

METHODOLOGY ARTICLE

Open Access



High-resolution, non-invasive animal tracking and reconstruction of local environment in aquatic ecosystems

Fritz A Francisco^{1,2,3†}, Paul Nührenberg^{1,2,3†*}  and Alex Jordan^{1,2,3}

Abstract

Background: Acquiring high resolution quantitative behavioural data underwater often involves installation of costly infrastructure, or capture and manipulation of animals. Aquatic movement ecology can therefore be limited in taxonomic range and ecological coverage.

Methods: Here we present a novel deep-learning based, multi-individual tracking approach, which incorporates Structure-from-Motion in order to determine the 3D location, body position and the visual environment of every recorded individual. The application is based on low-cost cameras and does not require the animals to be confined, manipulated, or handled in any way.

Results: Using this approach, single individuals, small heterospecific groups and schools of fish were tracked in freshwater and marine environments of varying complexity. Positional tracking errors as low as 1.09 ± 0.47 cm (RSME) in underwater areas up to 500 m² were recorded.

Conclusions: This cost-effective and open-source framework allows the analysis of animal behaviour in aquatic systems at an unprecedented resolution. Implementing this versatile approach, quantitative behavioural analysis can be employed in a wide range of natural contexts, vastly expanding our potential for examining non-model systems and species.

Keywords: 3D tracking, Collective behaviour, Aquatic ecosystems, Computer vision, Structure from motion, Machine learning

Background

Understanding the movement and behaviour of animals in their natural habitats is the ultimate goal of behavioural and movement ecology. By situating our studies in the natural world, we have the potential to uncover processes of selection acting on behaviour in natural populations. The ongoing advance of animal tracking and biologging

brings the opportunity to revolutionize not only the scale of data collected from wild systems, but also the types of questions that can subsequently be answered. Incorporating geographical data has already given insights, for example, into the homing behaviour of reef fish, migratory patterns of birds, or the breeding site specificity of sea turtles [1–3]. Great advances in systems biology have further been made through the study of movement ecology, for example understanding the decision-making processes at play within primate groups manoeuvring through difficult terrain or the collective sensing of birds traversing their physical environment [4, 5]. Unravelling these aspects of animal movement can vastly improve management strategies [6, 7], for example in the creation of protected areas

*Correspondence: paul.nuehrenberg@uni-konstanz.de

[†]Fritz A Francisco and Paul Nührenberg contributed equally to this work.

¹Centre for the Advanced Study of Collective Behaviour, University of Konstanz, Universitätsstraße 10, 78457 Konstanz, Germany

²Department of Collective Behavior, Max Planck Institute of Animal Behavior, Universitätsstraße 10, 78457 Konstanz, Germany

Full list of author information is available at the end of the article



© The Author(s). 2020 **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/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

that incorporate bird migratory routes [8] or by reducing by-catch with dynamic habitat usage models of marine turtles [9].

Yet the application of techniques that meet the challenges of working in naturally complex environments is not straightforward, with practical, financial, and analytical issues often limiting the resolution or coverage of data gathered. This is especially true in aquatic ecosystems, where approaches such as Global Positioning System (GPS) tags allow only sparse positioning of animals that surface intermittently, or Pop-up Satellite Archival Tags (PSATs) which integrate surface positions with logged gyroscope and accelerometer data to estimate movement of larger aquatic animals [10, 11]. Not only does the spatial resolution of respective tracking systems, e.g. currently 4.9 m for GPS, limit the possibilities of behavioural analyses on a fine scale, but also excludes almost all animals below a certain size class [12]. These methods also require animals to be captured and equipped with tags that should not exceed 5% of the animals weight [13], further limiting current generation GPS and PSATs to larger animals. This is problematic because in aquatic ecosystems, as in terrestrial systems, life is numerically dominated by small animals [14]. In contrast, ultrasonic acoustic telemetry is one methodology useful for underwater tracking of smaller animals and those in larger groups [11, 15]. This approach is limited to a stationary site through the positioning of the acoustic receivers, and the costs, maintenance, and installation of these systems preclude their effective use in the majority of systems and for many users. While acoustic tags are small enough for injection, even for smaller animals such as fish, the increased handling time associated with these invasive measures can lead to additional stress for the animals, whereas the tag itself may disturb the animals' natural behaviour [16]. Further, acoustic telemetry systems also face accuracy problems, with average positional errors in the range of multiple meters, and highly depend on the environment (such as low ambient noise or sufficient water depth) in which these systems are deployed in [17, 18]. Hence, approaches that facilitate collection of behavioural data in smaller animals, those in large groups, and those in varied aquatic habitats, are still lacking.

A lack of data becomes a fundamental problem if certain ecosystems, species, or habitat types are underrepresented in terms of adequate research, management, or discovery. Although the oceans constitute up to 90% of habitable ecosystems worldwide, as little as 5% have been explored [19–21]. Within the oceans, coastal inshore areas have the greatest species diversity, with approximately 80% of fish species (the most speciose group of vertebrates) inhabiting the shallow waters of the littoral zone [22], and providing over 75% of commercial seafood landings [23]. Coastal regions in both marine and freshwater

environments are also those that are of greatest interest for eco-tourism, community fisheries, and industry, while simultaneously being most affected by habitat degradation, exploitation, and anthropogenic pollution [24–26]. Knowledge of the coastal regions is essential for establishing sanctuaries and sustainable concepts of ocean preservation [27] and movement data plays a vital role in this process, insofar as it gives detailed information about the location, preferred habitat and temporal distribution of organisms [13]. Yet for reasons of animal size, species abundance, and habitat complexity, most available tracking methods are poorly suited to these inshore regions.

Application of appropriate tracking and behavioural analysis techniques in a flexible, accessible, and broadly applicable manner would alleviate these limitations in systems and species coverage, improving capacity for conservation, management, and scientific understanding of natural systems across scales and conditions. In pure research terms, the application of quantitative behavioural and movement analyses in natural settings would also help bridge the gap between quantitative lab-based research and often qualitative field-based research. Recent advances in computational analysis of behaviour [28, 29] may then be employed in field settings, vastly improving our understanding of behaviour and movement in aquatic ecosystems.

Here we present an open-source, low-cost approach based on consumer grade cameras to quantify the movement and behaviour of animals of various sizes in coastal marine and freshwater ecosystems. Our approach integrates two methodologies from the field of computer vision, object detection with deep neural networks and Structure-from-Motion (SfM). Object detection has been successfully employed in terrestrial systems for animal localization, yielding high resolution movement data through e.g. drone-based videos over broad environmental contexts [30]. While these aerial approaches may also be used in some aquatic systems, they are limited to extremely shallow water and large animals [31]. The approach we advocate allows data to be collected on any animal that can be visualized with cameras, enabling application in smaller fish, invertebrates, and other aquatic animals. In addition to providing animal trajectories, video-based observations also incorporate environmental data that adds the possibility to study interactions of mobile animals with their natural habitat [4]. Our approach synthesizes object detection with SfM into a coherent framework that can be deployed in a variety of systems without domain-specific expertise. SfM is commonly used for 3D environmental reconstructions, photogrammetry and camera tracking for visual effects in video editing [32, 33], and here allows the reconstruction of 3D models of the terrain through which the

animals move and interact with. Our open-source analysis pathway enables subsequent calculation of movement, interactions, and postures of animals. Set-up costs can be as small as two commonly available action cameras, and the proposed method can be taken into habitats which are otherwise explored by snorkeling, diving, or with the use of remotely operated underwater vehicles (ROVs). Analysis can be performed on local GPU-accelerated machines or widely-accessible computing services (e.g. Google Colaboratory). Overall, this method provides a low-cost approach for measuring the movement and behaviour of aquatic animals that can be implemented across scales and contexts.

Methods

Three datasets of varying complexity were used to demonstrate the versatility of the proposed method. These were chosen to range from single animals (*Conger conger*) and small heterospecific groups (*Mullus surmuletus*, *Diplodus vulgaris*) to schools of conspecific individuals (*Lamprologus callipterus*) under simple and complex environmental conditions, resulting in the datasets 'single', 'mixed' and 'school', respectively. Moreover, we used a dataset of repeated trials ($N = 4$, 'accuracy') to validate the accuracy of our tracking approach. This dataset was used to reconstruct the trajectories of a calibration wand of 0.5 m length and examine resulting tracking errors. The 'single' and 'mixed' datasets were created while snorkeling at the surface, using a stereo camera set-up at STARESO, Corsica (Station de Recherche Océanographiques et sous-marines). The remaining datasets were collected by SCUBA diving (5–8 m) with either multi or stereo camera set-ups in Lake Tanganyika (Tanganyika Science Lodge, Mpulungu, Zambia), or at STARESO. While the 'single' and 'mixed' datasets were recorded with untagged fish, we attached tags made of waterproof paper (8×8 mm) anterior to the dorsal fin of the individuals for the 'school' dataset to facilitate detection and individual identification, although the latter was not implemented. See Table 1 for a summary of the collected datasets and

respective environmental conditions. For a general guideline and comments on the practical implementation of our method, refer to Additional file 7.

Automated animal detection and tracking

Since all data was collected in the form of videos, image-based animal detection was required for subsequent trajectory reconstruction and analyses. First, the videos from the stereo or multi-camera set-ups were synchronized using a convolution of the Fourier-transformed audio signals to determine the video offsets. Second, the synchronized videos were tracked independently using an implementation of a Mask and Region based Convolution Neural Network (Mask R-CNN) for precise object detection at a temporal resolution of either 30 Hz ('single', 'mixed' and 'accuracy') or 60 Hz ('school') [34, 35]. To this end, we annotated the contours of the fish (or the tags in case of the 'accuracy' dataset) in a small subset of video frames to generate custom training datasets for each of the detection tasks. These subsets needed to be sufficiently diverse to represent the full videos for effective training and, therefore cover most of the variation in contrast, lighting and animal poses. Our training sets contained 171 ('single'), 80 ('mixed'), and 160 ('school') labeled images for each dataset. For the 'accuracy' dataset, we annotated a total of 73 images. We then trained Mask R-CNN models on these training sets using transfer learning from a model that was pre-trained on the COCO dataset ('Common Objects in Context') with more than 200K labeled images and 80 object classes [35, 36]. Here, transfer learning refers to a machine learning concept in which information gained from learning one task is applied to a different, yet related problem [37]. Therefore, the state of Mask R-CNN, previously trained on COCO, was fine-tuned to our specific problems of identifying fish or tags. The original image resolutions of 2704×1520 px ('single' and 'school') and 3840×2160 px ('mixed' and 'accuracy') were downsampled to a maximum width of 1024 px while training and predicting to achieve better performance. After training, the models were able to accu-

Table 1 Summary of acquired datasets

Dataset	Location	Approach	Species	Duration (m:ss)	N	Setup	Dist. (m)	Tags	Pose	Complexity
single	STARESO	snorkel	<i>Conger conger</i>	0:20	1	stereo	0.4	no	yes	high
mixed	STARESO	snorkel	<i>Mullus surmuletus</i> ,	7:09	1,	stereo	0.6	no	yes	low
			<i>Diplodus vulgaris</i>		2					
school	Tanganyika	dive	<i>Lamprologus callipterus</i>	0:35	11	multi (12)	0.2	yes	no	medium
accuracy	STARESO	dive	NA	3:14 - 5:00	NA	multi (4)	0.6	NA	NA	varying

N lists the number of tracked individuals, *Dist.* the minimum camera-to-camera distance in the setups, *Tags* whether individual animals were tagged and *Pose* if animal spine pose estimation was used during tracking. *Complexity* lists an estimate of overall complexity: *high* (single individual with complex posture, variable lighting and contrast, motile background elements), *medium* (multiple individuals, high turbidity and greater depth, visible tags), *low* (few individuals, good lighting, homogeneous background), *varying* (intentionally varied complexity). *NA*: not applicable

rately detect and segment the observed animals, which was visually confirmed with predictions on validation datasets.

The predicted masks were either used to estimate entire poses of the tracked animals (*'single'*, *'mixed'*) or to calculate centroids of the tags or calibration wand ends in case of the *'school'* and *'accuracy'* datasets. Established morphological image processing was used to skeletonize the Mask R-CNN predictions, producing a 1 px midline for each of the detected binary masks. A fixed number of points was equidistantly distributed on these midlines as an estimation of the animals' spine poses. Both the spine points and the tag centroids represent pixel coordinates of detected animals in further data processing. Partitioned trajectories were generated from detections with a simple combination of nearest-neighbors between subsequent frames or utilizing a cost-reduction algorithm (the Hungarian method [38]), and filtering for linear motion over a short time window, reducing later quality control and manual track identification for continuous trajectories to a minimum. For video and image annotations, trajectory and pose visualization, manual track corrections and other trajectory utility functions, we developed a GUI based on Python and Qt5 within the lab (*'TrackUtil'*, Additional file 4). The code for Mask R-CNN training and inference, video synchronization, fish pose estimation and automatic trajectory assignment is also available (Additional files 5 and 6). The training and tracking details are summarized in Table 2.

Structure from motion

The field of computer vision has developed powerful techniques that have found applications in vastly different fields of science [39–41]. The concept of Structure-from-Motion (SfM) is one such method that addresses the large scale optimization problem of retrieving three dimensional information from planar images [42]. This approach relies on a static background scene, from which stationary features can be matched by observing them from different perspectives. This results in a set of images, in which feature-rich key points are first detected and subsequently used to compute a 3D reconstruction of the scene and the corresponding view point positions. As shown in Eqs. 1 and (2), a real world 3D point M' (consisting of x, y, z) can be projected to the image plane of an observing camera by multiplying the camera's intrinsic matrix K (consisting of focal lengths f_x, f_y and principal point c_x, c_y), with the camera's joint rotation-translation matrix $[R|t]$ and M' , resulting in the corresponding image point m' (consisting of pixel coordinates u, v , scaled by s) [43]. By extension, this can be used to resolve the ray casting from a camera position towards the actual 3D coordinates of a point given the 2D image projection of that point with known camera parameters. Due to this projective geometry, it is not possible to infer at which depth a point is positioned on its ray from a single perspective. SfM is able to circumvent this problem by tracking mutually-observed image points (m') across images of multiple camera view points. As a result, the points can be triangulated in 3D space (M'), representing the optimal intersections of their respective

Table 2 Dataset parameters and accuracy metrics

Dataset	Annotations	Rate (Hz)	Resolution (px)	Coverage (%)	Accuracy metrics			
					Metric	Reconstruction (cm)	Reprojection (px)	Tracking (cm)
<i>single</i>	171	30	2.7k	97.79	median	0.30	9.65	NA
					RMSE	1.28	16.30	NA
<i>w/sv</i>	as above			100.00	as above			
<i>mixed</i>	80	30	4k	69.60	median	0.44	3.77	NA
					RMSE	1.09	7.77	NA
<i>school</i>	160	60	2.7k	78.38	median	0.06	2.57	NA
					RMSE	0.30	3.78	NA
<i>w/sv</i>	as above			94.02	as above			
<i>accuracy</i>	73	30	4k	80.64 ± 16.73	median	-0.14 ± 0.06	3.53 ± 1.96	0.14 ± 0.33
					RMSE	1.34 ± 0.79	8.56 ± 5.21	1.09 ± 0.47
<i>w/sv</i>	as above			97.29 ± 2.20	median	as above		0.28 ± 0.32
					RMSE			2.12 ± 1.37

'w/sv' indicates that trajectory points were also estimated from single-view projections at an interpolated depth component. *Annotations* lists how many frames were annotated for training Mask R-CNN, *Rate* the frames per second of each video set, i.e. the temporal tracking resolution. *Resolution* is video resolution, 2.7k: 2704 × 1520 px, 4k: 3840 × 2160 px. *Coverage* is the mean coverage off all individual trajectories of a dataset. *Reconstruction* metrics refer to the deviation of reconstructed camera-to-camera distances from the actual distance, *Reprojection* metrics to the reprojection of triangulated 3D tracks to the original video pixel coordinates and *Tracking* to the deviation of the tracked calibration wand length from its actual length. In case of the *'accuracy'* dataset, the accuracy results are listed as the mean and standard deviation of the four repeated trials. *NA*: not applicable

rays pointing from the cameras positions towards them. By minimizing reprojection errors, which are the pixel distances between the 3D points' reprojections to the image planes and their original image coordinates (u, v), SfM is also able to numerically solve the multi-view system of the cameras relative rotation (R), translation (t) and intrinsic (K) matrices and to retrieve the optimal camera distortion parameters (d).

$$m' = K[R|t]M' \quad (1)$$

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{21} & r_{31} & t_1 \\ r_{12} & r_{22} & r_{32} & t_2 \\ r_{13} & r_{23} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \quad (2)$$

Here, SfM was incorporated into data processing in order to gain information about exact camera positions, which was done using the general-purpose and open-source pipeline COLMAP [44, 45]. The synchronized videos were resampled as image sequences with a rate of 3 Hz. In case of the 'mixed' dataset, we removed frames that were recorded when the cameras were stationary. The resulting image sequences served as input into the reconstruction process during which the cameras were calibrated (K, d) and relative extrinsic parameters (R, t) computed, so that all camera projections relate to a shared coordinate system. Every input image resulted in a corresponding position along the reconstructed, 3D camera path of the recording, where the number of images determined the temporal resolution of resolved camera motion. Since only a subset of all video frames were used for the reconstructions, SfM optimized a smaller number of parameters, resulting in a reduced computational load. Additionally, this could improve reconstruction accuracy, as the images still had sufficient visual overlap, but increased angles between view points. Finally, the retrieved camera parameters were interpolated (the translations t linearly, rotations R using Slerp, spherical linear interpolation [46]) to match the acquisition rate of animal tracking, assuring that reference camera parameters are given for each recorded data point by simulating a continuous camera path.

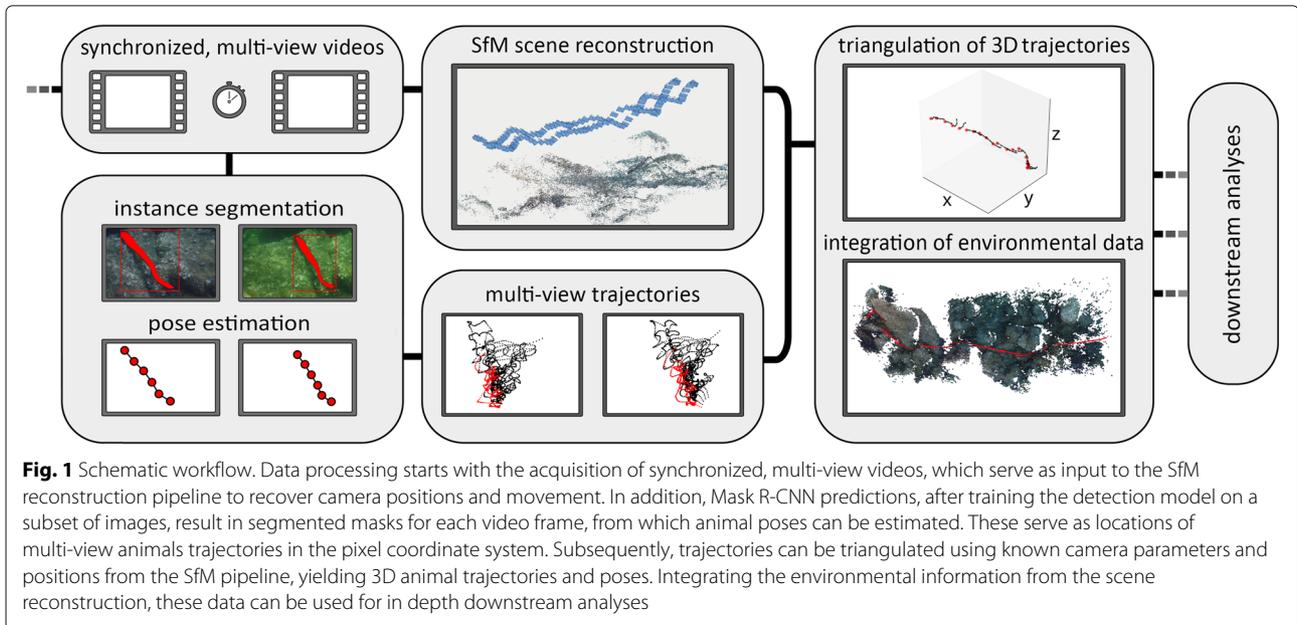
Reconstruction of animal trajectories

It is necessary to resolve the camera motion when tracking moving animals with non-stationary cameras, since the camera motion will also be represented in the pixel coordinate trajectories of the animals. With camera information (K, d) and relative perspective transformations (R, t) for entire camera paths retrieved from SfM, as well as multi-view animal trajectories from the Mask R-CNN detection pipeline available, a triangulation approach similar to SfM can be used to compute 3D animal trajectories.

Here, an animal's pixel coordinates represent m' (consisting of u and v) observed from more than one known view point (R, t), and the animals 3D positions M' (x, y, z) can be triangulated. Positions of animals observed in exactly two cameras were triangulated using an OpenCV implementation of the direct linear transformation algorithm, while positions of animals observed in more than two cameras were triangulated using singular value decomposition following an OpenCV implementation [39, 43]. Additionally, positions of animals temporarily observed in only one camera were projected to the world coordinate frame by estimating the depth component as an interpolation of previous triangulation results. Through the recovered camera positions, the camera motion is nullified in the resulting 3D trajectories. Thus, they provide the same information as trajectories recorded with a fixed camera setup (Fig. 1). Animal trajectories and the corresponding reconstructions were scaled, so that the distances between the reconstructed camera locations equal the actual distances within the multi-view camera setup. As a result, all observations are represented on a real world scale. The code for trajectory triangulation, camera path interpolation and visualizations is bundled in a Python module ('*multiviewtracks*'), accessible on GitHub [47].

Accuracy estimation

Given that the proposed method incorporates out-of-domain and novel approaches from computer vision, reliable accuracy measures are required. Therefore, a ground-truth experiment ('*accuracy*' dataset) was conducted in which two points of fixed distance to each other, the colored end points of a clear, rigid calibration wand (0.5 m), were filmed underwater over various backgrounds using four cameras. In total, four repeated trials were incorporated for the accuracy estimation, varying in environmental complexity (such as *Poseidonia* sea grass beds, large rock formations or sand), depth and lighting conditions. The trajectories of both calibration wand end points were reconstructed throughout the four trials, which enabled the calculation of a per-frame tracking error. The ground-truth distance (\hat{d}) between the two 3D positions is known from the wand length for each frame, hence the difference of the reconstructed distance of these two 3D positions (d) from the actual distance (\hat{d}) can be calculated as this tracking error. Additionally, since the cameras were arranged in a fixed, multi-view setup, the same calculation can be performed on the known camera-to-camera distances (\hat{d}) within the array and their reconstructed 3D positions (and respective distances d) to assess errors of the SfM reconstructions. A third measure of accuracy can be calculated as the reprojection error of triangulated trajectory points. Here, the 3D points are projected back to the image planes of their respective view points, resulting in



pixel coordinates for each 3D point and observing camera. The distance of these pixel coordinates to the tracked pixel coordinates m' (consisting of u and v) is the reprojection error. This is the error which is used by SfM for numeric optimization of the multi-view system, the camera parameters and the scene's 3D point cloud, and can be similarly used to estimate the precision of the acquired trajectories. We calculated the median errors and the standard deviations of the errors, i.e. the root-mean-square errors (RMSEs, equation 3) for all datasets and for each of the three accuracy metrics when applicable. In case of the 'accuracy' dataset, we calculated the mean and standard deviation of the accuracy metrics for the four trials.

$$RMSE = \frac{1}{N} \sqrt{\sum_{N=1}^N (\hat{d} - d_N)^2} \quad (3)$$

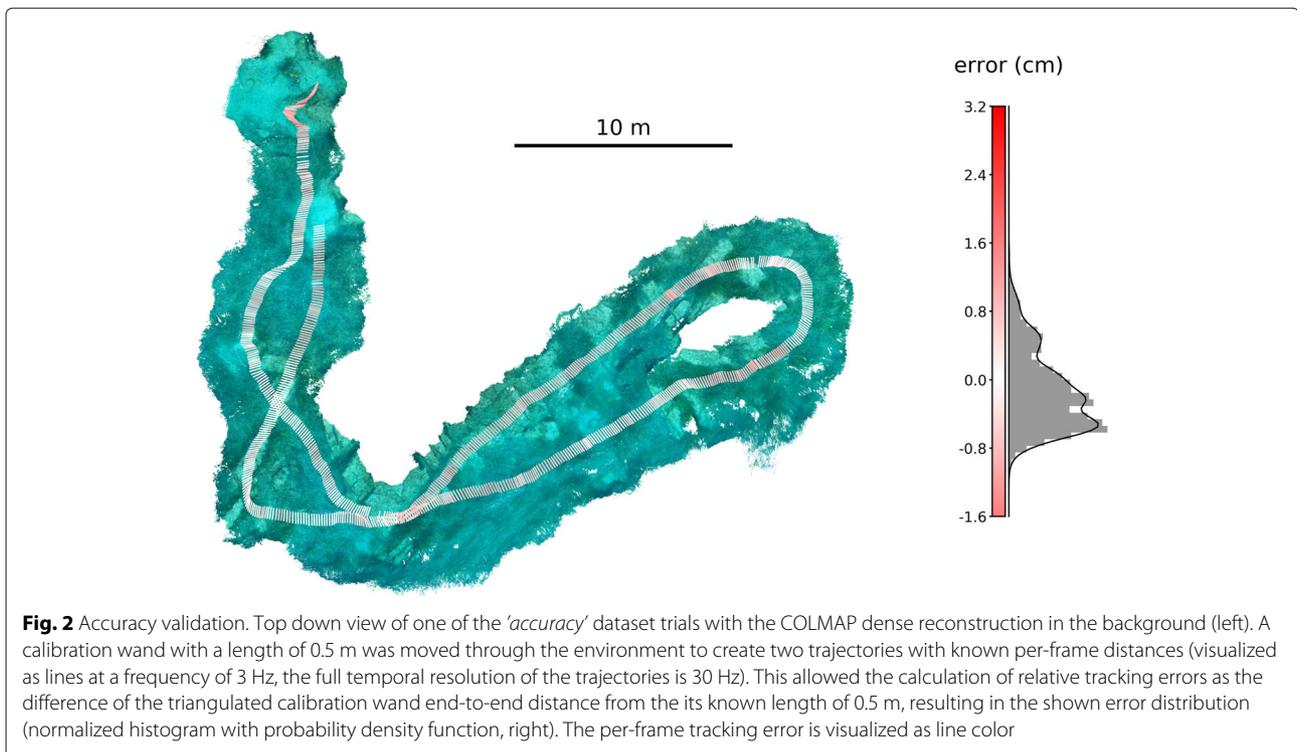
Results

Here we combined Mask R-CNN aided animal detection and tracking with SfM scene reconstruction and triangulation of 3D animal trajectories to obtain high resolution data directly from videos taken while snorkeling or diving in the field. Using this method, we were able to track freely moving, aquatic animals in their natural habitats without installation of infrastructure.

In order to ground truth our method, we performed an accuracy estimation for the four trials of the 'accuracy' dataset. Using our approach, we were able to retrieve both the 3D positions of the tracked calibration wand and the 3D trajectories of the cameras throughout the trials

(Fig. 2). The mean trajectory coverage was $80.64 \pm 16.73\%$ when only multiple-view triangulation was used, or $97.3 \pm 2.2\%$, when also projections from single views were used to estimate trajectory positions at an interpolated depth component. This resulted in a total of 19482 frames in which both wand ends were detected (or 26562 with the additional single-view projections). The known camera-to-camera distances within the camera array (0.6 m) and the known length of the calibration wand (0.5 m) allowed the calculation of respective per-frame reconstruction and tracking errors. The resulting RMSE for the camera-to-camera distances was 1.34 ± 0.79 cm (median error -0.14 ± 0.06 cm). The errors for the calibration wand length differed when calculated for only multi-view triangulated trajectories (RMSE 1.09 ± 0.47 cm, median error 0.14 ± 0.33 cm) or for trajectories with single-view projections (RMSE 2.12 ± 1.37 cm, median error 0.28 ± 0.32 cm). Further, we projected the triangulated 3D positions back to the original videos and computed the reprojection error as a RMSE of 8.56 ± 5.21 px (median error 3.53 ± 1.96 px). This was only done for the multi-view triangulations, since the reprojection of a point that was projected from a single view is, by definition, the same point (with a potentially misleading error of 0 px).

Trajectories were successfully obtained from large groups ('school'), small groups ('mixed') and single individuals ('single'). Due to the specific design of Mask R-CNN for instance segmentation, the network architecture was able to distinguish given object classes from the background and solved partial occlusions. However, differences in data acquisition remained across these datasets. For example, this resulted in varying track coverage, 97.79%, 69.60% and 78.34% for the 'single', 'mixed'



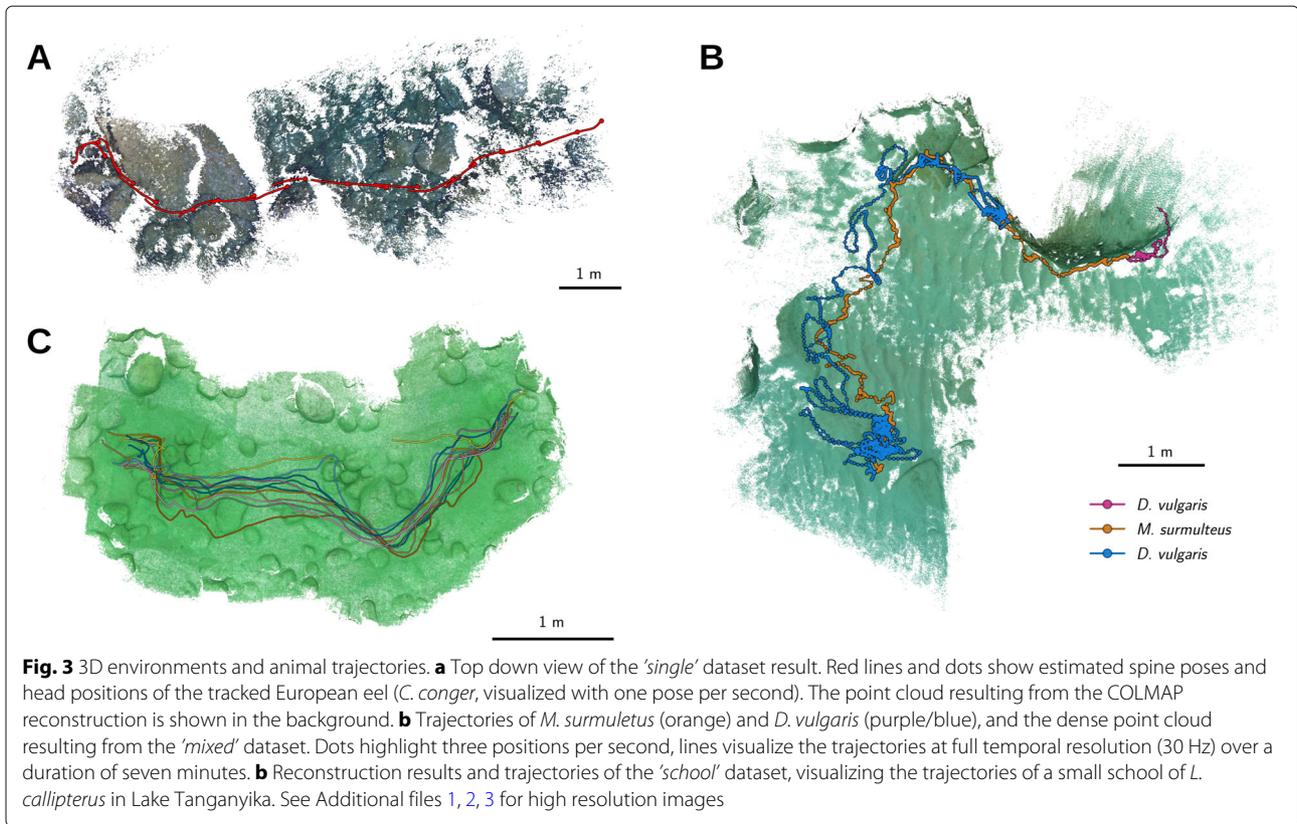
and 'school' datasets, respectively. When also single-view projections were included in the animal trajectories, the trajectory coverage increased to 100.00% ('single') and 94.02% ('school'). Additionally, the camera positions and corresponding environments through which the animals were moving were reconstructed. In case of the 'single' and 'mixed' datasets, the Mask R-CNN detection results were used to estimate fish body postures in 3D space by inferring spine points from the segmented pixel masks (Fig. 3). We computed the RMSEs of the camera-to-camera distances (1.28 cm 'single', 1.28 cm 'mixed' and -0.15 cm 'school') and reprojection errors (20.97 px 'single', 7.77 px 'mixed' and 6.79 px 'school') to assess the overall quality of the SfM reconstructions analogously to the calculation of reconstruction errors for the 'accuracy' dataset. The results of the accuracy estimations and respective median errors are listed in Table 2.

Discussion

Here we demonstrate a novel approach to collect highly resolved 3D information of animal motion, including interactions with the physical environment, in aquatic ecosystems. Although being based on relatively advanced computational techniques, the open-source workflow we present requires little domain expertise and can be implemented with low-cost consumer grade cameras. The

incorporation of these methods into an accessible framework will allow quantitative analyses of animal behaviour and ecology across systems, scales, and user groups, and can even be modified for use in terrestrial systems. Our approach allows data collection while swimming, snorkelling, or with the aid of ROVs, making it appropriate for general usage with minimal investment into infrastructure, equipment, or training. Although analyses are computationally demanding, they can be achieved on an average GPU or free cloud-based computing services. The lack of high-end hardware therefore does not interfere with any of the steps required for this method.

Many alternative techniques for tracking of small aquatic animals do exist, however, they often have the considerable drawback of tagging and handling the animals or high infrastructure costs. This is a major barrier to implementation when animals are protected, difficult to catch, or too small to carry tags. In many marine protected areas all three of these factors apply, meaning that many existing approaches are inappropriate. Some of these drawbacks will be alleviated, for instance with improvements in telemetry-based approaches [48] that reduce tag size and increase range. Nevertheless, these techniques cannot simultaneously measure or reconstruct local topography and environmental factors. Although here we do not provide any analyses of environmental structure, this topographical information collected with



our approach can be directly used to answer questions on e.g. habitat segmentation and environmental complexity [49, 50].

In highly complex social environments, encounters with numerous con- and heterospecifics can strongly affect behaviour and motion [51]. Using approaches that rely on tagging will unavoidably miss or under-sample these interactions because not all individuals can ever be tagged in wild contexts. In contrast, our approach does not require animals to be handled or tagged, nor does specialized equipment need to be deployed in the desired tracking area. Moreover, because the object detection and segmentation approach can take any image input, it is not tied to one particular animal form or visual scene. Our approach can therefore be used even in demanding conditions such as high turbidity or low-light conditions, within certain limits. While it has a lower spatial range than telemetry, underwater filming comes as an unintrusive alternative, with higher spatial resolution possible when small animals are moving over small areas, or when animals are highly site-specific, for example damselfish or cichlids living in close association with coral or rocky reef [52, 53].

While our approach offers many benefits in terms of applicability and data acquisition, it also suffers from some limitations. From the accuracy tests it became apparent that in cases where the background was composed of

moving objects, such as macrophytes or debris, the tracking accuracy dropped noticeably. The SfM approach relies on the reconstructed components to be feature-rich and static, because environmental key-points are assumed to have the same location over time. Moving particles and objects will result in higher reconstruction errors, rendering our approach problematic e.g. when the filmed animals occupy most of the captured images in case of very large fish schools. Complex environments, occlusions of the animals and highly variable lighting conditions are detrimental to the detectability of animals with Mask R-CNN. Observations at greater depths may face similar problems due to the high absorption of red light, although, in this case, detectability could be alleviated through image augmentation approaches such as SeaThru [54]. Similarly, water turbidity can greatly affect the detectability in aquatic systems by absorbing light and diffusing the scene. Therefore, although removing the benefits from measuring animal behaviour non-invasively, it can be advantageous to add clearly visible tags to the animals in cases of high turbidity, ensuring continuous tracking of all individuals.

Another aspect that needs consideration is that data acquisition is confined to the area captured by the multi-camera set-up. Animal trajectories can not be triangulated if individuals leave this area, and therefore are no longer visible from at least two camera view points. This

circumstance is apparent in the 'school' example, in which one individual left and re-entered the scene, leading to a discontinuity in its trajectory. To compensate the potential limitation in trajectory coverage, trajectory points can also be estimated from single-view detections by projecting them from this view point to an interpolated depth. However, this is only possible when filmed from above and for animals that do not drastically change the distance to the camera (otherwise, the estimated depth component would likely be erroneous). As a consequence, we report the lowest trajectory coverage in the 'mixed' dataset (69.60%), in which we filmed the animals with a semi-stationary tripod and isometric camera angles. Considering the temporal resolution of 30 Hz, this still resulted in a relatively high average detection rate of approximately 21 detections per second. Further, we could demonstrate with the 'accuracy' dataset, that although the single-view projections can increase track coverage significantly (from $80.64 \pm 16.73\%$ to $97.29 \pm 2.20\%$), they also come at a moderate accuracy cost (calibration wand length RMSE increased from 1.09 ± 0.47 to 2.12 ± 1.37 cm).

The estimation of 3D animal poses strongly relies on accurate detections and can therefore be compromised by poorly estimated animal shapes during Mask R-CNN segmentation. In these cases, a less detailed approximation of the animals' positions such as the mask centroids are favorable and can still be reliably employed as showcased with the 'school' dataset. The errors in estimating animal locations and poses can be partially explained by marginal detection errors of Mask R-CNN, but also by inaccuracies derived from trajectory triangulation using the SfM camera positions.

Aware of these error sources, users can incorporate accuracy metrics such as reprojection errors or relative camera reconstruction RMSEs into their own analytical pathways by using our proposed method. This enables the assessment of the overall reconstruction quality and required fine scale resolution for the specific scientific demands. We were able to demonstrate with the 'accuracy' dataset, that the combination of SfM and object detection yields highly accurate trajectories of moving objects over large spatial scales (RMSE tracking error of 1.34 ± 0.79 cm, median error -0.14 ± 0.06 cm, reconstructed areas up to 500 m^2) without prior manipulation of the underwater environment. Since these accuracy calculations are based on per-frame deviations from known distances, such as the length of a calibration wand or camera-to-camera distances in a stereo-camera setup, they are not suited for the assessment of large-scale SfM accuracy. However, rigorously ground-truthing SfM is of general interest in the field of computer vision, and various benchmarks showcase the high precision of 3D reconstructions that can be achieved using current SfM pipelines [55, 56].

An additional requirement of our approach is associated with the need to annotate images and train object detection networks. Further, manual correction of false trajectory assignments and overall quality-control are required, but can be reduced to a minimum with adequately-sized training sets and resulting, precise Mask R-CNN predictions. Reliable and automatic identification of unmarked individuals in large animal groups recently became possible in laboratory conditions [57], and future development and increasing robustness of similar methods might also enable them for field observations. However, at present, these tasks present an additional, mainly initial, time investment that is likely to be compensated by the time subsequently saved using high-throughput behavioural analyses on the acquired, highly-resolved animal trajectories. For example, this allows the classification of behavioural states by quantifying the behavioural repertoire of the animals using unsupervised machine learning techniques [28, 58]. The incorporation of 3D trajectory data in motion analyses has already improved the understanding of the phenotype and development of animal behaviours [59]. In addition, 3D pose estimation can now be achieved for wild animals, enabling exact reconstruction of the entire animal [60]. There has been a shift in how animal movement is analyzed in light of computational ethological approaches [61–63], with patterns of motion able to be objectively disentangled, revealing the underlying behavioural syntax to the observer. Automated approaches based on video, or even audio, recordings may also overcome sensory limitations of other systems, allowing a better understanding of the sensory *umwelt* of study species [64] and also facilitate novel experimental designs [61, 65] that can tackle questions of the proximate and ultimate causality of behaviour [60, 62, 63]. These methods are gaining interest and sharply contrast with the traditional approach of trained specialists creating behavioural ethograms, but can usefully be combined and compared to gain further insight into the structure of animal behaviour, potentially generating a more objective and standardized approach to the field of behavioural studies [63].

In order to incorporate these novel techniques into more natural scenarios, we aim to present a complete tracking pipeline, guiding the user through each step after the initial field observation. From video synchronization, object annotation and detection to the final triangulation of animal trajectories, we provide a set of open-source utilities and scripts. Although we heavily rely on other open-source projects (COLMAP for SfM and Mask R-CNN for object segmentation), these specific approaches can be replaced with other implementations by solely adopting respective in- and output data formatting for specific needs. We found COLMAP and Mask R-CNN to be easily employed, as they are well documented,

performant and purpose-oriented. However, many alternatives exist for both SfM and object detection, and the general approach of our pipeline is not limited to any particular implementation, thus future-proofing this approach as new and better methods are developed.

Conclusions

Computational approaches to analyze behaviour, including automated tracking of animal groups, deep-learning, supervised, and unsupervised classification of behaviour, are areas of research that have been extensively developed in laboratory conditions over the past decade. These techniques, in combination with sound evolutionary and ecological theory, will characterize the next generation of breakthroughs in behavioural and movement science, yet are still difficult to achieve in natural contexts, and are unobtainable for many researchers due to implementation and infrastructure costs. Here we present a framework to enable the utilization of these cutting-edge approaches in aquatic ecosystems, at low-cost and for users of different backgrounds. Our proposed tracking method is flexible in both the conditions of use, and the study species being examined, vastly expanding our potential for examining non-model systems and species. In combination with the genomic revolution, allowing sequencing in a matter of days, state-of-the-art behavioural sequencing under field conditions will revolutionize the field of movement ecology and evolutionary behavioural ecology. The approach we advocate here can further integrate the study of wild animal behaviour with modern techniques, facilitating an integrative understanding of movement in complex natural systems.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s40462-020-00214-w>.

Additional file 1: Tracking results of 'single' dataset. Top down view of 'single' results: dense COLMAP 3D reconstruction and trajectories of the tracked animal, *C. conger* (red).

Additional file 2: Tracking results of 'mixed' dataset. Top down view of 'mixed' results: dense COLMAP 3D reconstruction and trajectories of the tracked animals, *M. surmuletus* (orange) and *D. vulgaris* (purple/blue).

Additional file 3: Tracking results of 'school' dataset. Top down view of 'school' results: dense COLMAP 3D reconstruction and trajectories of the tracked animals, *L. callipterus*.

Additional file 4: TrackUtil. Python and Qt5 based GUI for image annotations, trajectory visualization and manual track corrections.

Additional file 5: Mask R-CNN training and inference. Our training regime for Mask R-CNN and inference on videos. For more information on Mask R-CNN, visit the original repository at https://github.com/matterport/Mask_RCNN.

Additional file 6: Video synchronization, trajectory assignment and pose estimation. Additional scripts for video synchronization, frame extraction, trajectory assignment and pose estimation.

Additional file 7: A practical guide. A general guideline and comments regarding the implementation of our method.

Abbreviations

GPS: Global Positioning System; PSAT: Pop-up satellite archival tag; ROV: Remotely operated underwater vehicle; Mask R-CNN: Mask and region based convolution neural network; SfM: Structure-from-motion; Slerp: Spherical linear interpolation; RMSE: Root-mean-square error; SCUBA: Self-contained underwater breathing apparatus

Acknowledgements

We would like to thank the entire Department of Collective Behaviour at the University of Konstanz for their support in making this project possible. We thank Philip Fourmann, Myriam Knöpfle and Jessica Ruff for contributing the 'single' and 'mixed' species footage. Special thanks also to Hemal Naik, Simon Gingins and Eduardo Sampaio for their suggestions and helpful input. We sincerely thank Etienne Lein for his substantial assistance and support during data acquisition in the field. Further, we thank the COLMAP team for making it open-source and easily accessible.

Authors' contributions

FAF, PN and AJ collected the raw data in the field. PN and FAF wrote code for data acquisition and performed analyses. PN, FAF, and AJ wrote the manuscript. All authors agree on the standards of authorship put forth by this journal. The author(s) read and approved the final manuscript.

Funding

This project was funded by the Deutsche Forschungsgemeinschaft (DFG, German research Foundation) under Germany's Excellence Strategy - EXC 2117 - 422037984.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on request. All code used for analysis is open-source and either accessible online ('multiviewtracks', <https://doi.org/10.5281/zenodo.3666726>) or provided as Additional files 4, 5, and 6 (TrackUtil, video synchronization, Mask R-CNN training and inference, trajectory assignment and pose estimation).

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Author details

¹Centre for the Advanced Study of Collective Behaviour, University of Konstanz, Universitätsstraße 10, 78457 Konstanz, Germany. ²Department of Collective Behavior, Max Planck Institute of Animal Behavior, Universitätsstraße 10, 78457 Konstanz, Germany. ³Department of Biology, University of Konstanz, Universitätsstraße 10, 78457 Konstanz, Germany.

Received: 26 February 2020 Accepted: 26 May 2020

Published online: 23 June 2020

References

- Gerlach G, Atema J, Kingsford MJ, Black KP, Miller-Sims V. Smelling home can prevent dispersal of reef fish larvae. *Proc Natl Acad Sci*. 2007;104(3):858–63.
- Sommeille M, Manica A, Rodrigues AS. Where the wild birds go: explaining the differences in migratory destinations across terrestrial bird species. *Ecography*. 2018. <https://doi.org/10.1111/ecog.03531>.
- Carr A, Carr MH. Site fixity in the caribbean green turtle. *Ecology*. 1972;53(3):425–9.
- Strandburg-Peshkin A, Farine DR, Crofoot MC, Couzin ID. Habitat and social factors shape individual decisions and emergent group structure during baboon collective movement. *eLife*. 2017;6:19505.
- Nagy M, Couzin ID, Fiedler W, Wikelski M, Flack A. Synchronization, coordination and collective sensing during thermalling flight of freely migrating white storks. *Phil Trans R Soc B*. 2018;373(1746):20170011.

6. Cooke SJ. Biotelemetry and biologging in endangered species research and animal conservation: relevance to regional, national, and iucn red list threat assessments. *Endanger Spec Res.* 2008;4(1-2):165–85.
7. Costa DP, Breed GA, Robinson PW. New insights into pelagic migrations: implications for ecology and conservation. *Ann Rev Ecol Evol Syst.* 2012;43:73–96.
8. Runge CA, Watson JE, Butchart SH, Hanson JO, Possingham HP, Fuller RA. Protected areas and global conservation of migratory birds. *Science.* 2015;350(6265):1255–8.
9. Lewison R, Hobday AJ, Maxwell S, Hazen E, Hartog JR, Dunn DC, Briscoe D, Fossette S, O'keefe CE, Barnes M, et al. Dynamic ocean management: identifying the critical ingredients of dynamic approaches to ocean resource management. *BioScience.* 2015;65(5):486–98.
10. Thys TM, Ryan JP, Dewar H, Perle CR, Lyons K, O'Sullivan J, Farwell C, Howard MJ, Weng KC, Lavaniegos BE, Gaxiola-Castro G, Bojorquez LEM, Hazen EL, Bograd SJ. Ecology of the ocean sunfish, mola mola, in the southern california current system. *J Exp Mar Biol Ecol.* 2015;471:64–76. <https://doi.org/10.1016/j.jembe.2015.05.005>.
11. Hussey NE, Kessel ST, Aarestrup K, Cooke SJ, Cowley PD, Fisk AT, Harcourt RG, Holland KN, Iverson SJ, Kocik JF, Mills Flemming JE, Whoriskey FG. Aquatic animal telemetry: A panoramic window into the underwater world. *Science.* 2015;348(6240):2478. <https://doi.org/10.1126/science.1255642>. <https://science.sciencemag.org/content/sci/348/6240/1255642.full.pdf>.
12. van Diggelen F, Enge P. The worlds first gps mooc and worldwide laboratory using smartphones. In: Proceedings of the 28th International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS+ 2015); 2015. p. 361–9.
13. Kays R, Crofoot MC, Jetz W, Wikelski M. Terrestrial animal tracking as an eye on life and planet. *Science.* 2015;348(6240):2478.
14. Andersen KH, Berge T, Gonçalves RJ, Hartvig M, Heuschele J, Hylander S, Jacobsen NS, Lindemann C, Martens EA, Neuheimer AB, et al. Characteristic sizes of life in the oceans, from bacteria to whales. *Ann Rev Mar Sci.* 2016;8:217–41.
15. Krause J, Krause S, Arlinghaus R, Psorakis I, Roberts S, Rutz C. Reality mining of animal social systems. *Trends Ecol Evol.* 2013;28(9):541–51.
16. Jepsen N, Koed A, Thorstad E, Baras E. Surgical implantation of telemetry transmitters in fish: How much have we learned? *Hydrobiologia.* 2002;483:239–48. <https://doi.org/10.1023/A:1021356302311>.
17. Bergé J, Capra H, Pella H, Steig T, Ovidio M, Bultel E, Lamouroux N. Probability of detection and positioning error of a hydro acoustic telemetry system in a fast-flowing river: intrinsic and environmental determinants. *Fish Res.* 2012;125:1–13.
18. James D, Fischer J, Laube J, Spindler M. An accuracy assessment of ultrasonic transmitter locations determined by mobile telemetry in aquatic systems. *Fish Manag Ecol.* 2014;21(5):421–5.
19. Mora C, Tittensor DP, Adl S, Simpson AG, Worm B. How many species are there on earth and in the ocean? *PLoS Biol.* 2011;9(8):1001127.
20. Hawaii Pacific University Oceanic Institute Aqua Facts. <https://www.oceanicinstitute.org/aboutoceans/aquafacts.html>. Accessed 24 Jan 2019.
21. National Oceanic and Atmospheric Administration Oceans & Coasts. <https://www.noaa.gov/oceans-coasts>. Accessed 24 Jan 2019.
22. Ray GC. Coastal-zone biodiversity patterns. *Bioscience.* 1991;41(7):490–8.
23. Food and Agriculture Organization of the United Nations. The state of food and agriculture 1991. 1992. <https://doi.org/10.18356/e3f71a7b-en>.
24. Clark JR. Coastal zone management for the new century. *Ocean Coast Manag.* 1997;37(2):191–216. [https://doi.org/10.1016/S0964-5691\(97\)00052-5](https://doi.org/10.1016/S0964-5691(97)00052-5). Lessons Learned in Integrated Coastal Management.
25. do Sul JAI, Costa MF. The present and future of microplastic pollution in the marine environment. *Environ Pollut.* 2014;185:352–64. <https://doi.org/10.1016/j.envpol.2013.10.036>.
26. Chust G, Ángel Borja, Liria P, Galparsoro I, Marcos M, Caballero A, Castro R. Human impacts overwhelm the effects of sea-level rise on basque coastal habitats (n spain) between 1954 and 2004. *Estuar Coast Shelf Sci.* 2009;84(4):453–62. <https://doi.org/10.1016/j.ecss.2009.07.010>.
27. Friedlander AM. Marine conservation in oceaania: Past, present, and future. *Mar Pollut Bull.* 2018;135:139–49.
28. Berman GJ, Choi DM, Bialek W, Shaevitz JW. Mapping the structure of drosophilid behavior. *bioRxiv.* 2014002873. <https://doi.org/10.1101/002873>.
29. Honegger K, de Bivort B. Stochasticity, individuality and behavior. *Curr Biol.* 2018;28(1):8–12.
30. Hughey LF, Hein AM, Strandburg-Peshkin A, Jensen FH. Challenges and solutions for studying collective animal behaviour in the wild. *Philos Trans R Soc B Biol Sci.* 2018;373(1746):20170005.
31. Raouf V, Tosetto L, Williamson J. Drone-based high-resolution tracking of aquatic vertebrates. *Drones.* 2018;2(4):37.
32. Willis M, Koenig C, Black S, Castaneda A. Archeological 3d mapping: the structure from motion revolution. *J Tex Archeology Hist.* 2016;3:1–36.
33. Barber A, Cosker D, James O, Waine T, Patel R. Camera tracking in visual effects an industry perspective of structure from motion; 2016. p. 45–54. <https://doi.org/10.1145/2947688.2947697>.
34. He K, Gkioxari G, Dollár P, Girshick R. Mask r-cnn. In: Computer Vision (ICCV), 2017 IEEE International Conference On. IEEE; 2017. p. 2980–8.
35. Abdulla W. Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow. GitHub repository. 2017. https://github.com/matterport/Mask_RCNN.
36. Lin T, Maire M, Belongie SJ, Bourdev LD, Girshick RB, Hays J, Perona P, Ramanan D, Dollár P, Zitnick CL. Microsoft COCO: common objects in context. *CoRR.* 2014;abs/1405.0312: 1405.0312.
37. Torrey L, Shavlik J. Transfer learning. In: Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques. IGI Global; 2010. p. 242–264.
38. Kuhn HW. The hungarian method for the assignment problem. *Nav Res Logist Q.* 1955;2(1-2):83–97.
39. Hartley R, Zisserman A. Multiple View Geometry in Computer Vision, 2nd edn. New York: Cambridge University Press; 2003.
40. Westoby MJ, Brasington J, Glasser NF, Hambrey MJ, Reynolds J. 'structure-from-motion' photogrammetry: A low-cost, effective tool for geoscience applications. *Geomorphology.* 2012;179:300–14.
41. Fonstad MA, Dietrich JT, Courville BC, Jensen JL, Carbonneau PE. Topographic structure from motion: a new development in photogrammetric measurement. *Earth Surf Process Landf.* 2013;38(4): 421–30.
42. Linda G, Shapiro CG. Stockman, Computer vision. Upper Saddle River: Prentice Hall; 2001.
43. Bradski G. The OpenCV Library: Dr. Dobb's Journal of Software Tools; 2000.
44. Schönberger JL, Frahm J-M. Structure-from-motion revisited. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016. <https://doi.org/10.1109/cvpr.2016.445>.
45. Schönberger JL, Zheng E, Pollefeys M, Frahm J-M. Pixelwise view selection for unstructured multi-view stereo. In: European Conference on Computer Vision (ECCV); 2016. https://doi.org/10.1007/978-3-319-46487-9_31.
46. Shoemake K. Animating rotation with quaternion curves. In: Proceedings of the 12th Annual Conference on Computer Graphics and Interactive Techniques; 1985. p. 245–54. <https://doi.org/10.1145/325334.325242>.
47. Nührenberg P. multiviewtracks: animal trajectories from multiple-view videos. Zenodo. 2020. <https://doi.org/10.5281/zenodo.3666727>.
48. Lennox RJ, Aarestrup K, Cooke SJ, Cowley PD, Deng ZD, Fisk AT, Harcourt RG, Heupel M, Hinch SG, Holland KN, et al. Envisioning the future of aquatic animal tracking: technology, science, and application. *BioScience.* 2017;67(10):884–96.
49. Kalacska M, Lucanus O, Sousa L, Vieira T, Arroyo-Mora J. Freshwater fish habitat complexity mapping using above and underwater structure-from-motion photogrammetry. *Remote Sens.* 2018;10(12):1912.
50. Figueira W, Ferrari R, Weatherby E, Porter A, Hawes S, Byrne M. Accuracy and precision of habitat structural complexity metrics derived from underwater photogrammetry. *Remote Sens.* 2015;7(12):16883–900.
51. Ward A, Webster M. Sociality: the behaviour of group-living animals. 2016. https://doi.org/10.1007/978-3-319-28585-6_1.
52. Ebersole JP. Niche separation of two damselfish species by aggression and differential microhabitat utilization. *Ecology.* 1985;66(1):14–20. <https://doi.org/10.2307/1941302>. <https://esajournals.onlinelibrary.wiley.com/doi/pdf/10.2307/1941302>.
53. Sturmbauer C, Fuchs C, Harb G, Damm E, Duftner N, Maderbacher M, Koch M, Koblmüller S. Abundance, distribution, and territory areas of rock-dwelling lake tanganyika cichlid fish species. In: Wilke T, Väinölä R, Riedel F, editors. Patterns and Processes of Speciation in Ancient Lakes. Dordrecht: Springer; 2009. p. 57–68.

54. Akkaynak D, Treibitz T. Sea-thru: A method for removing water from underwater images. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2019. p. 1682–91. <https://doi.org/10.1109/cvpr.2019.00178>.
55. Knapitsch A, Park J, Zhou Q-Y, Koltun V. Tanks and temples: Benchmarking large-scale scene reconstruction. *ACM Trans Graph*. 2017;36(4):.
56. Bianco S, Ciocca G, Marelli D. Evaluating the performance of structure from motion pipelines. *J Imaging*. 2018;4(8):98.
57. Romero-Ferrero F, Bergomi MG, Hinz RC, Heras FJ, de Polavieja GG. Idtracker.ai: tracking all individuals in small or large collectives of unmarked animals. *Nat Methods*. 2019;16(2):179–82.
58. Todd JG, Kain JS, de Bivort BL. Systematic exploration of unsupervised methods for mapping behavior. *Phys Biol*. 2017;14(1):015002.
59. Wiltschko AB, Johnson MJ, Iurilli G, Peterson RE, Katon JM, Pashkovski SL, Abaira VE, Adams RP, Datta SR. Mapping sub-second structure in mouse behavior. *Neuron*. 2015;88(6):1121–35.
60. Zuffi S, Kanazawa A, Black MJ. Lions and tigers and bears: Capturing non-rigid, 3d, articulated shape from images. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2018. p. 3955–63. <https://doi.org/10.1109/cvpr.2018.00416>.
61. Robie AA, Seagraves KM, Egnor SR, Branson K. Machine vision methods for analyzing social interactions. *J Exp Biol*. 2017;220(1):25–34.
62. Pereira TD, Aldarondo DE, Willmore L, Kislin M, Wang SS-H, Murthy M, Shaevitz JW. Fast animal pose estimation using deep neural networks. *bioRxiv*. 2018331181. <https://doi.org/10.1101/331181>.
63. Brown AE, de Bivort B. Ethology as a physical science. *Nat Phys*. 2018;14:1. <https://doi.org/10.1038/s41567-018-0093-0>.
64. Jordan LA, Ryan MJ. The sensory ecology of adaptive landscapes. *Biol Lett*. 2015;11(5):20141054.
65. Anderson DJ, Perona P. Toward a science of computational ethology. *Neuron*. 2014;84(1):18–31.

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

