rodlike grains with lengths of about one centimeter in microgravity (Harth et al. 2011, 2013a, b, 2015, 2017, 2018). Regarding particle detection and tracking, this is advantageous as the mean free path scales primarily with the *length* of the rods (Harth et al. 2013a, 2015, 2018), while the optical mean free path (i. e. the average depth a light ray penetrates the ensemble before being blocked by a particle) can be optimized by using small rod *diameters*. Thus, sufficiently frequent particle-particle collisions can be achieved at lower filling fractions than for spherical grains. This provides a better in-depth visibility of the rods for their identification and tracking. The sketch of the experimental geometry as well as two typical snapshots of the granular ensemble are shown in Fig. 1. In our previous study (Harth et al. 2018), Haff's cooling law and the equipartition of kinetic energy among the individual degrees of freedom were checked in a 3D ensemble.

Detailed studies of such statistical ensemble properties require not only repeatable microgravity experiments but also a reliable and accurate particle tracking in 3D. The latter was previously done completely, or at least to a large extent, manually. Although in the case of thin rodlike particles, the visual tracking is possible, automatic tracking is still complicated because of multiple technical difficulties. The main problems are frequent particle overlaps in the images as well as differing appearance,

Fig. 1 **a** Sketch of the experimental setup and definition of the coordinates. Two side walls can be vibrated mechanically; the top and front walls are transparent. **b, c** Typical frames of the top and front videos. Image reproduced from Ref. (Harth et al. 2018), copyright by the American Physical Society



e.g., of particles of identical color in different positions and orientations in the same video frame. The biggest of those challenges is the accurate detection of positions and orientations of particles given the numerous cases of overlapping in each camera perspective (Fig. 1). In particular, due to the overlaps, the particles frequently appear as several segments. Due to frequent collisions, predictions of the particle positions and orientations in subsequent images, as they are often used in particle tracking velocimetry at very low particle concentrations, is of limited applicability. Another problem is the correct assignment of the particles in the two different viewpoints, in order to extract their 3D coordinates. This problem is partially relieved by choosing subsets of particles with different colors. Finally, 3D tracking of particles has to be performed on the basis of a correct assignment of the particles in subsequent frames of the video sequences. We describe a machine learning (ML)-aided approach to detect the rod positions and orientations, which allows a fast acquisition of results with an accuracy comparable to manual visual analysis performed by human evaluators.

Detection in 2D

General Approach

The general approach which we applied so far for the extraction of statistical properties from experimental video footage mainly relied on manual tracking, using custom MATLAB code. The first step consisted of the detection and tracking of rods in (one or two) 2D perspectives (Harth et al. 2013a, b, 2015). This task was initially performed by human operators. First attempts to automatize the detection procedure involved simple Hough transforms (Hough 1962). For single jumping rods (Trittel et al. 2017), automatic reconstruction of 3D positions and orientation was successful, yet for multiparticle ensembles, this method proved to be insufficient by far. A few 3D trajectories could be reconstructed (Harth et al. 2013b) with massive efforts.

In the subsequent experiments, the setup and choice of particles were substantially adjusted to improve the ability to detect and distinguish the representative rods in 3D ensembles. The semi-automatic tracking approach used in Refs. Harth et al. (2017, 2018) was based on object detection/connection of partial objects in color segmented individual video frames, in combination with subsequent manual correction/addition of a large sub-set of identified rods Harth et al. (2017, 2018). It led to some improvement but was overall still unsatisfactory in terms of tracking efficiency. In that algorithm, the automatically detected rod projections required substantial manual correction, which caused a still low efficiency of data evaluation. Particle

trajectories were composed by conventional minimization of distances. The number of tracked particles was small.

In a second step, the projected images of the rods were matched. These steps were performed by trained individuals. In a third step, 3D positions of the rods' endpoints were calculated using the camera calibration toolbox for MATLAB, combined with proper calibration data. In a last step, the quality of reconstructed rods was assessed by comparing the detected lengths with the actual lengths of the particles. For purely manual tracking, the data were substantially more accurate than for the automatic detection where a considerable percentage of data, up to 75%, had to be discarded or manually corrected. Purely manual tracking was performed by initially choosing one of the rods, and subsequent viewing of sequences of video frames, where the evaluator had to manually click the endpoints of the rods. Typically, the standard rate at which these activities can be performed is around 2-3 full frames per hour for experienced individuals. This is not only a tedious task, but also leads to a significant demand of human resources.

We have thereafter performed extensive research into the applicability of traditional computer vision (CV) methods for the 3D setup. Various methods of color separation (by thresholds in color spaces or by clustering of colors) were followed by application of probabilistic Hough transform (Kiryati et al. 1991) and subsequent clustering of line segments. Overall, the results were proven to be inconsistent. First of all, the Hough transform contains a number of parameters (i. e. minimum and maximum length of line segment, size of possible obscured part, statistical threshold, etc.), which can be adjusted to specific configurations of rods on one frame, but are unstable on a different frame and even more in a different experiment. There are significant issues with color separation which works with highly varying efficiency for different rod colors in different color spaces.

The biggest challenge is clustering the line segments detected by Hough transform into “real” elongated structures. We have applied various metrics for clustering of lines based on their mutual position and orientation (see Ref. Wirtz and Paulus (2016) for an overview of segment distance functions). In all cases, significant problems of separation of line segments belonging to different rods of the same color occur if they are nearby or overlapping. The inefficient clustering of lines leads to the fact that the number of reconstructed elongated objects usually varies between 70% and 130% of the real number of particles. This leads to severe complications in the solution of the optimization problem of 3D line segment matching (see “[Matching and Tracking in 3D](#)”).

We arrived at the conclusion that a more accurate and adaptive object recognition method is required, which is

capable to parse the visual information in the sense of detection of particular objects (rods of different colors), unlike the traditional CV algorithms which analyse the color/geometrical information in the picture without a notion of which objects of interest are being tracked. Building a somewhat intelligent recognition system is possible with the use of modern techniques of artificial intelligence/machine learning (Russell and Norvig 2009; Bishop 2006). Namely, solutions based on machine learning usually require large amounts data for training, but provide great advantages in efficiency in many tasks such as classification and regression for various types of data. One of the crucial challenges in the application of machine learning is the choice of the fitting approach (e.g. supervised, unsupervised or reinforcement learning) and particular configuration of the system (most often the architecture of an artificial neural network).

In recent years, the application of Convolutional Neural Networks (CNNs) has led to significant breakthroughs in the solution of computer vision problems. CNNs are particularly effective in tasks which require the detection of objects of varying shapes under complex background and lighting conditions, including visual noise. Our task requires high precision recognition along with the detection of particular object shapes. This can be achieved with Mask R-CNN (He et al. 2017). The main feature of Mask R-CNN is the combination of layers of the network which detect the raw position of the object (bounding box) with additional layers providing the exact mask (region shape) for each detected object. This is instrumental for our task, since we need to detect the coordinates of rod endpoints. Moreover, in comparison to other CNNs, a sufficient increase in the accuracy of detection of small objects (which is particularly important for elongated rod-like particles) is observed. The trade-off is a relative slowness of Mask R-CNN: on modern GPUs, detection can take several seconds for one frame, depending on image resolution and network parameters. Thus, Mask R-CNN is not suited for real time applications (e.g., self-driving vehicles) with contemporary hardware. However, our particular problem does not require real-time data analysis.

After making the choice of the network architecture, two essential steps in building the ML-aided detection and tracking system follow: assembling a sufficient amount of annotated data and training of the network.

Assembling the Data Set

From the previous data acquisition by manual detection, around 600 images with particle data (endpoint coordinates) were already available. These data correspond to experiments with colored rods (6 colors, 12 rods of each color) which required tracking. Detectable rods are mixed with

several hundreds of background particles (gray metallic or black, see Fig. 1). A substantially smaller amount of data (50 frames) is available for experiments in a dilute regime (Knudsen gas). Here, all particles (6 colors, 20 rods of each color) need to be tracked to analyze the collision dynamics (see Fig. 4).

Mask R-CNN is a network which performs semantic segmentation of the image, i.e. for training it requires information about the particular shapes of the regions which the detected objects occupy in the image of interest (masks). Generally, a single “isolated” rod (a rod that does not overlap with another one) in the image occupies an approximately rectangular region (neglecting the curvature near the ends of the cylinders). This rectangular region can be rotated in arbitrary direction and have different length to width ratios. In our setup, the width of those rectangles does not depend substantially on the depth, i. e. the distance between the rod and the camera.

In case of overlap, the rod on the rear is seen with a more complicated shape, consisting of several not necessarily rectangular segments, see overlapping red and yellow rods in Fig. 2. One question that is not sufficiently discussed in the literature and the Mask R-CNN manuals is whether one should include the obscured part of the object into the corresponding object region if large numbers of partly overlapping objects must be detected. In other words, should one “see through” the closer rods to recognize the shapes of the rods in the back?

We found that, since we are interested in the rods’ endpoint coordinates, including overlapping parts of the rods into each rod’s region is the better choice in our experiment. It simplifies the creation of the data set dramatically, since one does not need to mark up all parts of obscured rods and label them as individual objects. Consequently, partially obscured rods are also represented as a single object in the detection stage. One can suggest that Mask R-CNN learns that the objects of interest look like the rectangles of certain width and color possibly obscured by other rectangles. Thus, the network generally reconstructs the set of the adjacent similarly-colored regions with certain shapes, mutual positions and orientations as a *single* elongated object, thus providing correct endpoint coordinates.

Finally, we use the fact that for the efficient training of Mask R-CNN, the shapes of the objects in the annotated data set do not require to be absolutely precise, allowing the mask of an object to include some adjacent areas of object background (He et al. 2017; Abdulla 2017). This gives us the opportunity to reconstruct the approximate regions for each rod from its endpoint coordinates into rectangles of fixed width. Thus, we were able to create the data set for Mask R-CNN automatically from available manual detection data. We have followed the instruction for creation

of a custom data set in Ref. Abdulla (2017) and constructed the first set using the VGG Image Annotator (VIA) (Dutta and Zisserman 2019). Each rod color was assigned a label (region property in VIA). Figure 2 shows an example of a zoomed part of an image with annotated regions for the colored rods.

The second essential issue is data set merging. Suppose one has data for different experiments (in our case this means that the geometry of the system or the lighting conditions may differ, or the box can be filled with a combination of rods of different colors and lengths). Suppose, for example, we have data for an experiment with a larger amount of background rods, which provide a higher filling fraction, as well as experiments with substantially fewer rods. Should the detection system then be trained on the combined data, or should one train several versions of detection on more specialized data sets? The answer to this question depends on the particular properties of the data set, viz. the amount of data available and its balance with respect to the different experiments.

For our experiments, the use of the combined data of all experiments increases the accuracy of the detection system, both formally by comparing the Mask R-CNN losses and in the more meaningful metric of 3D reconstruction accuracy (see “Matching and Tracking in 3D” for details). This boost in accuracy is evident for each of the tasks, even if the data set is unbalanced with respect to those tasks.

After training on the initial data set, we employed an iterative procedure for the improvement of the data set. Namely, the shapes of rods are detected in some relatively

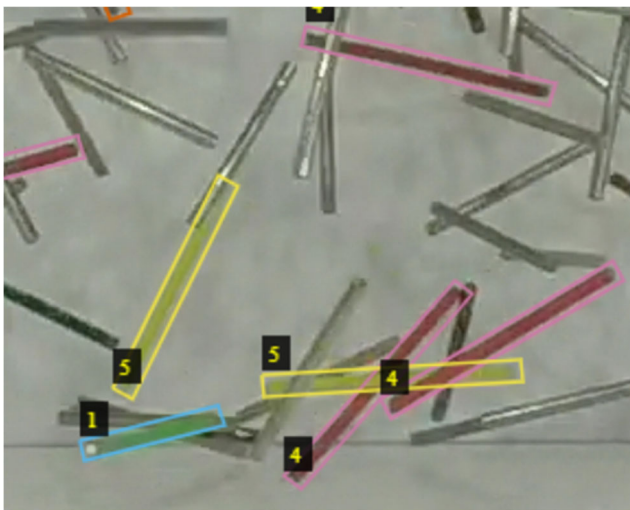


Fig. 2 Snapshot of annotated data set in VIA (zoomed in). One can see partially overlapping rods of different colors together with background gray metallic rods (not for detection). The network requires the mask shape (region) and class for each object of interest. Each rod color corresponds to a specific object class, denoted by region outline color and class identifier value (in black squares). If two or more rods overlap, the overlapping parts are included in each rod region

small number of images, and then they are manually corrected. These corrected data are then used to extend the original data set. Our current data set consists of around 1500 train and 150 validation images. Each image contains between 72 and 120 detected rods. It combines data for several experiments, where a total of eight different rod colors were used.

We are currently performing work in the direction of balancing and improving the accuracy of the data set and plan to publish the data set for open access. Images with corresponding masks and 2D endpoint data can be used for training a similar detection network, which can be adjusted to a modified experiment. Together with matched 3D endpoint coordinates, it will provide a useful database for statistical models of granular gases as well as a tool to calibrate numerical simulations. Additional investigations into the possibility of augmenting the data set with synthetic data (computer-generated images) are planned.

Training

In the current version of the software, the Mask R-CNN implementation of Matterport (Abdulla 2017) is used as a starting point. Migration to DETECTRON 2 is envisioned in the near future, see “Conclusions”.

We followed the approach described in Ref. Abdulla (2017) for training of Mask R-CNN on a custom data set. Since we do not possess a really large amount of data for training, we have used the technique of transfer learning (Thrun and Pratt 2012). Namely, we have initialized our network with the weights from the network trained on the large MS-COCO dataset, which contains tens of thousands of annotated images with millions of object instances (Lin et al. 2014). The available data are separated into train and validation partitions. The standard Mask R-CNN loss is used for optimization. Since ultimately we are interested in a custom 3D accuracy metric (see “Matching and Tracking in 3D”) and take the built-in loss metric of MASK R-CNN as an approximation, we use 10% of data for validation instead of the standard 20 %. We also have a sufficient amount of manually marked data in a separate test partition to estimate the accuracy of 3D matching. During training, we use the `imgaug` library (Jung et al. 2020), and apply the data augmentation in the form of translations, flips and rotations of images, occasionally with a small amount of sharpening or Gaussian blur.

After numerous experiments and adjustments, the training protocol reads as follows: we first train the network for 200 epochs with only head layers unfreezed with standard learning rate ($r_l = 0.001$). The aim is to reach the first plateau in Mask R-CNN losses. After seeing that the training and validation losses do not decrease for several epochs, we train the network with all trainable layers for 100 epochs

with the same learning rate. After that, we train the network with smaller learning rates (50 epochs with $r_l = 0.0004$ and 50 epochs with $r_l = 0.0001$).

Note that we have not encountered substantial overfitting problems during the training, with a training error being only around 10% smaller than the validation error in the worst case.

Matching and Tracking in 3D

Studying the dynamics of 3D granular gases naturally requires the 3D information on particle positions and velocities. After the 2D detection step is performed, the next task is to transform the two-view 2D endpoint coordinates into 3D coordinates. This is nontrivial since the correspondence between the detected line segments in both views is initially unclear.

We perform the matching of line segments in the following way: we use the MATLAB `triangulate` function which outputs the 3D coordinates from a pair of 2D coordinates and (experiment-specific) stereo camera calibration parameters. The output 3D coordinates minimize the sum of reprojection errors (difference between the 3D points projected back to the 2D camera coordinate systems and their input 2D coordinates). One can perform this procedure for any pair of coordinates in the 2D projections, albeit for non-matching points, the resulting reprojection error is exceptionally large. Thus, given two sets of endpoint coordinates, one can choose the combination of endpoints that minimizes the total reprojection error. This requires solving the optimal assignment problem (Burkard et al.

2009) for the number of rods (taking into account the rods' endpoint detection errors as well). For the current number of rods, any polynomial-time optimal assignment solver gives sufficiently fast and accurate results. We found that the rod matching procedure is indeed stable (we have not observed any discrepancy with our 'manual' matching procedure where each assignment is made by a human operator).

Tracking in 3D is subsequently performed with the `trackpy` package (Allan et al. 2019) for Python, which implements the well known Crocker-Grier (Crocker and Grier 1996) particle tracking algorithm.

After the tracking of particles, some additional post-processing steps to further reduce the effects of detection noise can be performed, before finally extracting the ensembles' statistical properties, such as the mean energy for each degree of freedom, spatial particle distributions, etc. For example, for the dilute case where all the rods are detected and tracked, one can find all the collisions between particles and particles and the walls. Between the collisions, the particles' center of mass moves with constant velocity and rotational velocity, and the corresponding segments of the trajectories can be fitted by an affine function. This immensely increases the accuracy of the detected velocities and rotational velocities, and thus also corresponding data in a statistical evaluation. In case the ensemble is far from the dilute regime, the collision detection is less straightforward. At the current stage, the task of collision detection, based on the displacement of the centers of mass and on the evolution of the particles' long axis orientations respective to the coordinate axes, is performed manually. However, this last manual post-processing step of evaluation is comparably fast. The implementation of collision detection is one of

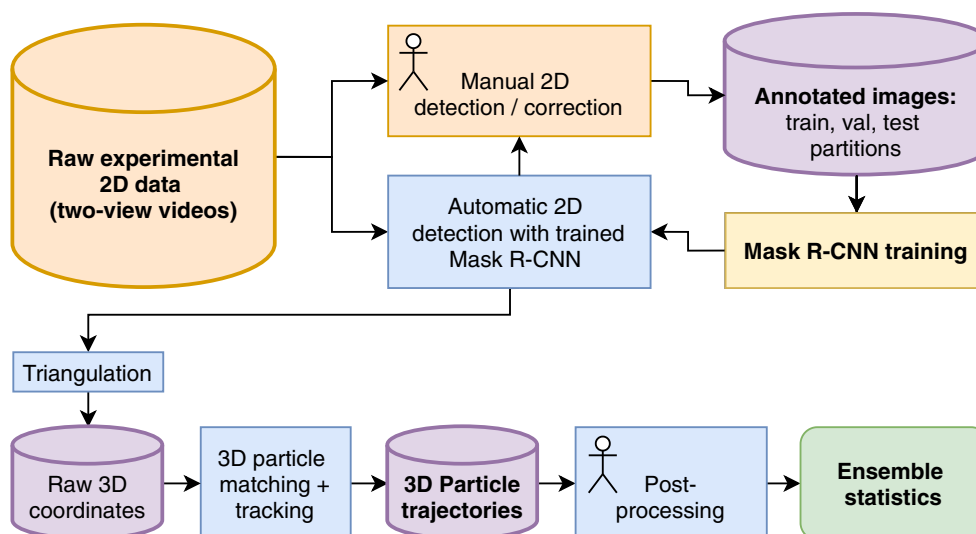
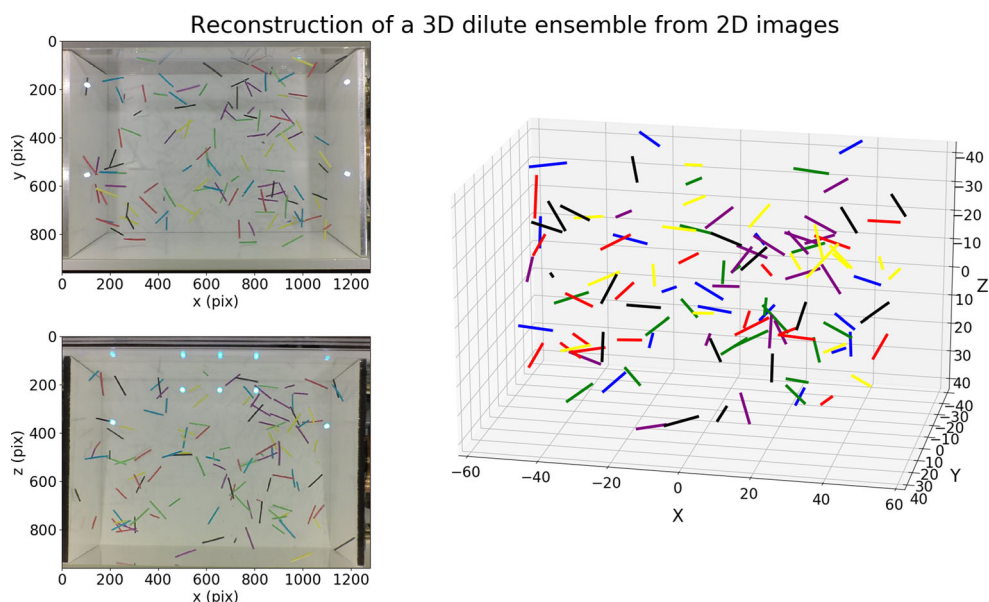


Fig. 3 Flowchart illustrating the steps performed by the detection system to extract 3D particle trajectories and ensemble statistics from the experiment. Currently, manual work of an operator (denoted by

pictograms) is required for data set construction and expansion, post-processing of trajectories and initialization of individual workflow steps

Fig. 4 Automatic reconstruction of a 3D scene for a dilute ensemble of rods based on two perpendicular camera views. The animated video of the same 3D scene from an orbiting viewpoint is included in [Supplementary Material](#)



the proposed steps for an improvement of the automated analysis.

Figure 3 shows the flowchart which summarizes the aforementioned steps in data analysis which lead from raw experimental video data to extraction of 3D particle trajectories and finally obtaining ensemble statistical properties.

Results

In the following, we demonstrate some results of the current version of our detection and tracking algorithm. Figure 4 shows frames from two cameras along with the corresponding automatically reconstructed 3D scene of the dilute ensemble of rods. The animated video of a 3D scene from an orbiting viewpoint is provided as [Supplementary Material](#). From the inspection of the 3D scenes, one can immediately assess the overall quality of the 3D reconstructions and identify obvious artifacts. One of the issues is that in our experiments described above, one does not have direct access to the *ground truth* information (i.e. to the exact positions and velocities of particles). Therefore, one has to devise a suitable measure to assess the detection accuracy both for the manual and automatic methods. We chose the reprojection error from 3D to 2D as the measure of accuracy: We assume that if the triangulated rod endpoint coordinate has small reprojection error, it must correspond to the “real” position of the rod endpoint in 3D. Naturally, it can happen that both frames taken from different camera positions have errors in the endpoint direction which assemble in a wrong 3D coordinate with a small reprojection error. However, those events happen to occur sufficiently rarely.

Figure 5 shows the distributions of reprojection errors for automatic and manual recognition. One can see, that the majority of rods (80-90%) is recognized with similar accuracy. There is a longer tail in the error distribution in the automatic system. One can propose the following ways to deal with this effect: a) First of all, based on reprojection error, we can predict which rod coordinates are extracted accurately. This part of the data can already be substantial and sufficient to extract ensemble statistical measures like the mean kinetic energy. However, we should be aware of the possibility of some systematic error which appears with neglecting some part of data. We are working on the

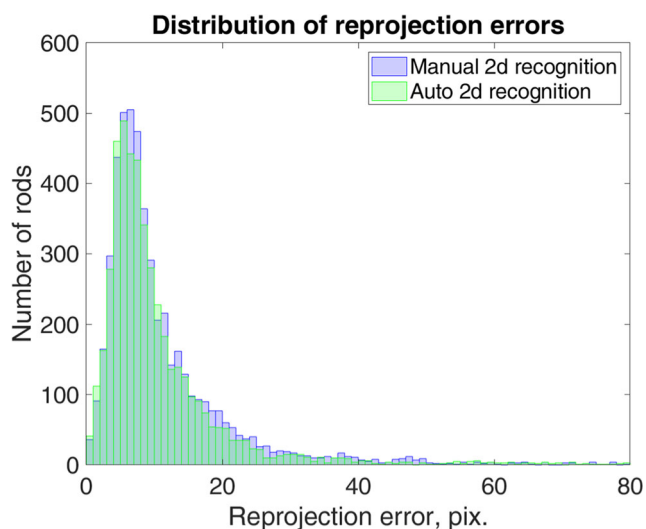


Fig. 5 Distribution of reprojection errors for manual and automatic particle recognition. A similar accuracy can be observed for the majority of the rods. The long tail of the distribution corresponding to large detection errors for few rods is a focus of ongoing optimization

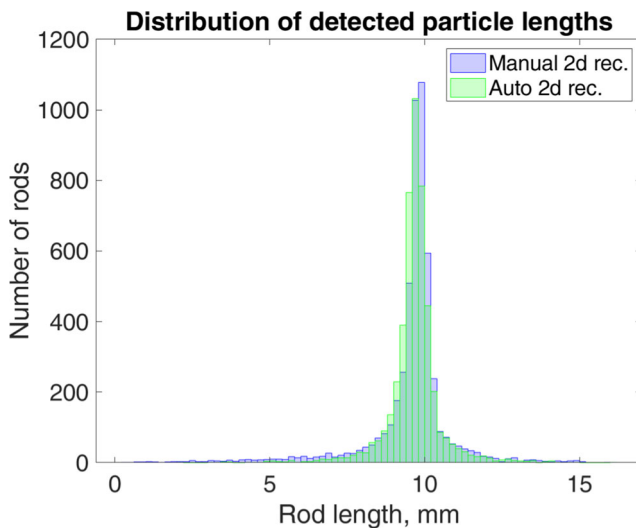


Fig. 6 Distribution of reprojection errors for manual and automatic particle recognition. A similar accuracy can be observed for the majority of the rods. The long tail of the distribution corresponding to large automatic detection errors for certain rods is a focus of ongoing optimization

investigation of this problem. b) The rods which belong to the long tail with high reprojection error can be marked for manual correction. This allows to obtain the most accurate results available at the moment, still greatly improving the overall analysis efficiency.

Another measure, which can be applied to assess the accuracy of the reconstruction, is to compare how well the detected particle lengths compare with the real rod lengths in the experiment (10 mm in the presented data set). Figure 6 shows the distribution of detected rod lengths for manual and automatic detection. One can observe comparable accuracy, however, automatic detection tends to slightly underestimate the lengths of the particles. We are investigating the source of this discrepancy and possible improvement of results.

Other measures can be also considered to assess accuracy quantitatively, for example, one can check how well the rods maintain constant translational and rotational velocities in the periods between the collisions with other rods and walls. We plan to include more accuracy measures in the next versions of the detection software.

Conclusions

We have investigated the potential of a ML-aided analysis of structural and dynamical properties of highly mobile multiple object ensembles. Namely, the software for detection and tracking of rod-like particles constituting a 3D granular gas in microgravity was developed. We have described here the data set creation and expansion procedure as well as the

training protocol for the Mask R-CNN network. The recognition accuracy is comparable to manual detection, while the efficiency of the analysis has been improved dramatically. These features are particularly desirable when one needs a large throughput of experimental data to obtain a reasonable basis for the statistical ensemble characterization. While few random errors in the raw output data of the automated approach will hardly influence the statistical results, it is important that the evaluation is not affected by bias (for example, larger reprojection errors in regions with higher particle number density). In that respect, the automatic detection is practically coequal to the manual approach.

In recent years, many improvements of machine learning methods in visual analysis of experimental data have been demonstrated. Most of those investigations focus on the improvement of recognition results on several generic data sets. However, there is a deficit of well documented cases in successful applications of such methods to study real physical model systems. In the present work, we aimed at the demonstration and comprehensive description of the efficient ML-aided statistical analysis, applied to a cutting edge microgravity experiment. The 3D reconstruction of trajectories, and particularly the identification and quantitative evaluation of collision events will allow the experimenters to explore the relations between microscopic individual particle properties and particle-particle interaction events and the macroscopic statistical properties of the ensembles such as loss of kinetic energy by dissipation, the evolution of spatial inhomogeneities and the formation of structures and collective dynamic patterns.

The present software can be adjusted to various applications in the statistical analysis of multi-particle ensembles in 2D and 3D, especially when particles have complex shapes and one is interested in their spatial orientations. One can envision an application to experimental studies of clogging problems (Zuriguel et al. 2015; Ashour et al. 2017; Shi et al. 2019), the collective behaviour of ellipsoidal colloids (Zheng et al. 2011; Crassous et al. 2012; Weeks 2011), granular matter (Aranson et al. 2007; Kudrolli et al. 2008) including active granular matter (Kumar et al. 2014), or bacteria colonies (Copeland and Weibel 2008; Xu et al. 2019) and other microswimmer organisms (Koch and Subramanian 2011). The investigation of multi-particle ensembles under microgravity conditions might be a particularly fruitful area of interest due to a high mobility of particles in three dimensions in the absence of gravity. We plan to publish the updated version of the detection software as well as the data set for open access.

There are multiple potential directions in sight for the improvement of the detection software. Significant

advances in machine-learning aided computer vision which can benefit our research are currently taking place on a regular basis. Namely, we plan to use the newly published DETECTRON 2 detection platform (Girshick et al. 2018; Wu et al. 2019) which includes advanced versions of the Mask R-CNN network. A promising advantage for our application would be the inclusion of rotating bounding boxes, as well as the use of object keypoints along with the masks, which are both among the newly supported features of DETECTRON 2.

The application of ML techniques in microgravity research is not restricted to the detection of positions and orientations of floating grains, it may be adapted to the observation of other floating objects in 3D (Vega-Martínez et al. 2017). ML-based techniques have also been proposed to reconstruct shapes of free-falling liquid droplets (Zou et al. 2018). There are also ongoing investigations into a direct ML-aided reconstruction of 3D scenes from 2D data. For example, additional layers which output 3D coordinates can be added to the MASK R-CNN network (Danielczuk et al. 2019; Lang et al. 2019). In this case, it might be possible to make use of the 3D reprojection error as a metric for network losses in a training stage, together with 2D losses. A promising direction is to include the spatio-temporal dynamics directly into the recognition workflow instead of analysing each slide separately. One can start with the simultaneous analysis of several consecutive frames by means of networks of a similar architecture as Mask R-CNN, with an expanded output layer. Furthermore, it might be possible to build a hybrid RNN (Recurrent Neural Network)-CNN architecture for the analysis of longer video sequences.

Finally, based on the investigations of detection and tracking, the future improvements of the microgravity experiment are envisioned. Those might include optimized geometrical and lighting conditions, camera setup and particle colors and materials.

Declarations

Acknowledgements The authors acknowledge funding by DLR (German Aerospace Center). Helpful discussions with Prof. Wenping Wang (University of Hong Kong) and Alexey Tikhonov (Yandex) regarding 3D line matching are gratefully acknowledged.

Funding Open Access funding provided by Projekt DEAL. Funding is provided by DLR (German Aerospace Center) within project Equipage II, 50WM1842.

Availability of Data and Material Authors plan to make data sets generated and analysed during the current study available for open access.

Compliance with Ethical Standards

Conflict of interests Authors declare no conflict of interest/competing interests.

Code Availability Authors plan to open custom code developed in the current study for open access in the near future.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abdulla, W.: Mask R-CNN for object detection and instance segmentation on keras and tensorflow. https://github.com/matterport/Mask_RCNN (2017)
- Allan, D., van der Wel, C., Keim, N., Caswell, T.A., Wieker, D., Verweij, R., Reid, C., Thierry, L., Grueter, K., Ramos, A., Zoeith, R.W., Perry, F., Boulogne, P., Sinha, P., Bruot, N., Uieda, L., Katins, J., Mary, H., Ahmadi, A.: Soft-matter/trackpy: Trackpy v0.4.2 (2019)
- Aranson, I.S., Olafsen, J.S.: . Phys. Rev. E **66**, 061302 (2002)
- Aranson, I.S., Volfson, D., Tsimring, L.S.: . Phys. Rev. E **75**, 051301 (2007)
- Ashour, A., Wegner, S., Trittel, T., Börzsönyi, T., Stannarius, R.: . Soft Matter **13**, 402 (2017)
- Aumaitre, S., Behringer, R.P., Cazaubiel, A., Clément, E., Crassous, J., Durian, D.J., Falcon, E., Fauve, S., Fischer, D., Garcimartín, A., Garrabos, Y., Hou, M., Jia, X., Lecoutre, C., Luding, S., Maza, D., Noirhomme, M., Opsomer, E., Palencia, F., Pöschel, T., Schockmel, J., Sperl, M., Stannarius, R., Vandewalle, N., Yu, P.: . Rev. Sci. Instr. **89**, 075103 (2018)
- Bishop, C.M.: Pattern recognition and machine learning (Information Science and Statistics). Springer-Verlag, Berlin (2006)
- Brilliantov, N.V., Formella, A., Poeschel, T.: . Nature Commun. **9**, 797 (2018)
- Burkard, R., Dell'Amico, M., Martello, S.: Assignment problems, Revised Reprint. Other titles in applied mathematics Society for Industrial and Applied Mathematics (SIAM, 3600 Market Street, Floor 6, Philadelphia, PA 19104) (2009)
- Burton, J.C., Lu, P.Y., Nagel, S.R.: . Phys. Rev. E **88**, 062204 (2013)
- Copeland, M.F., Weibel, D.B.: . Nature Comm. **5**, 1174 (2008)
- Crassous, J.J., Dietsch, H., Pfliederer, P., Malik, V., Diaz, A., Hirshi, L.A., Drechsler, M., Schurtenberger, P.: . Soft Matter **8**, 3538 (2012)
- Crocker, J.C., Grier, D.G.: . J. Colloid Interface Sci. **179**, 298 (1996)
- Danielczuk, M., Matl, M., Gupta, S., Li, A., Lee, A., Mahler, J., Goldberg, K.: . (2019)
- Daniels, L.J., Park, Y., Lubensky, T.C., Durian, D.J.: . Phys. Rev. E **79**, 041301 (2009)
- Dutta, A., Zisserman, A.: . In: Proceedings of the 27th ACM international conference on multimedia. ACM, New York (2019). MM '19

- Falcon, E., Aumaitre, S., Evesque, P., Palencia, F., Lecoutre-Chabot, C., Fauve, S., Beysens, D., Garrabos, Y.: *Eur. Phys. Lett.* **74**, 830 (2006)
- Falcon, E., Wunenburger, R., Évesque, P., Fauve, S., Chabot, C., Garrabos, Y., Beysens, D.: *Phys. Rev. Lett.* **83**, 440 (1999)
- Girshick, R., Radosavovic, I., Gkioxari, G., Dollár, P., He, K.: *Detectron* <https://github.com/facebookresearch/detectron> (2018)
- Goldhirsch, I., Zanetti, G.: *Phys. Rev. Lett.* **70**, 1619 (1993)
- Grasselli, Y., Bossis, G., Morini, R.: *Eur. Phys. J. E* **38**, 8 (2015)
- Haff, P.K.: *J. Fluid Mech.* **134**, 401 (1983)
- Harth, K., Höme, S., Trittel, T., Kornek, U., Strachauer, U., Will, K.: In: *Proceedings of 20th ESA symposium on rocket and balloon science: (Europeanspace agency, Hyere, 2011)*, SP-700, p. 493 (2011)
- Harth, K., Kornek, U., Trittel, T., Strachauer, U., Höme, S., Will, K., Stannarius, R.: *Phys. Rev. Lett.* **110**, 144102 (2013a)
- Harth, K., Trittel, T., Kornek, U., Höme, S., Will, K., Strachauer, U., Stannarius, R.: *AIP Conference Proceedings* **1542**, 807 (2013b)
- Harth, K., Trittel, T., May, K., Wegner, S., Stannarius, R.: *Adv. Space Res.* **55**, 1901 (2015)
- Harth, K., Trittel, T., Wegner, S., Stannarius, R.: *EPJ web of conferences: Powders. Grains* **2017(140)**, 04008 (2017)
- Harth, K., Trittel, T., Wegner, S., Stannarius, R.: *Phys. Rev. Lett.* **120**, 214301 (2018)
- He, K., Gkioxari, G., Dollár, P., Girshick, R.B.: *arXiv:1703.06870* (2017)
- Hestroffer, D., Sanchez, P., Staron, L., Campo Bagatin, A., Eggl, S., Losert, W., Murdoch, N., Opsomer, E., Radjai, F., Richardson, D.C., Salazar, M., Scheeres, D.J., Schwartz, S., Taberlet, N., Yano, H.: *Astron. Astrophys. Rev.* **27**, 6 (2019)
- Hopkins, M.A., Louge, M.Y.: *Phys. Fluids A* **3**, 47 (1991)
- Hou, M., Liu, R., Zhai, G., Sun, Z., Lu, K., Garrabos, Y., Evesque, P.: *Micrograv. Sci. Technol.* **20**, 73 (2008)
- Hough, P.V.: *Method and means for recognizing complex patterns*. US Patent 3,069,654 (1962)
- Jung, A.B., Wada, K., Crall, J., Tanaka, S., Graving, J., Reinders, C., Yadav, S., Banerjee, J., Vecsei, G., Kraft, A., Rui, Z., Borovec, J., Vallentin, C., Zhydenko, S., Pfeiffer, K., Cook, B., Fernández, I., De Rainville, F.M., Weng, C.H., Ayala-Acevedo, A., Meudec, R., Laporte, M., et al.: *imgaug*. <https://github.com/aleju/imgaug>. Online; accessed 01-Feb-2020 (2020)
- Kiryati, N., Eldar, Y., Bruckstein, A.: *Pattern Recogn.* **24(4)**, 303 (1991). [https://doi.org/10.1016/0031-3203\(91\)90073-E](https://doi.org/10.1016/0031-3203(91)90073-E)
- Koch, D.L., Subramanian, G.: *Annu. Rev. Fluid Mech.* **43**, 637 (2011)
- Kudrolli, A., Lumay, G., Volfson, D., Tsimring, L.S.: *Phys. Rev. Lett.* **100**, 058001 (2008)
- Kumar, N., Soni, H., Ramaswamy, S., Sood, A.K.: *Nature Comm.* **5**, 4688 (2014)
- Lang, Y., Wang, L., Yap, P.T., Lian, C., Deng, H., Thung, K.H., Xiao, D., Yuan, P., Shen, S.G.F., Gateno, J., Kuang, T., Alfi, D.M., Xia, J.J., Shen, D.: In: *Zhang, D., Zhou, L., Jie, B., Liu, M. (eds.) Graph learning in medical imaging*, pp. 130–137. Springer International Publishing, Cham (2019)
- Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: In: *Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) Computer Vision – ECCV 2014*, pp. 740–755. Springer International Publishing, Cham (2014)
- Maaß, C.C., Isert, N., Maret, G., Aegerter, C.M.: *Phys. Rev. Lett.* **100**, 248001 (2008)
- Nichol, K., Daniels, K.E.: *Phys. Rev. Lett.* **108**, 018001 (2012)
- Noirhomme, M., Cazaubiel, A., Darras, A., Falcon, E., Fischer, D., Garrabos, Y., Lecoutre-Chabot, C., Merminod, S., Opsomer, E., Palencia, F., Schockmel, J., Stannarius, R., Vandewalle, N.: *EPL* **123**, 14003 (2018)
- Opsomer, E., Ludewig, F., Vandewalle, N.: *Phys. Rev. E* **84**, 051306 (2011)
- Pöschel, T., Luding, S. (eds.): *Granular gases. Lecture Notes in Physics*. Springer, Berlin (2001)
- Pöschel, T., Brilliantov, N.V. (eds.): *Granular gas dynamics. Lecture Notes in Physics*. Springer, Berlin (2003)
- Pöschel, T., Brilliantov, N.V.: *Kinetic theory of granular gases*. Oxford University Press, Berlin (2004)
- Russell, S., Norvig, P. *Artificial intelligence: A modern approach*, 3rd edn. Prentice Hall Press, USA (2009)
- Sack, A., Heckel, M., Kollmer, J.E., Zimber, F., Pöschel, T.: *Phys. Rev. Lett.* **111**, 018001 (2013)
- Shi, X., Ye, Z., Shiwakoti, N., Tang, D., Lin, J.: *Physica A: Statistical Mechanics and its Applications* **522**, 350 (2019)
- Thrun, S., Pratt, L.: *Learning to learn*. Springer, US (2012)
- Trittel, T., Harth, K., Stannarius, R.: *Phys. Rev. E* **95**, 062904 (2017)
- van Zon, J.S., Kreft, J., Goldman, D.I., Miracle, D., Swift, J.B., Swinney, H.L.: *Phys. Rev. E* **70**, 040301 (2004)
- Vega-Martínez, P., Rodríguez-Rodríguez, J., van der Meer, D., Sperl, M.: *Microgravity Sci. Technol.* **29**, 297–304 (2017)
- Weeks, E.R.: *Phys. Rev. Lett.* **107**, 065702 (2011)
- Wirtz, S., Paulus, D.: *Pattern Recognit: Image Anal.* **26(2)**, 354–359 (2016)
- Wu, Y., Kirillov, A., Massa, F., Lo, W.Y., Girshick, R.: *Detectron2* <https://github.com/facebookresearch/detectron2> (2019)
- Xu, H., Dauparas, J., Das, D., Lauga, E., Wu, Y.: *Nature Comm.* **10**, 1792 (2019)
- Yanpei, C., Evesque, P., Hou, M., Lecoutre, C., Palencia, F., Garrabos, Y.: *J. Phys.* **327**, 012033 (2011)
- Zheng, Z., Wang, F., Han, Y.: *Physics* **4**, 61 (2011)
- Zou, Z., Luo, X., Yu, Q.: *Microgravity Sci. Technol.* **30**, 321 (2018)
- Zuriguél, I., Parisi, D.R., Hidalgo, R.C., Lozano, C., Janda, A., Gago, P.A., Peralta, J.P., Ferrer, L.M., Pugnali, L.A., Clément, E., Maza, D., Pagonabarraga, I., Garcimartín, A.: *Sci. Rep.* **4**, 7324 (2015)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.