WILDPOSE: A LONG-RANGE 3D WILDLIFE MOTION CAPTURE SYSTEM

A PREPRINT

Naoya Muramatsu
African Robotics Unit
University of Cape Town
Cape Town, 7700, Western Cape, South Africa
mrmnao001@myuct.ac.za

Sangyun Shin
Department of Computer Science
University of Oxford
7 Parks Road, Oxford, OX1 3QG, United Kingdom
sangyun.shin@worc.ox.ac.uk

Qianyi Deng
Department of Computer Science
University of Oxford
7 Parks Road, Oxford, OX1 3QG, United Kingdom
qianyi.deng@wolfson.ox.ac.uk

Andrew Markham
Department of Computer Science
University of Oxford
7 Parks Road, Oxford, OX1 3QG, United Kingdom
andrew.markham@cs.ox.ac.uk

Amir Patel
African Robotics Unit
University of Cape Town
Cape Town, 7700, Western Cape, South Africa
amir.patel@uct.ac.za

February 5, 2024

ABSTRACT

Understanding and monitoring wildlife behavior is crucial in ecology and biomechanics, yet challenging due to the limitations of current methods. To address this issue, we introduce WildPose, a novel long-range motion capture system specifically tailored for free-ranging wildlife observation. This system combines a electronically controllable zoom-lens camera with a solid-state LiDAR to capture both 2D videos and 3D point cloud data, thereby allowing researchers to observe high-fidelity animal morphometrics, behavior and interactions in a completely remote manner. Field trials conducted in Kgalagadi Transfrontier Park have successfully demonstrated WildPose's ability to quantify morphological features of different species, accurately track the 3D movements of a springbok herd over time, and observe the respiratory patterns of a distant lion. By facilitating non-intrusive, long-range 3D data collection, WildPose marks a significant leap forward in ecological and biomechanical studies, offering new possibilities for conservation efforts and animal welfare, and enriching the prospects for interdisciplinary research.

Keywords Motion capture · Field research · Biomechanics · Ethology

1 Introduction

Animal motion data play a pivotal role in multiple research disciplines, but current motion capture systems are largely confined to laboratory settings [1]. To achieve a holistic comprehension of natural animal behavior equivalent to the insights garnered within laboratory settings, the acquisition of high-resolution and precise 3D data becomes imperative. However, studying animals in their natural habitats introduces a unique set of challenges that remain unaddressed.

Existing methodologies for observing free-roaming animals have used active (e.g. GPS-IMU collars) or passive/focal sensing (e.g. camera traps, unmanned aerial vehicles (UAVs), and zoom-camera photography) — refer to Tab. 1 for details. GPS-IMU collars are effective for long-term tracking of animals over large areas and their inertial data is useful in maneuverability analyses [2, 3]. However, a neck-mounted collar lacks the ability to capture fine-grained information like detailed body pose and may disrupt natural animal behavior [4, 5]. Furthermore, researchers need to capture and sedate the animal to fit the collar, a procedure that can be costly and dangerous for both the animal and the researchers. Camera traps offer an inexpensive, non-invasive alternative and have seen significant advancements by using depth camera data [6] and vision-based software for post-processing [7]. Nevertheless, their utility is restricted to high-traffic areas, such as waterholes, making them unsuitable for capturing specific behaviors like hunting or migration. UAVs equipped with cameras could serve as another option [8, 9], but the noise generated when flying within 100 meters of subjects can alter animal behavior [10]. As the most common approach, many researchers, citizen-scientists, and wildlife photographers commonly use zoom-camera photography, proving it as a convenient approach for high-resolution focal observations at long range. Using reference objects in the scene, researchers can measure some morphometric data [11, 12, 13, 14, 15, 16]. However, these methods fall short in capturing detailed 3D data, like surface contours or body asymmetry, mainly due to their limited resolution in rendering intricate spatial information. DeMargerie et al. have developed an interesting application, the zoom-lens stereo-videography system, for 3D tracking of birds [17]. This system is capable of tracking animals at a distance of 100 m with an resolution of 0.1 m. This accuracy is sufficient for tracking gross motion of birds but for fine-grained position resolution required to extract whole-body kinematics, the baseline distance for the disparity view would need to increase substantially.

To address these limitations comprehensively, we introduce WildPose, a novel motion capture system specifically tailored for 3D wildlife observation in natural environments (Fig. 1A). This system meets three critical criteria and provides additional functionalities:

1. 3D Data Acquisition: Vital for modern biomechanics research [18, 19, 20, 21].
2. Long-range Capability: Designed to minimize behavioral influence, allowing data capture from considerable distances.
3. Portability: Easy to transport and set up.

Unlike existing solutions, WildPose features a compact, mobile design optimized for unobtrusive data collection in challenging terrains. As shown in Fig. 1E, existing popular, portable 3D scanners have difficulty observing animals at a distance, but WildPose can acquire detailed 3D information from much longer distances. We validated the system's capabilities through field tests, capturing a diverse range of animals and generating colored point clouds.

This paper further explores the implications of WildPose for wildlife research and identifies potential directions for future enhancements. One notable application is the precise measurement of animal morphometrics, achievable with a precision of less than 50 mm, across a broad range of species — from giraffe to eagles. Precise individual tracking of animals, e.g., within herds, is another application of WildPose. In a case study involving a springbok herd, WildPose captured intricate behaviors such as sparring from a distance of approximately 150 meters. Leveraging the image foundation model, Segment Anything [22], our system employs high-resolution video data for automated individual animal segmentation. Integration with the corresponding 3D points enables precise tracking of individual movements with a typical precision of ± 42 mm at 10 Hz — far exceeding GPS collars [23]. Moreover, as a novel example, we show how WildPose can be used to capture subtle body movements, such as the breathing patterns during sleep.

Through these example applications, we demonstrate that WildPose constitutes a significant step forward in animal behavior research, offering a unique and effective tool for high-quality 3D data acquisition in natural habitats.

Table 1: **Comparison of approaches for recording free-ranging animal behavior.**

	Noninvasive?	Morphometrics	Portability	3D whole-body data
GPS-IMU collar [3]	✗	✗	✓	✗
Camera trap [13, 6]	✓	✓	✗	✓
UAV [8]	✓	✓	✓	✗
Photography [12]	✓	✓	✓	✗
Zoom-lens stereo vision [17]	✓	✓	✗	✗
WildPose	✓	✓	✓	✓

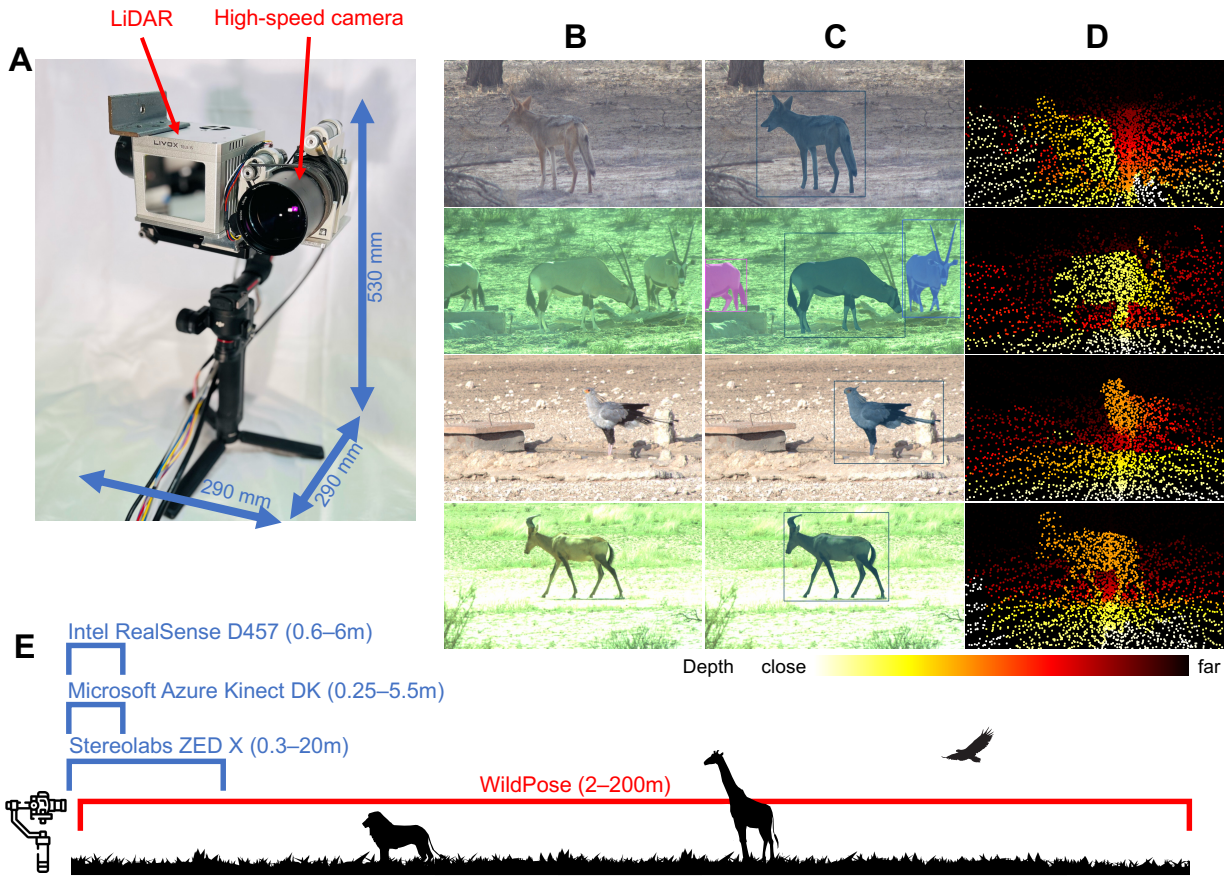

Figure 1: **The WildPose long-range 3D motion capture system.** (**A**), The main hardware components of the WildPose system, which was mounted on a car door in field work. (**B–D**), Examples frames with jackal, gemsbok, secretary bird and red hartbeest. Given a zoom-lens camera image (**B**), the image foundation model [22] outputs segmentation masks of the wild animals with 2D bounding box annotations (**C**). The corresponding point cloud frames provide the depth information (**D**). (**E**), Available range comparison of portable 3D vision modules and WildPose.

2 Results

2.1 Implementation of the WildPose System

WildPose integrates solid-state LiDAR technology with a zoom-lens high-speed camera, both enhanced by a stabilizer for reliable wildlife observation in natural settings (Fig. 1A). This system is distinguished by three core attributes. The long-range sensing capability minimizes human impact on animal behavior, providing a more authentic observational context without disturbance. The use of 3D point-cloud LiDAR ensures the accurate acquisition of metric data — i.e., the ability to capture (x, y, z) points with less than 20 mm precision (`https://www.livoxtech.com/tele-15/specs`). This is indispensable for analyzing inter-individual interactions and making quantifiable, repeatable morphological measurements. Additionally, a rugged design coupled with an active stabilizer makes the system both durable and efficient under natural conditions. We substantiated the system's capabilities through field trials in Kgalagadi Transfrontier Park (South Africa), capturing a wide spectrum of animal data.

LiDAR serves as the premium depth sensor, offering long-range outdoor utility, precise measurements, and system resilience. Due to its repetitive scanning pattern of mechanical spinning LiDAR, e.g., Velodyne (`https://velodynelidar.com/`), the reflected point cloud tends to be sparse for the interest area. To address the sparsity concerns associated with traditional LiDAR systems, we selected a prism-scanning LiDAR from Livox (`https://www.livoxtech.com/`), which provides dense point clouds with a narrow field of view [24]. In particular, the device features a unique scanning mechanism that produces non-repetitive point cloud patterns, improving the richness of the data captured compared with traditional line scanning LiDARs. For this study, we used the TELE-

15 model, which has the narrowest field of view ($14.5° \times 16.2°$) in the product series and a high density of points ($240,000$ points/s). Motors and encoders were affixed to the camera for remote zoom-lens control, and upon calibrating both sensor positions, colored point cloud data (i.e. each (x, y, z) point is associated with a RGB color) is produced. Representative frames of the camera and LiDAR illustrating this are depicted in Fig. 1B, D.

To orchestrate the multitude of sensors and actuators in WildPose, we built the software architecture on ROS2 [25]. However, we substituted the default ROS2 middleware, Real-Time Publish Subscribe Data Distribution Service (RTPS DDS), with eCAL RMW (`https://github.com/eclipse-ecal/rmw_ecal`) to facilitate quicker inter-process communication. This decision was motivated by the inefficiency of RTPS DDS in long-duration data recording, which led to frame loss (see the "eCAL RMW" section). On the other hand, eCAL RMW prioritizes performance and simplicity, even though this comes at the expense of increased storage requirements. This is the key software design to record standard high definition video (resolution of 720×1280) with 170 FPS and LiDAR points with 10 FPS. In summary, the WildPose system accurately records the animal's 3D data even at a long distance, and provides a high-resolution close-up view of the animal's behavior.

2.2 Pipeline

The accurate extraction of animal bodies from a scene, complete with precise 3D coordinates, forms the cornerstone for leveraging both camera and LiDAR data in the WildPose system. We address this segmentation challenge by combining machine learning models with the calibrated 3D sensors capable of accurately detecting animals at a distance.

In the initial phase, 2D bounding boxes are annotated on the first frames of a monocular RGB image. These bounding boxes are subsequently propagated to successive frames using an image-based box tracking method [26], facilitated by DeepLabel [27]. This methodology substantially expedites the annotation process, as it obviates the need for extensive manual labeling beyond the first frame. However, 2D bounding boxes offer only a coarse-grained representation, encapsulating merely the center, width, and height of the object while treating it as a rigid body. Additionally, these boxes often encompass background areas, thereby introducing extraneous data.

To refine this representation, we employ Segment Anything [22], a foundational model for image segmentation, as the next stage in our pipeline. Using the 2D bounding boxes generated in the first phase as input, Segment Anything produces finely-tuned segmentation masks for the predominant object within each box, as illustrated in Fig. 1C.

The final step in the process involves back-projecting these foreground pixels from the segmentation masks to 3D coordinates in the camera's frame of reference. To establish the correspondence between points in the camera coordinates and pixels in the image plane, we first transform points from the LiDAR coordinate system to the camera coordinate system using known extrinsic parameters. These transformed points are subsequently projected onto pixel coordinates through perspective projection, employing known intrinsic parameters. Following this projection, we retain only those pixels that have a direct mapping to the 3D space (Supplementary Video 1).

Through this multi-step process, we successfully obtain accurate 3D segmented masks of animals. This segmentation methodology serves as a foundational approach for various downstream analyses, which we elaborate upon in the subsequent sections. In particular, we highlight three exemplar applications.

2.3 Animal morphology & pose

WildPose's first application aims to radically enhance the remote measurement capabilities for wildlife morphological features, which are invaluable for a wide array of biological investigations. Typical methods to measure the key biological characteristics is to use invasively tranquilized animals which is dangerous and risky for the animals and investigators. Alternatively, post-mortem [28] or museum specimens [29] can be used.

Conventional long-range morphological measurements in the wild primarily utilize digital photogrammetry, broadly classified into two approaches. The first involves a distance meter, where measurements require precise distance determination between the camera and the target, typically using a laser range finder [12, 13, 14]. The second method employs parallel lasers, projecting equidistant laser points onto the target for calibration [15, 16]. Both techniques, however, oversimplify by treating the animal's body as a vertical flat plane, affecting accuracy.

WildPose leverages LiDAR technology to capture metrically accurate point cloud data of animals, with point density increasing if the animal remains stationary. To prove the validation for morphometrics measurement, we measured the 3D lengths of body parts of a giraffe and martial eagle from 2D keypoints (Fig. 2A,B). The measured lengths are cataloged in Tab. 2. Notably, the system's long-range capabilities allow for the precise measurement of a diverse range of species, including smaller ones such as the martial eagle.

Furthermore, the challenge of 3D pose estimation is critical, with various methodologies being proposed in numerous studies [30, 31, 32]. Our system, which merges 3D point cloud data with camera frames, advances the estimation of dynamic wildlife poses in three dimensions by utilizing scene data alongside 2D keypoint labels. Fig. 2C demonstrates a representative result, showcasing the system's ability to capture a lion's movement from standing to walking, accurately capturing joint positions. However, the limitation of a single viewpoint for data capture results in occlusions, preventing the marking of some keypoints in the 3D visualization.

Table 2: **Morphometrics of wild animals.** Each item indicates the measured mean and standard deviation of 17 measurements of the distance between key points. The definitions of keypoint are detailed in the "Measurement of morphometrics" section.

	Martial eagle	Giraffe
Overall height (ground to parietal)	493.9 ± 3.3 mm	5.50 ± 0.02 m
Head length (`nose` to `r_eye`)	56.7 ± 3.5 mm	0.42 ± 0.02 m
Neck length (`r_eye` to `neck`)	123.4 ± 2.9 mm	2.23 ± 0.02 m
Body length (`neck` to `hip`)	530.0 ± 11.2 mm	1.30 ± 0.04 m
Distance from system	18.5 m	90.2 m

2.4 Tracking individual animals in 3D

The ability to track the movements of wild animals in three dimensions is critical for addressing a diverse array of questions in animal ecology, behavior, and cognition. Long-term tracking, spanning days or even longer, provides ecologists with insights into large-scale spatial behaviors such as home ranges, migration patterns, and dispersal mechanisms. The data can be provided from GPS-IMU collar modules. On the other hand, short-term, high-frequency, and high-resolution tracking illuminates localized animal pathways during specific activity phases [3]. This granularity of data is invaluable for investigating an animal's exploratory strategy, orientation skills, and biomechanical interactions with its environment.

The WildPose system surpasses GPS-IMU collar accuracy by integrating LiDAR with zoom-lens cameras and post-processing techniques described in the "Pipeline" section. Fig. 3 exemplifies this capability by demonstrating the tracking of the centers of segmented 3D animal point clouds along with their headings in Bird-Eye-View (top-down) over time. The result achieves an average precision of ± 42 mm, as measured by stationary individuals (ID 4 and 8 from 3.3 to 9.9 seconds in Fig. 3).

This illustration underscores the advantages of the WildPose system in scenarios where traditional approaches may fall short. For instance, the majority of the springboks are seen moving in the negative direction of the x-axis. However, two specific springboks (ID 6 and 7 in Fig. 3) initially engage in sparring and subsequently diverge at different times. Capturing and analyzing such nuanced behaviors with GPS-IMU modules [33] would require sedating and instrumenting several animals.

2.5 Fine scale deformation monitoring

The third example application of WildPose serves as an innovative instrument for the nuanced study of breathing patterns in wildlife, an essential indicator of animal well-being tightly correlated with energetic expenditure across taxa [34]. By leveraging LiDAR technology, WildPose enables a pioneering method for remotely tracking detailed breathing dynamics. Building on the preliminary research that explored non-invasive remote vital data collection using thermal imagers [35], WildPose significantly expands these capabilities, offering unprecedented granularity and accuracy in monitoring both the rate and amplitude of respiration in wildlife by directing measuring the expansion distance of the chest. This is notably demonstrated in the system's ability to capture subtle thoracic motions in a resting lion from 31 meters away, a task traditionally constrained by the requirement for close proximity.

To measure the thorax height of a lion resting on its side, we first estimated the ground plane from the point cloud data. As the lion was situated on a flat road, we aggregated point cloud frames across the scene and employed the RANSAC algorithm to identify the largest plane as the ground [36]. The chest height of the lion was then defined as the average height of 3D points in a frame relative to this identified ground plane. To isolate the lion's breathing frequency amid data noise, a band-pass filter was applied, configured to allow a frequency range of 0.75 to 1.4 Hz.

Fig. 4 and Supplementary Video 2 illustrate the result of this process, as the derived size changes in the lion's body align closely with manual observations. For validation, we manually annotated inhalation and exhalation phases. The filtered breathing pattern not only corroborated these labels but also revealed variations in breathing depth, which were

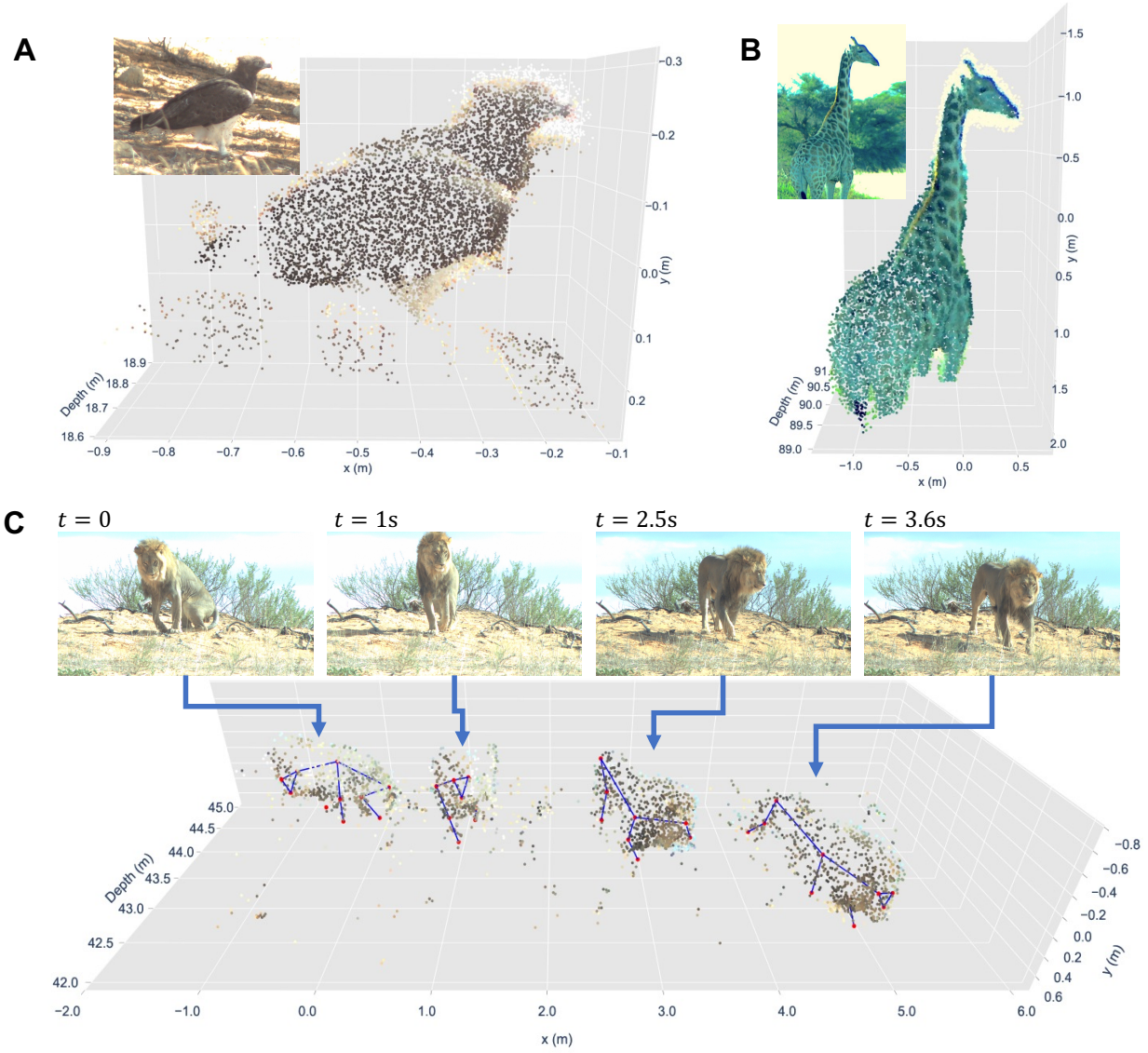

Figure 2: **Colored point cloud of wildlife.** (**A**, **B**), dense point cloud and the corresponding camera frames of static animals (martial eagle and giraffe), where 20 LiDAR frames (2 seconds) were accumulated for each scene. (**C**), point cloud frames of a male lion standing motion, where the red dots indicate the keypoints visible on the frame (`nose`, `eyes`, `shoulder`, `elbow`, `wrist`, `hip`, `knee` and `ankle`) and blue lines show the skeleton. In the 3D plot, the origin of each point cloud frame has 1.5 m margin along the x-axis for the appearance.

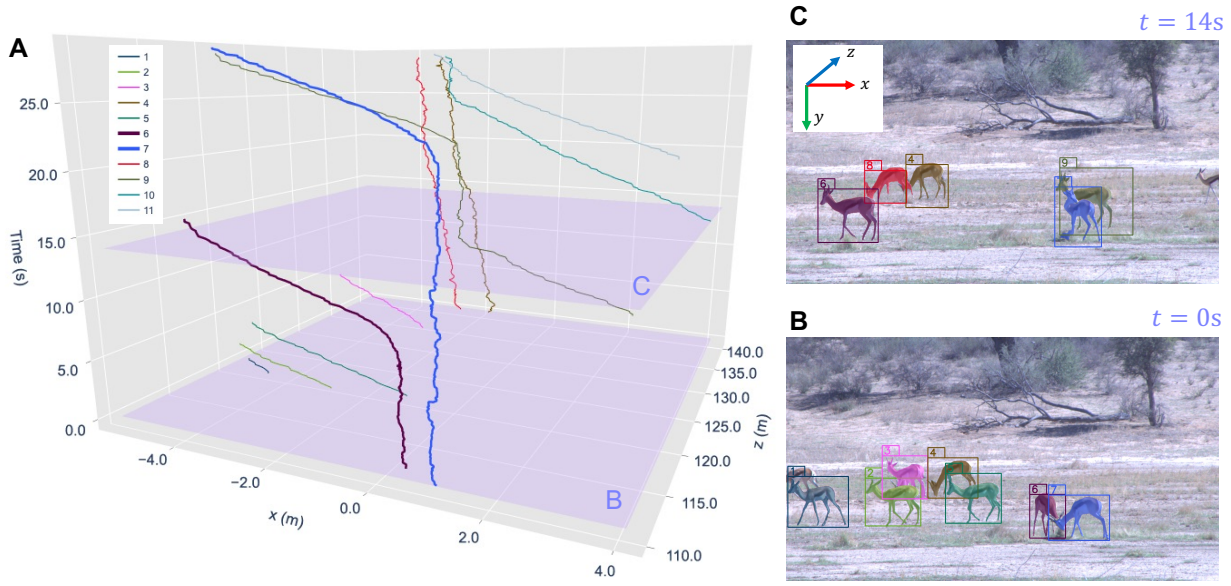

Figure 3: **3D temporal tracking of individual springboks in a herd using the WildPose system.** (**A**), 3D trajectories of the individuals, the z-axis represents the depth information from the camera. (**B**, **C**), the camera frames corresponding with the purple planes in **a**. The x-axis in the 3D plot (**A**) aligns with the x-axis of the images (**B** and **C**), while the z-axis represents the depth information. After the springboks (ID 6 and 7) were sparring in the first time (**B**), ID 6 was the first to leave (**C**).

not readily apparent in the visual data. In the scene, the measured respiration rate was approximately 80 breaths per minute, which is consistent with past observations [37].

An analysis of the video data suggested that the lion's breathing was initially shallow and irregular between 4 and 5 seconds, stabilizing into a more consistent and deep breathing pattern afterward. This observation was supported by the plotted data, which also showed a relatively small thoracic motion in the initial interval, becoming more pronounced and regular subsequently.

3 Discussion

WildPose serves as a long-range 3D motion capture system, enabling researchers to study animals and their behavior within their natural habitats at an unprecedented level. Its capacity for behavioral recording through a portable system unlocks avenues for diverse experiments, from investigating individual interactions within a herd to addressing broader concerns in wildlife health management.

The application examples shown in the previous section distinguish our novel system from existing approaches for wild animal observation. The GPS-IMU module and camera trap have been established as de facto methods for recording wildlife behavior, but approaches to record fine grained animal motion and morphology are still not possible. WildPose overcomes the limitations of existing methods; it can record 3D data from a long distance and it does not require the animal to be instrumented.

The colored 3D data can lead to advanced research work through connecting ethology and computer vision techniques. For example, the colored 3D shape could be used to enhance identification of particular individuals over time, increasing accuracy of existing 2D pelage/spot trackers [38]. Another potential usefulness of the 3D data of wild animals is to provide metric data for 3D modeling and reconstruction. Computer vision researchers have been generating 3D animal models from a 2D image [39, 40], however, the problem is how to evaluate the scale/precise shape of the estimated models. In addition, data of this form could be used to better train monocular-to-3D algorithms by acting as a form of ground-truth, like be done for human pose estimation [41, 42]. WildPose data can thus provide an important repository of baseline data for 2D-to-3D works for wildlife.

Moreover, WildPose contributes to the diversification of animal motion data. Many studies in animal motion analysis have traditionally been conducted within controlled laboratory settings or zoological gardens, often overlooking less-

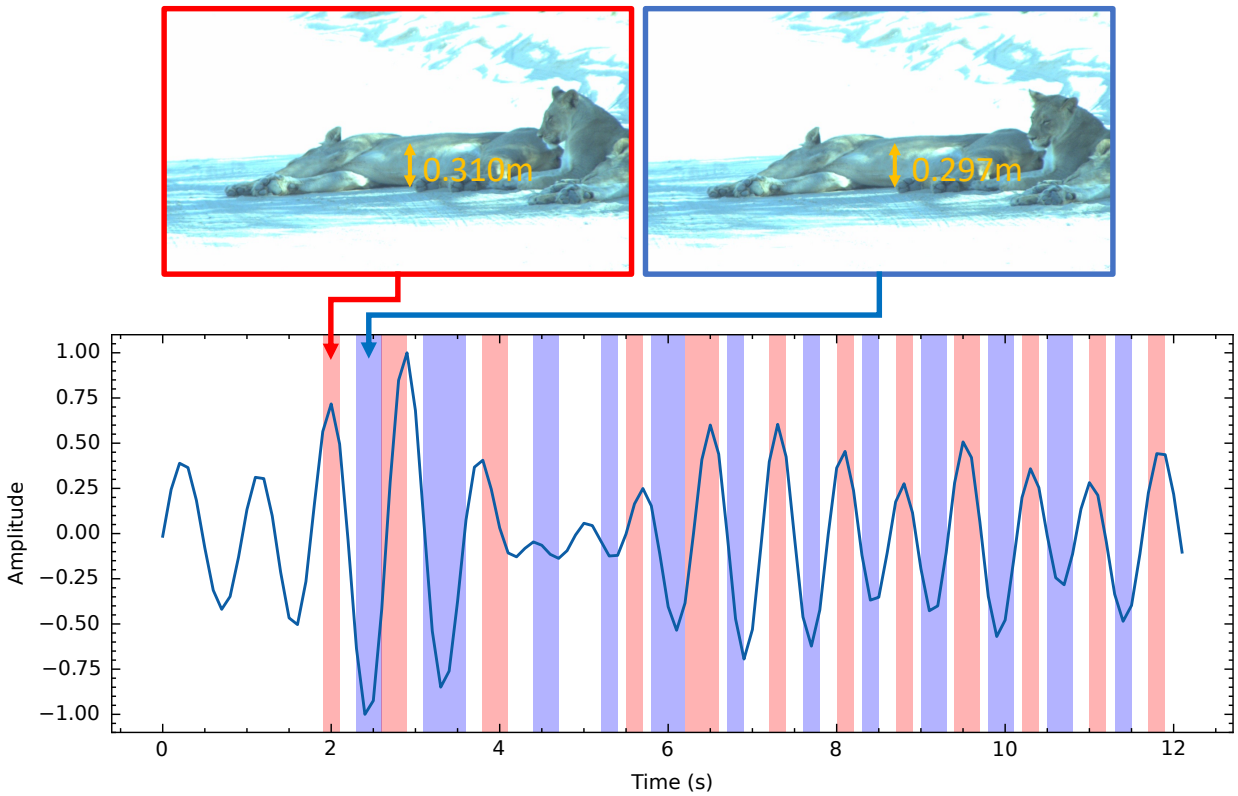

Figure 4: **The respiration pattern of a female lion on a road with video frames.** The colored shadows shows the human labels of inhalation (red) and exhalation (blue). In pictures of the two example frames, the actual height lengths measured from point cloud data are shown.

popular or hard-to-access animals. Additionally, certain natural phenomena, such as springbok herd migrations, also pose challenges when attempting to observe them in laboratory environments. In contrast, we can easily record these scenes with WildPose, allowing precise 3D reconstruction and replay.

We note that camera-equipped UAVs have a great potential for wildlife observation, and these can be potentially combined with WildPose system e.g. through the addition of LiDAR. UAVs interfere with the animals behavior if they are too close (< 50 m) [10], however, a UAV with WildPose would be able to provide 3D data from further distances. We note that the change of viewing angle (nadir) from horizontal to vertical will improve some capabilities (e.g. tracking individual animals) whilst others will be impaired (e.g. monitoring respiration).

While WildPose has such inherent advantages, there is significant room for improvement in the current configuration. In its current iteration, the system relies heavily on manual controls, making data quality contingent on the operator's expertise. The development of an automated tracking system would be especially beneficial for capturing high-speed and expansive behavioral patterns. Although the system faces challenges in measuring the morphometrics of smaller animals at greater distances, it remains proficient in determining the individual 3D positions, as shown in the "Tracking individual animals in 3D" section.

In terms of data acquisition, incorporating multiple camera types could enhance the value of the data captured, as observed in previous motion capture systems designed for free movement [18]. For instance, we encountered scenarios where the narrow field-of-view of a zoom-lens camera was insufficient for capturing behaviors, such as a cheetah spotting a springbok from approximately 30 meters away. The addition of a wide-view camera could mitigate such limitations.

Regarding post-processing, the system offers ample scope for improvements. The LiDAR's laser beams are occasionally obstructed by tall grass in savanna environments, compromising 3D position accuracy near an animal's feet. Advanced algorithms, potentially leveraging machine learning, are needed to offset the limitations inherent to each sensor. While this challenge is difficult to address solely through post-processing —especially when grass may not appear in the video due to the zoom-lens camera's narrow depth of field— it remains a focal point for future research.

Another significant data-processing challenge arises from the disparity between the sensors on WildPose. While the sensor frame rate is constrained by the lowest one in the current process to generate the colored point cloud, advances in multimodal sensor fusion techniques, similar to those used in autonomous driving [43], may offer a potential solution.

The system's robust design, complemented by its active stabilizer and remote connectivity, positions it as a useful tool for capturing genuine behaviors in expansive and natural settings. Its portability also extends the potential for multi-view studies of wild animal behavior. Moreover, WildPose holds promise in the realm of wildlife health monitoring. Algorithms that have proven effective in livestock health monitoring through computer vision and depth cameras in indoor settings [44, 45, 46] could be adapted for use with WildPose.

In summary, WildPose represents a critical technological advancement, integrating LiDAR and zoom-lens cameras to fill the existing void between lab-based motion capture and field study instrumentation. We envision this system pioneering a new frontier in the reliable collection of wildlife motion data, thereby catalyzing future interdisciplinary research efforts.

4 Materials and Methods

4.1 Hardware

As shown in Fig. 5, WildPose consists of stabilized multi-sensor module mounted on a car door during the field work and the control module including the display, gamepad controller, mini-computer and electrical circuits.

The multi-sensor module has three main components: a LiDAR, motorized zoom-lens camera and active gimbal. As the LiDAR, we used Livox TELE-15 (recorded at 10 FPS) for the narrow field-of-view is suitable for combining with zoom-lens camera. A 170 FPS camera (MQ022CG-CM, XIMEA) was used for capturing the video data. The cameras were equipped with zoom lens (Super16 Zoom Camera Lens, 16–160mm FL, NAVITAR), which is controlled by three encoder-attached DC motors (9.7:1 Metal Gearmotor 25D×63L mm HP 12V with 48 CPR Encoder, Pololu) through pulleys. To automatically stabilize the sensor system on a car, we used DJI RS3 Pro, which holds the LiDAR and motorized zoom-lens camera.

The control module consists of two computers. As the main computer recording all the sensor data, controlling the active gimbal and display the visual status of the system, we used NVIDIA Jetson AGX Orin Development Kit with 1TB V-NAND SSD (970 EVO Plus NVMe M.2 2280, Samsung). Teensy 3.2 microcomputer board was used for controlling a motor driver of the zoom lens, correspond to commands from the main computer. The operator can control the zoom-lens parameters (focus, zoom level and aperture), gimbal posture (yaw and pitch), recording states using by a gamepad controller (F710 Wireless Gamepad, Logitech).

4.2 Software

The previously mentioned sensors are seamlessly integrated into a compact and robust platform. This platform operates on an Nvidia Jetson AGX Orin ensuring efficient capturing and synchronization of sensor data with high throughput. The WildPose software, which is built upon ROS2 Foxy, utilized a combination of C++ and Python. For 3D visualization of data points, we harnessed the capabilities of Open3D (`www.open3d.org`). The image sensor system was managed using XiAPI libraries for C++ (`www.ximea.com/support/wiki/apis/XiAPI`). Additionally, a Teensy 3.2 board, programmed with the Arduino IDE, was implemented to manage zoom-lens motor rotations and record encoder steps.

4.2.1 eCAL RMW

Capturing high-speed HD video alongside 240,000 points per second LiDAR data with ROS2 poses a significant challenge in our software system. The standard ROS2 RTPS DDS often grapples with being resource-intensive and exhibits bandwidth limitations and latency issues [47]. To overcome this hurdle, we employed eCAL RMW, an efficient ROS2 middleware developed from the open-source eCAL framework (`https://github.com/eclipse-ecal/rmw_ecal`). This middleware enables streamlined communication among ROS2 components utilizing the eCAL framework. Contrasting the ROS2 RTPS DDS, eCAL RMW underscores simplicity and peak performance. It diverges from adherence to any DDS standard, instead concentrating on fostering efficient data communication, reducing storage and bandwidth requirements, and ensuring low-latency data transfer [48].

For data storage, we opted for the HDF5 file format, due to its enhanced accessibility relative to ROS2 Bags. Serialized data from each sensor is directly streamed into HDF5 files via an eCAL subscriber, allocating one subscriber per sensor topic. Subsequently, this data is deserialized and translated into human-readable formats, such as images and PCD, providing a comprehensive representation of the dataset.

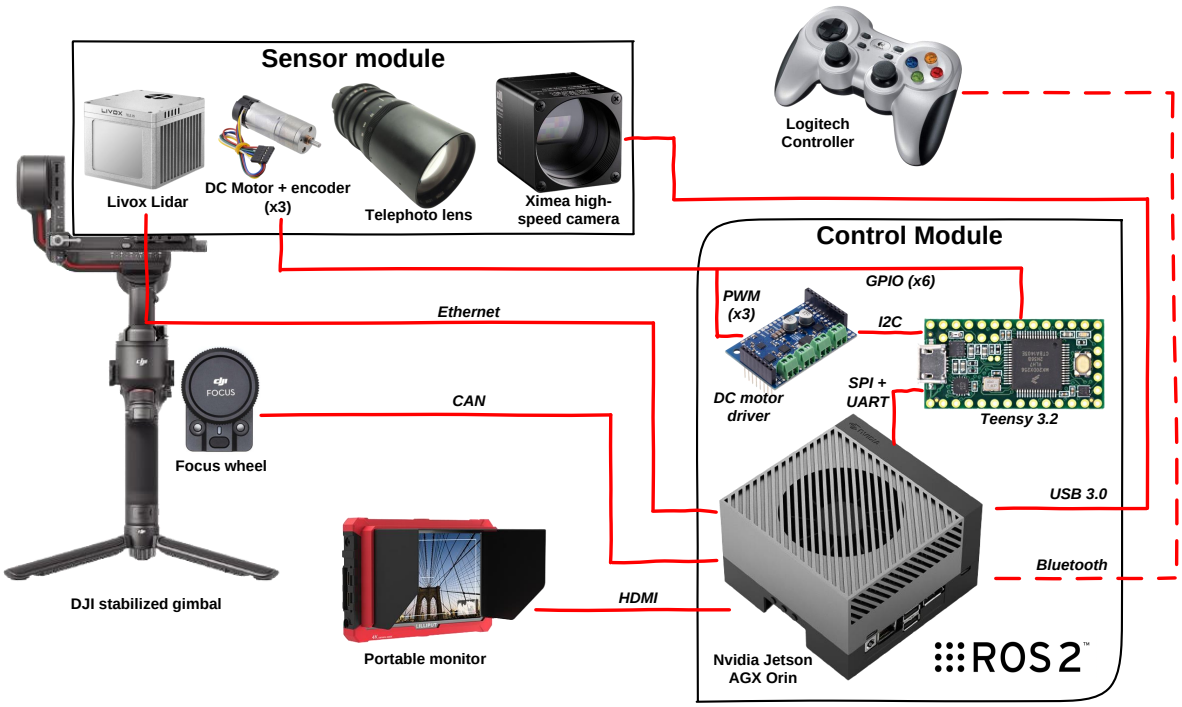

Figure 5: **An overview of the hardware sub-systems which comprise WildPose.** The sensor module contains primary sensors, the solid-state lidar (Livox Tele-15) and the high-speed camera (Ximea MQ022CG-CM), connected to a telephoto lens driven by three brushed DC motors for zoom, focus and aperture. These are housed on a stabilized gimbal (DJI RS3 Pro). The control module contains an NVIDIA Jetson AGX Orin with customized software running on the ROS 2 Foxy and handles the time-synchronization, data logging, motion control of the gimbal as well as real-time display of the video on a portable monitor. The microcontroller (Teensy 3.2) handles the low-level hardware interfacing to the motors.

4.2.2 Synchronization

Synchronization is a critical component for constructing an integrated multi-modal animal dataset. All sensor data in our system are timestamped at every frame. These timestamps are instrumental in identifying the nearest frame across sensors to construct a cohesive multi-modal frame. Notably, the camera video operates at 170 FPS, thereby limiting the maximum error margin with LiDAR data to 3 milliseconds.

4.3 Textured point cloud reconstruction

Given the dynamic nature of camera parameters, we employed a manual calibration approach for intrinsic camera parameters [49]. This process utilized an in-house developed visualization tool that projected the 3D point cloud from LiDAR onto corresponding image frames (`https://github.com/African-Robotics-Unit/wildpose-self-calibrator`). Operators could fine-tune intrinsic parameters by using laser reflectivity and depth cues from the camera.

Utilizing the estimated camera parameters, we achieved the creation of colored point clouds, a result of merging camera and LiDAR data. As illustrated in Fig. 1, animal segmentation was attained by combining camera and LiDAR data with the Segment Anything Model [22]. The process involved identifying corresponding 3D points using camera parameters.

4.4 Data collection in Kgalagadi Transfrontier Park

In December 2022, we embarked on a wildlife recording expedition in Kgalagadi Transfrontier Park, South Africa, using the WildPose system mounted on a vehicle door. Over the course of 13 days, we conducted regular surveys along

fixed routes within the road network surrounding the Two Rivers and Nossob camp sites. Whenever land animals or large birds were encountered, we captured their movements for duration ranging from 30 to 900 seconds. To optimize image quality at a fixed frame rate, we adjusted the zoom lens aperture according to the varying lighting conditions within the park.

4.5 Measurement of morphometrics

To measure of morphometrical data of a martial eagle and giraffe in Tab. 2, we identified and defined four critical keypoints for this analysis: `nose`, `r_eye` (right eye), `neck`, and `hip`. For the martial eagle, the `nose` keypoint corresponds to the tip of its beak. The `r_eye` keypoint is situated at the right eye. The `neck` keypoint is located at the base of the neck. Finally, the `hip` keypoint is identified as the base of the tail in the giraffe and the tip of the tail in the martial eagle.

Conversely, for the scene of a walking lion depicted in Fig. 2C, a more comprehensive set of 15 keypoints was defined, including `nose`, `eye`, `shoulder`, `elbow`, `wrist`, `hip`, `knee`, and `ankle`. These points were initially identified manually on 2D video frames, with their corresponding 3D positions marked on the LiDAR frame. While keypoints other than `nose` were designated on both the right and left sides, occlusions often obscured one side in most frames.

4.6 Static Analysis

4.6.1 Measurement Precision

To evaluate the calibration accuracy of the WildPose system, we measured known object lengths using an 8×11 checkerboard with dimensions of 720×990 mm. Specifically, we assessed the length of three points on the checkerboard as depicted in Fig. 6.

Initially, we calibrated the LiDAR and camera as described in the "Textured point cloud reconstruction" section. Subsequently, we identified two checker pattern points in the image frame and located their corresponding 3D points within the LiDAR point cloud data. We compared these measurements against the reference lengths on the calibration board at varying object distances.

The results, as displayed in Fig. 7, show the percentage error of measured lengths at different distances. Fig. 7 highlights the capability of WildPose to measure large animals at long distances with an error below 20 mm from a distance of 120 meters.

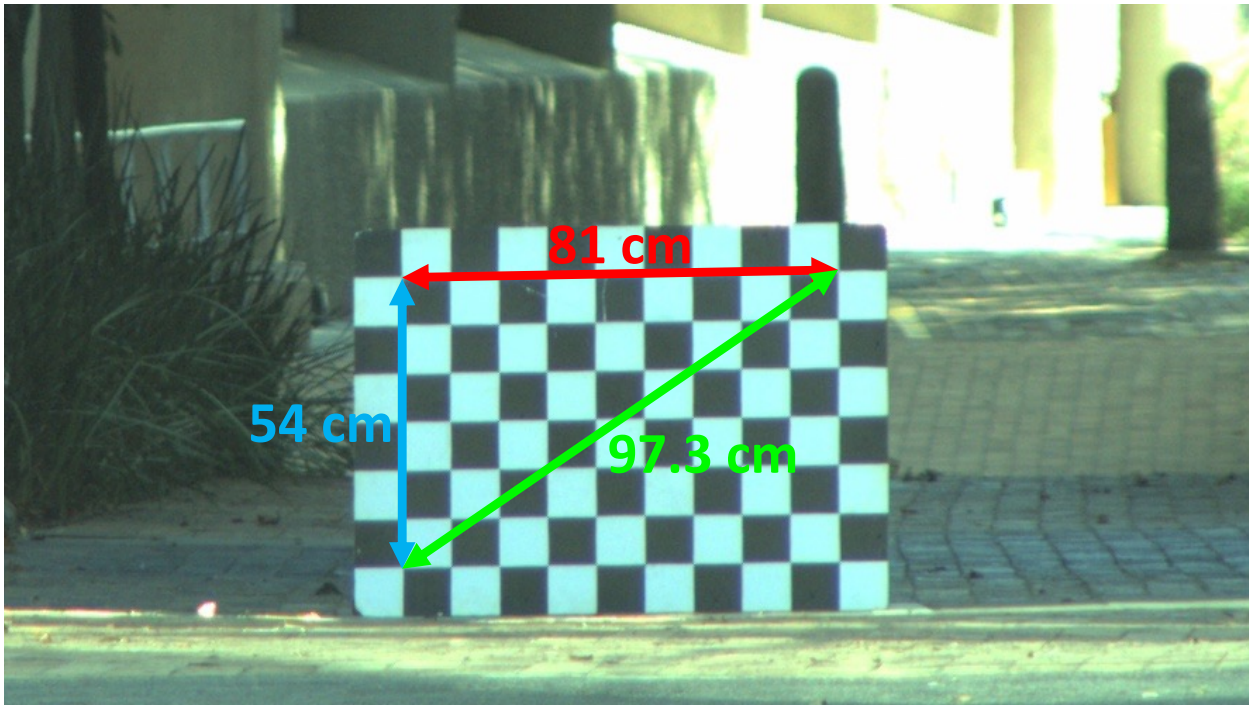

Figure 6: Reference lengths on the checker board.

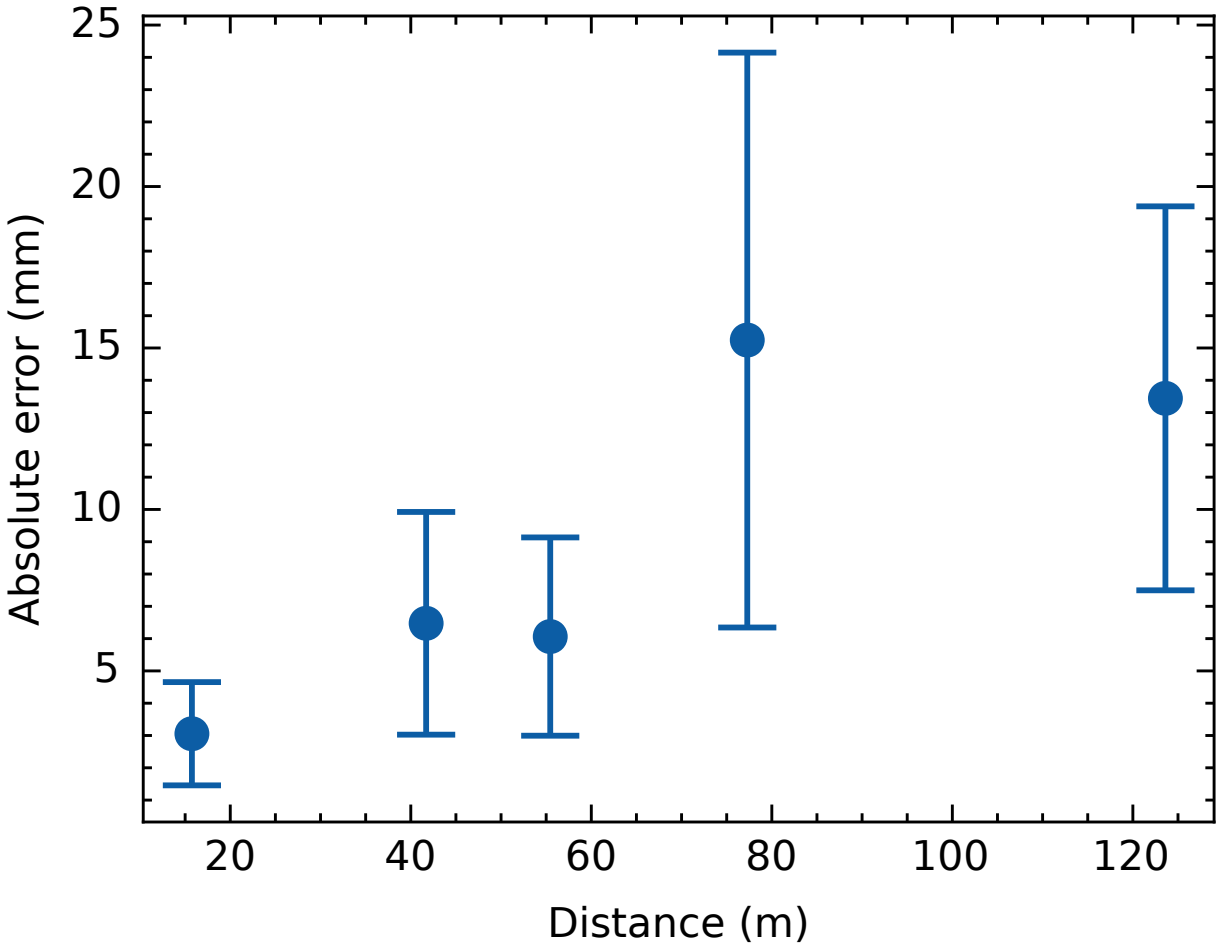

Figure 7: **The validation of WildPose calibration process.** The x-axis denotes the distance to the target object, while the y-axis quantifies the absolute error of the measured lengths as illustrated in 6. Values expressed as mean $\pm$ SD. The graph indicates an upward trend in error corresponding to increased object distance.

4.6.2 Statistics and Reproducibility

Suitable statistical method was used to predetermine the sample size. No data were excluded from the analyses. The experiments were randomized. All values reported as mean$\pm$SD unless otherwise stated. Plot and analyses were performed in Python 3.7.13 (with the aid of numpy 1.21.5 and matplotlib 3.5.3).

Acknowledgments

We thank all people that helped in designing and setting up the hardware and during the fieldwork procedure, especially Nkabeng Mzileni, Correra Links and Michael Katsoulis.

Funding: This work was supported by the Google Research Scholar program (for AP), Oppenheimer Memorial Trust 20881/02 (to AP) and Microsoft Research PhD Fellowship (to NM).

Author contributions: AP and NM conceptualized the system. NM designed and developed the hardware system. NM designed and developed the general reconstruction and analysis framework. SS and NM desgined and wrote the segmentation and tracking software. NM, AP andAM designed all the case studies. NM performed the field work and the experiment and captured all the data. NM analyzed data for all results. NM and SS wrote the original draft of manuscript. NM, SS, QD, AM and AP reviewed and edited the manuscript.

Competing interests: The authors declare no competing interests.

References

[1] M. Nagy, H. Naik, F. Kano, N. V. Carlson, J. C. Koblitz, M. Wikelski, and I. D. Couzin, "SMART-BARN: Scalable multimodal arena for real-time tracking behavior of animals in large numbers," *Science Advances*, vol. 9, p. eadf8068, Sept. 2023.

[2] J. W. Wilson, M. G. L. Mills, R. P. Wilson, G. Peters, M. E. J. Mills, J. R. Speakman, S. M. Durant, N. C. Bennett, N. J. Marks, and M. Scantlebury, "Cheetahs, Acinonyx jubatus, balance turn capacity with pace when chasing prey," *Biology Letters*, vol. 9, p. 20130620, Oct. 2013.

[3] A. M. Wilson, T. Y. Hubel, S. D. Wilshin, J. C. Lowe, M. Lorenc, O. P. Dewhirst, H. L. A. Bartlam-Brooks, R. Diack, E. Bennitt, K. A. Golabek, R. C. Woledge, J. W. McNutt, N. A. Curtin, and T. G. West, "Biomechanics of predator–prey arms race in lion, zebra, cheetah and impala," *Nature*, vol. 554, pp. 183–188, Feb. 2018.

[4] C. Brooks, C. Bonyongo, and S. Harris, "Effects of Global Positioning System Collar Weight on Zebra Behavior and Location Error," *The Journal of Wildlife Management*, vol. 72, no. 2, pp. 527–534, 2008.

[5] C. E. Coughlin and Y. van Heezik, "Weighed down by science: Do collar-mounted devices affect domestic cat behaviour and movement?," *Wildlife Research*, vol. 41, pp. 606–614, Mar. 2015.

[6] M. Klasen and V. Steinhage, "Improving wildlife tracking using 3D information," *Ecological Informatics*, vol. 68, p. 101535, May 2022.

[7] T. D. Pereira, J. W. Shaevitz, and M. Murthy, "Quantifying behavior to understand the brain," *Nature Neuroscience*, vol. 23, pp. 1537–1549, Dec. 2020.

[8] C. K. Basu, F. Deacon, J. R. Hutchinson, and A. M. Wilson, "The running kinematics of free-roaming giraffes, measured using a low cost unmanned aerial vehicle (UAV)," *PeerJ*, vol. 7, p. e6312, Feb. 2019.

[9] B. Koger, A. Deshpande, J. T. Kerby, J. M. Graving, B. R. Costelloe, and I. D. Couzin, "Quantifying the movement, behaviour and environmental context of group-living animals using drones and computer vision," *Journal of Animal Ecology*, vol. 92, no. 7, pp. 1357–1371, 2023.

[10] E. Bennitt, H. L. A. Bartlam-Brooks, T. Y. Hubel, and A. M. Wilson, "Terrestrial mammalian wildlife responses to Unmanned Aerial Systems approaches," *Scientific Reports*, vol. 9, p. 2142, Feb. 2019.

[11] J. Galbany, T. S. Stoinski, D. Abavandimwe, T. Breuer, W. Rutkowski, N. V. Batista, F. Ndagijimana, and S. C. McFarlin, "Validation of two independent photogrammetric techniques for determining body measurements of gorillas," *American Journal of Primatology*, vol. 78, no. 4, pp. 418–431, 2016.

[12] M. B. Brown and E. Wells, "Skeletal dysplasia-like syndromes in wild giraffe," *BMC Research Notes*, vol. 13, p. 569, Dec. 2020.

[13] S. Cui, D. Chen, J. Sun, H. Chu, C. Li, and Z. Jiang, "A simple use of camera traps for photogrammetric estimation of wild animal traits," *Journal of Zoology*, vol. 312, no. 1, pp. 12–20, 2020.

[14] T. Breuer, M. M. Robbins, and C. Boesch, "Using photogrammetry and color scoring to assess sexual dimorphism in wild western gorillas (Gorilla gorilla)," *American Journal of Physical Anthropology*, vol. 134, no. 3, pp. 369–382, 2007.

[15] J. L. Richardson, E. J. Levy, R. Ranjithkumar, H. Yang, E. Monson, A. Cronin, J. Galbany, M. M. Robbins, S. C. Alberts, M. E. Reeves, and S. C. McFarlin, "Automated, high-throughput image calibration for parallel-laser photogrammetry," *Mammalian Biology*, vol. 102, pp. 615–627, June 2022.

[16] J. M. Rothman, C. A. Chapman, D. Twinomugisha, M. D. Wasserman, J. E. Lambert, and T. L. Goldberg, "Measuring physical traits of primates remotely: The use of parallel lasers," *American Journal of Primatology*, vol. 70, no. 12, pp. 1191–1195, 2008.

[17] E. de Margerie, M. Simonneau, J.-P. Caudal, C. Houdelier, and S. Lumineau, "3D tracking of animals in the field using rotational stereo videography," *Journal of Experimental Biology*, vol. 218, pp. 2496–2504, Aug. 2015.

[18] A. Nourizonoz, R. Zimmermann, C. L. A. Ho, S. Pellat, Y. Ormen, C. Prévost-Solié, G. Reymond, F. Pifferi, F. Aujard, A. Herrel, and D. Huber, "EthoLoop: Automated closed-loop neuroethology in naturalistic environments," *Nature Methods*, vol. 17, pp. 1052–1059, Oct. 2020.

[19] N. Muramatsu, Z. da Silva, D. Joska, F. Nicolls, and A. Patel, "Improving 3D Markerless Pose Estimation of Animals in the Wild using Low-Cost Cameras," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 3770–3776, Oct. 2022.

[20] V. Lobato-Rios, S. T. Ramalingasetty, P. G. Özdil, J. Arreguit, A. J. Ijspeert, and P. Ramdya, "NeuroMechFly, a neuromechanical model of adult Drosophila melanogaster," *Nature Methods*, pp. 1–8, May 2022.

[21] X. B. Peng, E. Coumans, T. Zhang, T.-W. E. Lee, J. Tan, and S. Levine, "Learning agile robotic locomotion skills by imitating animals," in *Robotics: Science and Systems*, July 2020.

[22] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, P. Dollár, and R. Girshick, "Segment Anything," Apr. 2023.

[23] A. M. Wilson, J. C. Lowe, K. Roskilly, P. E. Hudson, K. A. Golabek, and J. W. McNutt, "Locomotion dynamics of hunting in wild cheetahs," *Nature*, vol. 498, pp. 185–189, June 2013.

[24] Z. Liu, F. Zhang, and X. Hong, "Low-Cost Retina-Like Robotic Lidars Based on Incommensurable Scanning," *IEEE/ASME Transactions on Mechatronics*, vol. 27, pp. 58–68, Feb. 2022.

[25] S. Macenski, T. Foote, B. Gerkey, C. Lalancette, and W. Woodall, "Robot operating system 2: Design, architecture, and uses in the wild," *Science Robotics*, vol. 7, no. 66, p. eabm6074, 2022.

[26] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," *arXiv:1804.02767 [cs]*, Apr. 2018.

[27] J. Veitch-Michaelis, "Jveitchmichaelis/deeplabel: V0.16.1." Zenodo, Oct. 2021.

[28] L. L. Marker and A. J. Dickman, "Morphology, physical condition, and growth of the cheetah (acinonyx jubatus jubatus)," *Journal of Mammalogy*, vol. 84, pp. 840–850, Aug. 2003.

[29] S. Ravi, D. Kolomenskiy, T. Engels, K. Schneider, C. Wang, J. Sesterhenn, and H. Liu, "Bumblebees minimize control challenges by combining active and passive modes in unsteady winds," *Scientific Reports*, vol. 6, p. 35043, Oct. 2016.

[30] T. Nath, A. Mathis, A. C. Chen, A. Patel, M. Bethge, and M. W. Mathis, "Using DeepLabCut for 3D markerless pose estimation across species and behaviors," *Nature Protocols*, vol. 14, pp. 2152–2176, July 2019.

[31] P. C. Bala, B. R. Eisenreich, S. B. M. Yoo, B. Y. Hayden, H. S. Park, and J. Zimmermann, "Automated markerless pose estimation in freely moving macaques with OpenMonkeyStudio," *Nature Communications*, vol. 11, p. 4560, Sept. 2020.

[32] P. Karashchuk, K. L. Rupp, E. S. Dickinson, S. Walling-Bell, E. Sanders, E. Azim, B. W. Brunton, and J. C. Tuthill, "Anipose: A toolkit for robust markerless 3D pose estimation," *Cell Reports*, vol. 36, p. 109730, Sept. 2021.

[33] S. Grünewälder, F. Broekhuis, D. W. Macdonald, A. M. Wilson, J. W. McNutt, J. Shawe-Taylor, and S. Hailes, "Movement Activity Based Classification of Animal Behaviour with an Application to Data from Cheetah (Acinonyx jubatus)," *PLOS ONE*, vol. 7, p. e49120, Nov. 2012.

[34] N. L. Ramanathan, "Reliability of estimation of metabolic levels from respiratory frequency," *Journal of Applied Physiology*, vol. 19, pp. 497–502, May 1964.

[35] C. L. Rzucidlo, E. Curry, and M. R. Shero, "Non-invasive measurements of respiration and heart rate across wildlife species using Eulerian Video Magnification of infrared thermal imagery," *BMC Biology*, vol. 21, p. 61, Mar. 2023.

[36] P. H. S. Torr and A. Zisserman, "MLESAC: A New Robust Estimator with Application to Estimating Image Geometry," *Computer Vision and Image Understanding*, vol. 78, pp. 138–156, Apr. 2000.

[37] A. Fahlman, A. Loveridge, C. Wenham, C. Foggin, J. M. Arnemo, and G. Nyman, "Reversible anaesthesia of free-ranging lions (Panthera leo) in Zimbabwe," *Journal of the South African Veterinary Association*, vol. 76, pp. 187–192, Dec. 2005.

[38] D. T. Bolger, T. A. Morrison, B. Vance, D. Lee, and H. Farid, "A computer-assisted system for photographic mark–recapture analysis," *Methods in Ecology and Evolution*, vol. 3, no. 5, pp. 813–822, 2012.

[39] M. S. Ahmed, *A Study on 3D Reconstruction of Animals in the Wild*. Honours, Murdoch University, 2020.

[40] M. Badger, Y. Wang, A. Modh, A. Perkes, N. Kolotouros, B. G. Pfrommer, M. F. Schmidt, and K. Daniilidis, "3D Bird Reconstruction: A Dataset, Model, and Shape Recovery from a Single View," in *Computer Vision – ECCV 2020* (A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, eds.), Lecture Notes in Computer Science, (Cham), pp. 1–17, Springer International Publishing, 2020.

[41] C. Ionescu, D. Papava, V. Olaru, and C. Sminchisescu, "Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, pp. 1325–1339, July 2014.

[42] S. Saxena, J. Hur, C. Herrmann, D. Sun, and D. J. Fleet, "Zero-Shot Metric Depth with a Field-of-View Conditioned Diffusion Model," Dec. 2023.

[43] Y. Almalioglu, M. Turan, N. Trigoni, and A. Markham, "Deep learning-based robust positioning for all-weather autonomous driving," *Nature Machine Intelligence*, vol. 4, pp. 749–760, Sept. 2022.

[44] A. F. A. Fernandes, J. R. R. Dórea, and G. J. d. M. Rosa, "Image Analysis and Computer Vision Applications in Animal Sciences: An Overview," *Frontiers in Veterinary Science*, vol. 7, 2020.

[45] L. Huang, S. Li, A. Zhu, X. Fan, C. Zhang, and H. Wang, "Non-Contact Body Measurement for Qinchuan Cattle with LiDAR Sensor," *Sensors*, vol. 18, p. 3014, Sept. 2018.

[46] A. Pezzuolo, D. Giora, H. Guo, Q. Ma, S. Guercini, and F. Marinello, "A comparison of low-cost techniques for three-dimensional animal body measurement in livestock buildings," *IOP Conference Series: Earth and Environmental Science*, vol. 275, p. 012015, May 2019.

[47] T. Kronauer, J. Pohlmann, M. Matthé, T. Smejkal, and G. Fettweis, "Latency Analysis of ROS2 Multi-Node Systems," in *2021 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, pp. 1–7, Sept. 2021.

[48] "eCAL - enhanced Communication Abstraction Layer." https://github.com/eclipse-ecal/ecal, Sept. 2023.

[49] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge, UK ; New York: Cambridge University Press, Mar. 2004.