

Towards Estimation of 3D Poses and Shapes of Animals from Oblique Drone Imagery

Vandita Shukla^{1,2}, Luca Morelli^{1,3}, Fabio Remondino¹, Andrea Micheli⁴, Devis Tuia⁵, Benjamin Risse²

¹3D Optical Metrology Unit (3DOM), Bruno Kessler Foundation (FBK), Trento, Italy

²Computer Vision & Machine Learning Systems Group, Institut Für Geoinformatik, University of Münster, Germany

³Dept. of Civil, Environmental and Mechanical Engineering, University of Trento, Italy

⁴Planning, Scheduling and Optimization Unit (PSO), Bruno Kessler Foundation (FBK), Trento, Italy

⁵Environmental Computational Science and Earth Observation Laboratory (ECEO), EPFL, Switzerland

Email: (vshukla, lmorelli, remondino, amicheli)@fbk.eu, devis.tuia@epfl.ch, b.risse@uni-muenster.de

Commission II - ICWG II/Ia

Keywords: Drones, Wildlife Conservation, 3D Pose Estimation, 3D Shape Estimation, Individual Characteristics, Photogrammetry

ABSTRACT

Wildlife research in both terrestrial and aquatic ecosystems now deploys drone technology for tasks such as monitoring, census counts and habitat analysis. Unlike camera traps, drones offer real-time flexibility for adaptable flight paths and camera views, thus making them ideal for capturing multi-view data on wildlife like zebras or lions. With recent advancements in animals' 3D shape & pose estimation, there is an increasing interest in bringing 3D analysis from ground to sky by means of drones. The paper reports some activities of the EU-funded WildDrone project and performs, for the first time, 3D analyses of animals exploiting oblique drone imagery. Using parametric model fitting, we estimate 3D shape and pose of animals from frames of a monocular RGB video. With the goal of appending metric information to parametric animal models using photogrammetric evidence, we propose a pipeline where we perform a point cloud reconstruction of the scene to scale and localize the animal within the 3D scene. Challenges, planned next steps and future directions are also reported.

1. INTRODUCTION

Drones, also known as Unmanned Aerial Vehicles (UAV) or Systems (UAS), have become an indispensable asset for wildlife conservation research (Wirsing et al., 2022; Tuia et al., 2022; Duffy et al., 2020). Their application is now widespread in ecological studies as they enable upscaled, replicable (Schroeder et al., 2020) and non-invasive acquisition of high-quality data (Fust and Loos, 2020; Jiménez López and Mulero-Pázmány, 2019). They are also more easily available, safer, and more cost-effective than traditional ground-based or aerial data collection methods (Duffy et al., 2020). Drones can help conduct flexible studies with more intricately designed flight paths as compared to remote sensing, thus making room for adding more complexity and nuances in surveys with varying altitudes (Fust and Loos, 2020). There is also the potential to access difficult or inaccessible terrains, and drones can bring to such locations a variety of high-resolution payloads, ranging from visual to environmental sensors, depending on the objectives of the study (Krishnan et al., 2023; Mou et al., 2023; Corcoran et al., 2021). The use of photogrammetry and computer vision on drone-acquired data can boost quantitative products that can be derived from collected data while reducing the processing time and human effort. For instance, Koger et al. (2023) demonstrate how high-altitude top-view RGB videos and photogrammetry could support habitat reconstruction and herd movement estimation, linking group action to the ecological context. Drones are also employed to create approximations of visual field of each animal within its habitat by learning pose and orientation when observed overhead (Schad and Fischer, 2023; Walter and Couzin, 2021). Most drone-based animal studies, particularly the ones involving some level of automation, work with nadir data whereas oblique imagery remains mostly unexplored (Chabot and Bird, 2015). This is due to the added ground distortion and viewing complexity of oblique videography, e.g. for morphometric analyses. Oblique views, however, have potential to be utilized for gaining visual insights that nadir imagery cannot be used for,

especially when studying terrestrial animal characteristics, appearances, or behavior (Shero et al., 2021). Oblique views can capture data on animals that are occluded by surrounding vegetation (Tuia et al., 2022) and contain additional information about shape, coat patterns, gait, and activity such as grazing (Figure 1). A relevant dataset of oblique videos on animals is the KABR dataset (Kholiavchenko et al., 2024): it comprises high-resolution footages of Grevy's zebras, plains zebras and giraffes. Authors have also presented a computer vision-based pipeline for simultaneously focal sampling several terrestrial animals in the scene at the same time, using deep learning for behavior recognition. Focal sampling (Altmann, 1974) refers to the observation of one specific individual animal for a set time duration, done with the purpose of gaining behavioral insights. Monitoring health, movement, behavior, and responses of the animal not only sheds light on the individual but helps form a more refined understanding of the collective behavior (Koger et al., 2023).

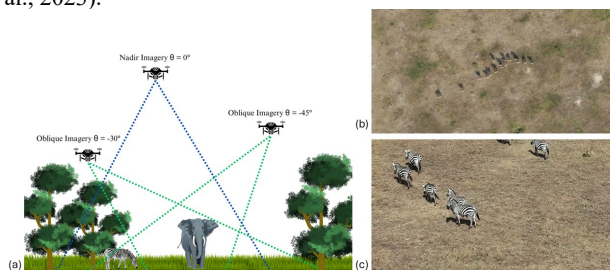


Figure 1: (a) Nadir vs oblique views for animal's surveying. (b) Nadir imagery sees animals only from the top. (c) Oblique views offer better scope for studying individual animal characteristics.

1.1 Paper aims

In this paper, we propose a methodology (Figure 2) to perform a scaled 3D pose and shape extrapolation on individual zebras building upon 3D knowledge of the surveyed scene and the skinned parametric model SMAL (Zuffi et al., 2017). Our work

acts as a first step towards drone-based research on capturing scale in addition to 3D pose and shape of animals using oblique views and photogrammetric knowledge of the scene. Since the parametric model can in principle characterize several species, this will be extendable to oblique drone data of different types of animals. The contributions of our paper include:

- to perform animals' 3D pose and shape estimation based on oblique drone imagery;
- to combine photogrammetrically extracted camera poses and flights logs to geolocate and scale the 3D animal models;
- to bring together parametric fitting and photogrammetry for comprehensive scene recovery.

2. RELATED WORK

2.1 Animal 3D pose and shape estimation

3D shape of animals can provide valuable knowledge on their health, reproductive status, and age (Postma et al., 2015). The 3D postural information allows for several types of kinematic analyses (Tuia et al., 2022). However, it remains an under-explored problem when compared to the analogous challenge of 3D human shape and pose estimation (Xu et al., 2024). This is true because of various reasons such as inter-species shape and appearance diversity, and shortage of datasets. 3D reconstruction through neural representations is progressing rapidly and can be generalized to several species. Some instances of research in this direction include 3D Fauna (Li et al., 2024), LASSIE (Li et al., 2024), MagicPony (Wu et al., 2023), 3D Style Birds (Wang et al., 2023), BANMo (Yang et al., 2022), TAVA (Li et al., 2022), DOVE (Wu et al., 2022) and LASR (Yang et al., 2021). However, the accuracy with which they represent animals, particularly for ecological use-cases, is not yet comparable to statistical fitting methods (Rüegg et al., 2023). Since large-scale 3D digitization operations of animals are not practically achievable. 3D parameterized quadruped models, such as SMAL (Zuffi et al., 2017), were proposed by learning a low-dimensional shape space using scans of toys. The manual process of labelling pose key-points and segmentation mask required for SMAL fitting was supplemented by Biggs et al. (2018) with a deep learning based front-end for 2D joint predictions. SMALR (Zuffi et al., 2018) improves SMAL fitting through incorporation of multi-view imagery to refine the shape and extract texture information. SMALST (Zuffi et al., 2019) uses an end-to-end network to regress SMAL parameters to extract shape, pose and texture from Grevy's zebras' images in the wild. Li and Lee (2021) add a layer of refinement to the SMAL fitting by integrating per-vertex deformation prediction using graph convolutional networks.

Generalization over diverse species is the main challenge for parametric methods. SMAL-like models can be used with high realism for animals belonging to the Felidae, Canidae, Equidae, Bovidae and Hippopotamidae family with a highly specific shape prior; the problem with generalization becomes more pronounced with animals such as elephants and giraffes. Their particularity and precision have, however, been taken advantage of for studies not focusing on generalization but instead where breed priori is exploited, such as SMBLD in Biggs et al. (2018). This is an extension of SMAL for representing dogs with better shape accuracy using augmented shaped parameters. These augmented shape parameters are picked up in BARC (Rüegg et al., 2022), a method incorporating breed-awareness, which has been followed by BITE (Rüegg et al., 2023) that additionally utilizes ground contact information for modelling more realism in dog shape and posture. Kanazawa et al. (2018) deform spherical meshes to birds for extracting shape, texture and pose. The deformation is done, however, without parameterizing the posture information. Therefore Badger et al. (2020) introduced a low-dimensional bird

shape space, further developed by Wang et al. (2021). The most recent advancement has been the Animal3D Dataset (Xu et al., 2024), which is a dataset consisting of over 3000 images of 40 mammal species with annotations of not just 2D pose key points, but also the accompanying pose and shape parameters of high-quality SMAL fitting. Note that none of these methods have been utilized on drone-based imagery, which introduces a new domain of modelling and posture challenges.

From an application standpoint, the use of 3D statistical models can be seen where Stennett et al. (2022) combined deep learning, 3D shape analysis with parametric modelling process SMALST and metric learning for re-identification of individual Grevy's zebra on camera-trap data. The authors highlight how 3D model fitting can improve re-identification results as compared to widely used 2D bounding box methods, but even though they lay foundation for an animal identification system which could be applicable to open population settings, there is still no solution to the two-side problem. The two-side problem refers to how one lateral view of an animal has no identifiable correspondence to the other lateral view and one zebra can be assigned with two separate ids when viewed from different sides, an issue prevalent with camera-trap data. Flight paths designed to acquire a combination of oblique and nadir views, combined with real-time tracking, can be the solution to the two-side problem. Our application of parametric 3D pose analysis to drone data lays the perfect groundwork for testing such re-identification pipelines with drones.

2.2 Drones, Photogrammetry and Computer Vision in Wildlife Conservation

The contribution of drones to wildlife conservation efforts has multiplied with the assimilation of machine learning, computer vision and photogrammetry in both real-time analysis and post-processing. This can be seen across a variety of applications such as census surveys and animal counts (Rahman et al., 2023; Burke et al., 2019; Kellenberger et al., 2019), social and individual behavioral analysis (Jagielski et al., 2022; Hartman et al., 2020; Torney et al., 2018), morphometric analysis (Torney et al., 2018), sample collection (Álvarez-González et al., 2023; Aucone et al., 2023), individual re-identification (Andrew et al., 2019), environment analysis (Koger et al., 2023), large-scale ground truthing of remote-sensing data (Wirsing et al., 2022) and security related applications such as poacher detection (Bhatia et al., 2024; Anbalagan et al., 2023; Doull et al., 2021). This combination has helped produce valuable and layered datasets comprising of quantified metrics on movement and social interactions at high-resolution (Koger et al., 2023; Haalck et al., 2023; Torney et al., 2018). Drones have helped capture the 3D structure of surrounding habitats using photogrammetry, for analyzing patterns in animal grouping and decision making and linking it to spatial knowledge (Maeda and Yamamoto, 2023; Koger et al., 2023). Data fusion of visible and thermal spectrum has been utilized for studying and distinguishing individuals from their environment in drone data (Krishnan et al., 2023).

Most of these instances study ways of disentangling several strands of information, for instance gauging animal movement from drone videos or reconstructing the environmental context around the monitored animal that is continuously engaging with its environment and changing positions. For videos acquired in the wild, photogrammetry can provide metric information of the scene but cannot recover 3D shape and pose of the animals due to their continuous movement and change in posture. Through statistical fitting, 3D shapes and pose of the moving animals in the scene can be recovered but without scale information. Our methodology thus brings together these two 3D analysis methods - photogrammetry and parametric modelling - for a more comprehensive form of terrestrial animals analyses.

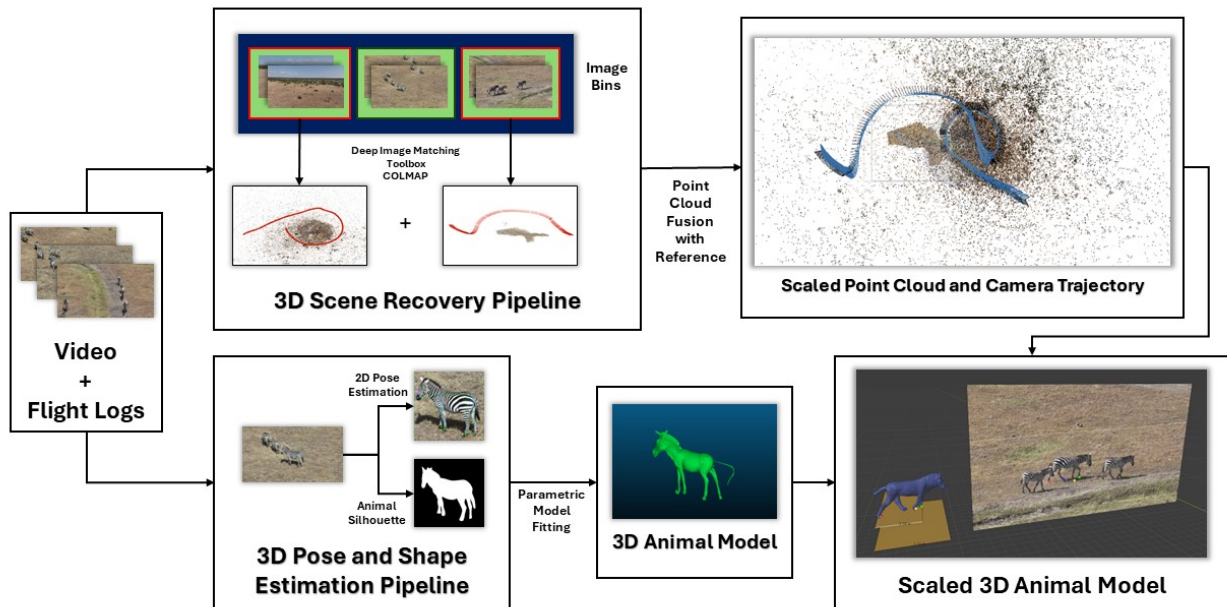


Figure 2: The proposed methodology which includes 3D environment reconstruction as well as 3D shape and pose estimation of animals seen in oblique drone imagery.

3. METHODOLOGY AND RESULTS

3.1 Data

The used data are standard videos collected for manual surveillance at the Ol Pejeta Conservancy, located in Laikipia County in Kenya, using a DJI Mavic 3E (Figure 3). The videos (3840 x 2160 pixels resolution) feature a herd of Plains Zebras (*Equus quagga*) and were captured from relative altitude ranging from 15 to 35 meters above ground. Generally, the flight starts with a circular trajectory around the herd and subsequently proceeds to follow the herd. The zebras can be distinctly observed in their environment and engage in a variety of activities such as grazing and self-grooming. Working with these videos poses a different set of challenges when compared to working with traditional nadir data acquired for wildlife studies, as they feature oblique viewpoints, varying distance from the herd, changing altitude from the ground, zooming effects, continuous movement of animals, etc.



Figure 3: Ol Pejeta Conservancy area as seen in Google Maps (left). UAV views from the WildDrone data acquisition in July 2023 (right).

3.2 Photogrammetric scene reconstruction

For our experiments, we selected a video that showcases all the above-mentioned challenges. Given the continuous movement of both herd and drone, we extracted frames at 12 fps. This allowed us to create a set of images suitable for photogrammetric purposes, while also to correctly determine poses and shape of the zebras. During the drone flight, frequent camera zooms are required to better examine the herd or single animals. Therefore, using the DJI flight logs and frame metadata, the frames were automatically split into bins of similar focal lengths to assist the photogrammetric processing. For the image orientation, the SuperPoint (DeTone et al., 2018) feature extractor and the LightGlue (Lindenberger et al., 2023) feature matcher, both available in the Deep Image Matching toolbox (Morelli et al., 2024)¹, were used. Camera trajectories and sparse 3D reconstruction of the scenes were then recovered in COLMAP (Figure 4a-b). After individual scene recovery (for each bin), the scenes were co-registered in Agisoft Metashape exploiting the GNSS information stored in the flight logs (Figure 4c). This process allowed to create a scaled 3D result of the scene.

3.3 3D pose and shape estimation of zebras

SMAL (Zuffi et al., 2017) is a statistical 3D shape space descriptor mesh model represented as $M(\beta, \theta, \gamma)$, where β is the shape, θ is the pose and γ is the translation. These parameters collectively describe the modulation in shape (or postural representation), thus making them suitable for integration in graphic-based pipelines. Shape β is a descriptor for the coefficients of the low-dimensional shape space of the animal that are learned through Principal Component Analysis. The joints are denoted via a kinematic tree analogous to the skeletal structure, the root of which undergoes the translation γ . Pose θ is described through joints rotation. Starting from the SMAL mesh model, the fitting based SMALR method (Zuffi et al., 2018) is then applied for high accuracy shape retrieval.

Silhouette masks and 2D pose keypoints are extracted from the UAV frames as inputs for the parametric fitting:

¹ <https://github.com/3DOM-FBK/deep-image-matching>

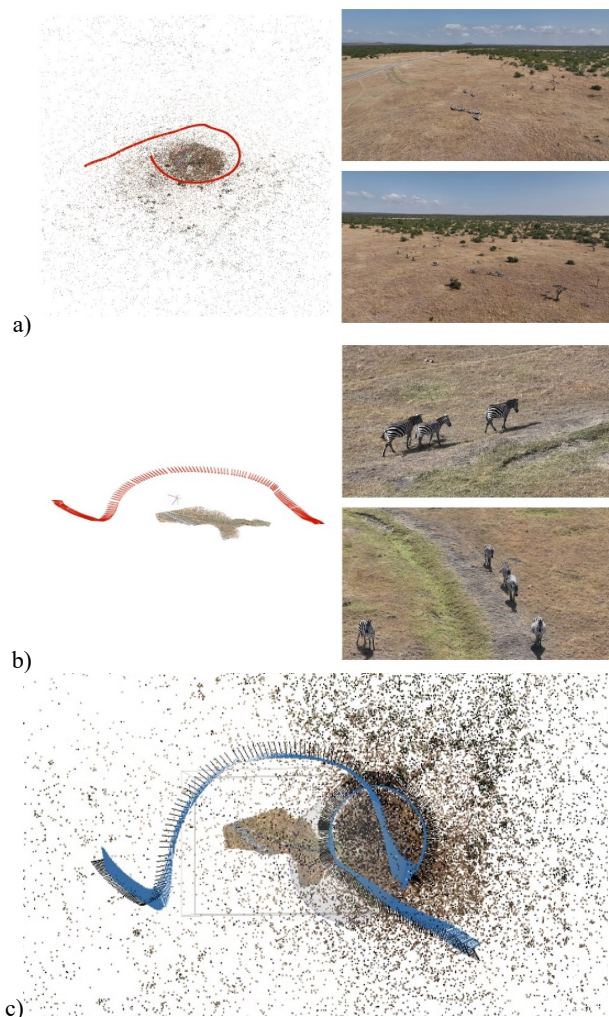


Figure 4: Recovered camera poses and sparse 3D scene from bins of 194 (a) and 227 (b) frames, respectively. Fusion of the separate processing results exploiting GNSS information extracted from the flight logs (c).

- for extracting zebra silhouettes, a MaskRCNN (He et al., 2018) architecture is trained using PyTorch’s Detectron2. Once masks are extracted, the selection of frames for SMALR fitting is performed by calculating the ratio between mask and image size and considering frames with a mask detection accuracy above 80%. For the former, frames with ratio above the median value are considered.
- for 2D pose estimation, the MMPose² toolkit is used, applying transfer learning from a pre-trained HRNet (Wang et al., 2020) backbone. Pose estimation with HRNet calculates the maximum likelihood of body keypoints through a ‘top-down approach’ i.e. body detection followed by joints estimation. In its first stage, the HRNet architecture processes the input image through parallelized network architecture, each branch corresponding to a different resolution scale with the goal to preserve both granular local details and global semantic contextual information. The multi-scale features are aggregated in the pose estimation head of the network through a fusion mechanism - each representation takes input from its immediate neighboring scales and from other parallel branches iteratively. This aggregation method ultimately generates keypoint heatmaps

² <https://github.com/open-mmlab/mmpose>

predicting the joint presence likelihood from the multi-spatial representations (Figure 5).

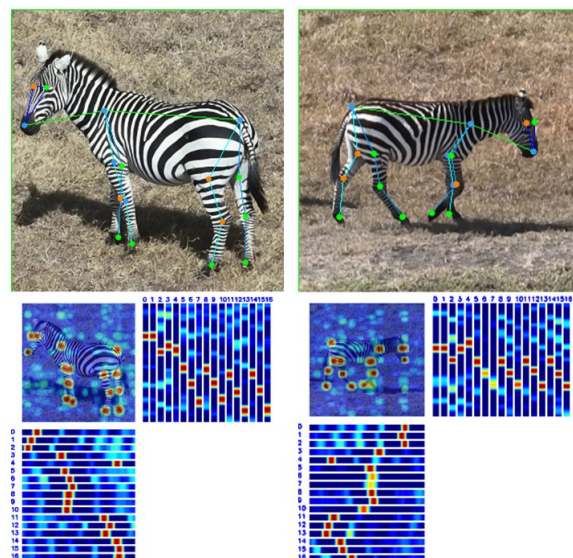


Figure 5: 2D pose estimation with HRNet.

Finally, using the Equidae or Horse family specific shape prior, we input 2D key points and silhouette-based image evidence of the zebra’s shape in the frame as pose and shape target for iterative 3D model fitting optimization process inspired by SMALR fitting method (Figure 6). The 3D model is aligned via error minimization on an objective function which represents both pose key points errors and the silhouette errors.

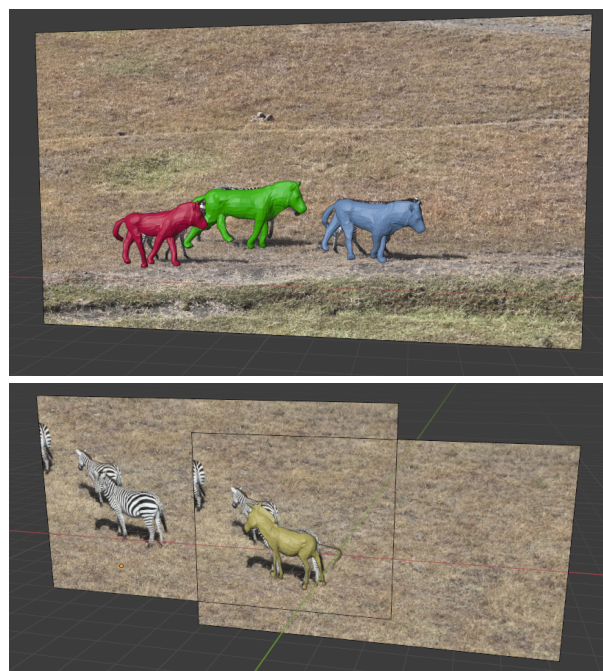


Figure 6: Visualization of SMALR fitting results in Blender.

3.4 Metric parametric model

To maintain the accuracy of the scaling, we take advantage of the extracted 2D and 3D keypoints of the hooves of the focal zebra. We chose hooves as animal reference because they have ground contact, they show good image replication trend when using SMALR and can help gauge orientation. We locate the 2D hoof

keypoints within the image using the already obtained 2D pose estimates and label them as markers across a set of five frames. These points are labelled as H_{rf} (image right front hoof), H_{lf} (image left front hoof), H_{rb} (image right back hoof), and H_{lb} (image left back hoof) in x,y image coordinates. The next step involves initiating a ground plane that denotes the maximum length covered within the x range:

$$\text{dist}2D_X = | \text{distance} (\max_{\text{ext}} (H_{rf}, H_{lf}), \max_{\text{ext}} (H_{rb}, H_{lb})) |$$

where \max_{ext} represents the maximum extension in either front or back to calculate distance from the most extended front hoof to the most extended back hoof either left or right. This distance is chosen because the SMALR model fitting, on rare occasions, can show distortions when differentiating between the left and right limb in both the front and back of the animal model, but, irrespective of the selection of limbs, it mimics extension along the head-to-tail length very accurately. Our goal here is to preserve this length and therefore go with the maximally extended limb in both front and back. We calculate the mean of this distance $\text{dist}2D_{X_{\text{mean}}}$ across a set of five frames to minimize the error. We then perform rigid scaling on the model mesh using the Python library Trimesh³. With the 3D pose estimates from the model hoof keypoints, labelled as H_{RF} (right front hoof: yellow in Figure 7), H_{LF} (left front hoof: green in Figure 7), H_{RB} (right back hoof: red in Figure 7), and H_{LB} (left back hoof: blue in Figure 7), a rigid mesh scaling is performed through the following steps:

- key point distances computation between P_i and the most extended back hoof, which in Figure 7 is H_{RB} . P_i is the x -axis component of the furthest extended front hoof, in this case is H_{LF} , along H_{RB} ;
- scaling factor computation as $\text{dist}2D_{X_{\text{mean}}} / | \text{distance} (P_i, \max_{\text{ext}3D} (H_{RB}, H_{LB})) |$
- rigid transformation application and mesh scaling with the computed factor. In case artefacts are created, a mesh regularization is applied.

After this scaling process, we calculated the distance between nose keypoint and tail-start keypoint using the corresponding 3D pose coordinate mesh vertices. In the sample case shown in Figure 7, the zebra length is 232 cm. In the literature (Kingdon, 1988), the average head-body length of a Plains Zebras can range anywhere between 217-244 cm. Figure 8 shows examples of the fitting results for 3D shape recovery.

4. CONCLUSIONS

In this paper, we perform parametric model fitting on zebras using monocular videos captured by drones in-the-wild.



Figure 8: Fitting results and 3D shape recovery on zebras from multiple drone frames: automatically extracted joints and 3D shape without/with superimposed original image.

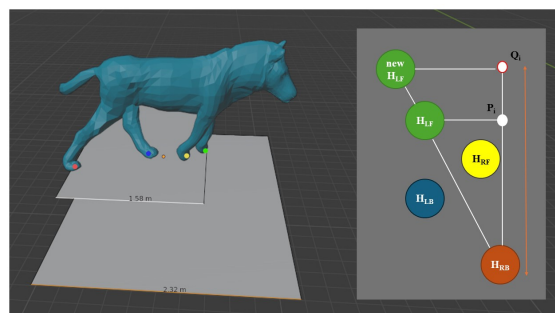


Figure 7: Scaling results visualized in Blender. Visualization of the rigid scaling factor deduction using the properties of similar triangles in the plane of $\Delta H_{LF} P_i H_{RB}$.

The results suggest that combining photogrammetric processing and parametric model fitting to oblique monocular drone footage is an effective technique for quantifying the posture and shape of zebras observed in the wild. At the moment, we do not have ground truth to verify the error in this metric estimation, but such ground truthing will be pursued in further research activities in the framework of WildDrone project. We plan to use toy animals moved around while drones are surveying the area and acquiring multi-view images.

A standard issue, as seen with SMALR fitting, is estimation of shape on dorsal viewpoints and missed gaze direction and these problems were observed with drone footage as well. To address this, we plan to collect aerial multi-view data with simultaneous multi-drone flights to understand the scope of improving fitting to odd poses and dorsal view with multiple perspectives. This was an important reason for choosing SMALR as the parametric model fitting algorithm in this pipeline. Since the model can represent several other species of interest, we will be looking at collecting data for different species as well.

Aerial high-resolution oblique viewpoints open doors for 3D insights in wildlife conservation and introduce opportunities to increase the granularity and depth in pipelines such as pose estimation. The performance of most current methods worsens when they are used outside the domain they were developed for (Jiang et al., 2022) therefore it is highly crucial that drone data start to be considered as an essential domain for these 3D methods to be experimented on and adapted to wildlife monitoring and conservation.

³ <https://trimesh.org/>

ACKNOWLEDGMENTS

The WildDrone project (<https://wilddrone.eu/>) has received funding from the European Union's Horizon Europe research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 101071224. The WildDrone project has also received funding in part from the EPSRC funded Autonomous Drones for Nature Conservation Missions grant (EP/X029077/1). We imported and operated drones in Kenya with permission from the Kenya Civil Aviation Authority (import permits: KCAA/OPS/2117/TIP/01 Vol.1-039, KCAA/OPS/2117/TIP/01 Vol.1-040, KCAA/OPS/2117/TIP/01 Vol.1-041, KCAA/OPS/2117/TIP/01 Vol.1-042, KCAA/OPS/2117/TIP/01 Vol.1-043, KCAA/OPS/2117/TIP/01 Vol.1-044, KCAA/OPS/2117/TIP/01 Vol.1-046; operator certificate number: KCAA/OPS/2117/ROC/10). We conducted research in Kenya with permission from the Kenyan National Commission for Science, Technology and Innovation (License No: NACOSTI/P/23/27796), and in affiliation with the Kenya Wildlife Service (Research Permit No. WRTI-0311-05-23). All research activities were performed with the knowledge and support of management and security staff at Ol Pejeta Conservancy.

REFERENCES

- Altmann, J., 1974. Observational Study of Behavior: Sampling Methods. *Behaviour* 49, 227–266.
- Álvarez-González, M., Suarez-Bregua, P., Pierce, G.J., Saavedra, C., 2023. Unmanned Aerial Vehicles (UAVs) in Marine Mammal Research: A Review of Current Applications and Challenges. *Drones* 7, 667.
- Anbalagan, S., Alhakami, W., Manoharan, A., S, S.G., Nehra, A., Alhakami, H., Raja, G., 2023. TVUB: Thermal Vision-based UAV and Blockchain-aided Poaching Prevention System. Proc. *GLOBECOM*, pp. 1296–1301.
- Andrew, W., Greatwood, C., Burghardt, T., 2019. Aerial Animal Biometrics: Individual Friesian Cattle Recovery and Visual Identification via an Autonomous UAV with Onboard Deep Inference. Proc. *IROS*, pp. 237–243.
- Aucune, E., Kirchgeorg, S., Valentini, A., Pellissier, L., Deiner, K., Mintchev, S., 2023. Drone-assisted collection of environmental DNA from tree branches for biodiversity monitoring. *Science Robotics* 8, eadd5762.
- Badger, M., Wang, Y., Modh, A., Perkes, A., Kolotouros, N., Pfrommer, B.G., Schmidt, M.F., Daniilidis, K., 2020. 3D Bird Reconstruction: A Dataset, Model, and Shape Recovery from a Single View. Proc. *ECCV*, pp. 1–17.
- Biggs, B., Roddick, T., Fitzgibbon, A., Cipolla, R., 2018. Creatures GREAT and SMAL: Recovering the shape and motion of animals from video. Proc. *Asian Conference on Computer Vision*, 2018.
- Burke, C., Rashman, M.F., Longmore, S.N., McAree, O., Glover-Kapfer, P., Ancrenaz, M., Wich, S.A., 2019. Successful observation of orangutans in the wild with thermal-equipped drones. *J. Unmanned Veh. Sys.* 7, 235–257.
- Chabot, D., Bird, D.M., 2015. Wildlife research and management methods in the 21st century: Where do unmanned aircraft fit in? *J. Unmanned Veh. Sys.* 3, 137–155.
- Corcoran, E., Winsen, M., Sudholz, A., Hamilton, G., 2021. Automated detection of wildlife using drones: Synthesis, opportunities and constraints. *Methods in Ecology and Evolution* 12, 1103–1114.
- DeTone, D., Malisiewicz, T., Rabinovich, A., 2018. SuperPoint: Self-Supervised Interest Point Detection and Description. Proc. *CVPRW*, pp. 337–33712.
- Doull, K.E., Chalmers, C., Fergus, P., Longmore, S., Piel, A.K., Wich, S.A., 2021. An Evaluation of the Factors Affecting 'Poacher' Detection with Drones and the Efficacy of Machine-Learning for Detection. *Sensors* 21, 4074.
- Shapiro, A., Anderson, K., Duffy, J., Spina Avino, F., DeBell, L., Glover-Kapfer, P., 2020. *Conservation Technology Series Issue 5: DRONES FOR CONSERVATION*.
- Fust, P., Loos, J., 2023. Increasing the accuracy and efficiency of wildlife census with unmanned aerial vehicles: a simulation study. *Wildl. Res.* 50, 1008–1020.
- Fust, P., Loos, J., 2020. Development perspectives for the application of autonomous, unmanned aerial systems (UASs) in wildlife conservation. *Biological Conservation* 241, 108380.
- Gorkin, R., Adams, K., Berryman, M.J., Aubin, S., Li, W., Davis, A.R., Barthelemy, J., 2020. Sharkeye: Real-Time Autonomous Personal Shark Alerting via Aerial Surveillance. *Drones* 4, 18.
- Haalck, L., Mangan, M., Wystrach, A., Clement, L., Webb, B., & Risse, B. (2023). Cater: Combined animal tracking & environment reconstruction. *Science Advances*, 9(16).
- Hartman, K., van der Harst, P., Vilela, R., 2020. Continuous Focal Group Follows Operated by a Drone Enable Analysis of the Relation Between Sociality and Position in a Group of Male Risso's Dolphins (Grampus griseus). *Front. Mar. Sci.* 7.
- He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask R-CNN, in: 2017 IEEE International Conference on Computer Vision (ICCV). Proc. *ICCV*, pp. 2980–2988.
- Inman, V.L., Leggett, K.E.A., 2022. Hidden Hippos: Using Photogrammetry and Multiple Imputation to Determine the Age, Sex, and Body Condition of an Animal Often Partially Submerged. *Drones* 6, 409.
- Jagielski, P.M., Barnas, A.F., Grant Gilchrist, H., Richardson, E.S., Love, O.P., Semeniuk, C.A.D., 2022. The utility of drones for studying polar bear behaviour in the Canadian Arctic: opportunities and recommendations. *Drone Syst. Appl.* 10, 97–110. h
- Jiménez López, J., Mulero-Pázmány, M., 2019. Drones for Conservation in Protected Areas: Present and Future. *Drones* 3, 10.
- Kanazawa, A., Tulsiani, S., Efros, A.A., Malik, J., 2018. Learning Category-Specific Mesh Reconstruction from Image Collections. Proc. *ECCV*, pp. 386–402.
- Kellenberger, B., Marcos, D., Lobry, S., Tuia, D., 2019. Half a Percent of Labels is Enough: Efficient Animal Detection in UAV Imagery using Deep CNNs and Active Learning. *IEEE Trans. Geosci. Remote Sensing* 57, 9524–9533.
- Kholiavchenko, M., Kline, J., Ramirez, M., Stevens, S., Sheets, A., Babu, R., Banerji, N., Campolongo, E., Thompson, M., Tiel,

- N.V., Miliko, J., Bessa, E., Duporge, I., Berger-Wolf, T., Rubenstein, D., Stewart, C., 2024. KABR: In-Situ Dataset for Kenyan Animal Behavior Recognition from Drone Videos. *Proc. IEEE WACVW*, pp. 31–40.
- Kingdon, J., 1988. *East African Mammals: An Atlas of Evolution in Africa*, Volume 3, Part D: Bovids, East African Mammals. University of Chicago Press, Chicago, IL.
- Koger, B., Deshpande, A., Kerby, J.T., Graving, J.M., Costelloe, B.R., Couzin, I.D., 2023. Quantifying the movement, behaviour and environmental context of group-living animals using drones and computer vision. *Journal of Animal Ecology* 92, 1357–1371.
- Krishnan, B.S., Jones, L.R., Elmore, J.A., Samiappan, S., Evans, K.O., Pfeiffer, M.B., Blackwell, B.F., Iglay, R.B., 2023. Fusion of visible and thermal images improves automated detection and classification of animals for drone surveys. *Sci Rep* 13, 10385.
- Li, C., Lee, G.H., 2021. Coarse-to-fine Animal Pose and Shape Estimation, in: *Advances in Neural Information Processing Systems*. Curran Associates, Inc., pp. 11757–11768.
- Li, R., Tanke, J., Vo, M., Zollhöfer, M., Gall, J., Kanazawa, A., Lassner, C., 2022. TAVA: Template-free Animatable Volumetric Actors. *Proc. ECCV*, pp. 419–436.
- Li, X., Huang, H., Savkin, A.V., 2022. Autonomous Navigation of an Aerial Drone to Observe a Group of Wild Animals With Reduced Visual Disturbance. *IEEE Systems Journal* 16, 3339–3348.
- Li, Z., Litvak, D., Li, R., Zhang, Y., Jakab, T., Rupprecht, C., Wu, S., Vedaldi, A., Wu, J., 2024. Learning the 3D Fauna of the Web. *Proc. CVPR*.
- Lindenberger, P., Sarlin, P.-E., Pollefeys, M., 2023. LightGlue: Local Feature Matching at Light Speed. *Proc. ICCV*, pp. 17581–17592.
- Maeda, T., Yamamoto, S., 2023. Drone Observation for the Quantitative Study of Complex Multilevel Societies. *Animals*, 13, 1911.
- Morelli, L., Ioli, F., Maiwald, F., Mazzacca, G., Menna, F., Remondino, F., 2024. Deep-image-matching: a toolbox for multiview image matching of complex scenarios. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLVIII-2/W4-2024*, 309–316.
- Mou, C., Liu, T., Zhu, C., Cui, X., 2023. WAID: A Large-Scale Dataset for Wildlife Detection with Drones. *Applied Sciences* 13, 10397.
- Postma, M., Tordiffe, A.S.W., Hofmeyr, M.S., Reisinger, R.R., Bester, L.C., Buss, P.E., de Bruyn, P.J.N., 2015. Terrestrial mammal three-dimensional photogrammetry: multispecies mass estimation. *Ecosphere* 6, 1–16.
- Rahman, D.A., Herliansyah, R., Subhan, B., Hutasoit, D., Imron, M.A., Kurniawan, D.B., Sriyanto, T., Wijayanto, R.D., Fikriansyah, M.H., Siregar, A.F., Santoso, N., 2023. The first use of a photogrammetry drone to estimate population abundance and predict age structure of threatened Sumatran elephants. *Sci Rep* 13, 21311.
- Rueegg, N., Zuffi, S., Schindler, K., Black, M.J., 2023. BARC: Breed-Augmented Regression Using Classification for 3D Dog Reconstruction from Images. *Int J Comput Vis* 131, 1964–1979.
- Rüegg, N., Tripathi, S., Schindler, K., Black, M.J., Zuffi, S., 2023. BITE: Beyond Priors for Improved Three-D Dog Pose Estimation. *Proc. CVPR*, pp. 8867–8876.
- Schad, L., Fischer, J., 2023. Opportunities and risks in the use of drones for studying animal behaviour. *Methods in Ecology and Evolution* 14, 1864–1872.
- Schroeder, N.M., Panebianco, A., Gonzalez Musso, R., Carmanchahi, P., 2020. An experimental approach to evaluate the potential of drones in terrestrial mammal research: a gregarious ungulate as a study model. *Royal Society Open Science* 7, 191482.
- Shero, M.R., Dale, J., Seymour, A.C., Hammill, M.O., Mosnier, A., Mongrain, S., Johnston, D.W., 2021. Tracking wildlife energy dynamics with unoccupied aircraft systems and three-dimensional photogrammetry. *Methods in Ecology and Evolution* 12, 2458–2472.
- Stennett, M., Rubenstein, D.I., Burghardt, T., 2022. Towards Individual Grevy's Zebra Identification via Deep 3D Fitting and Metric Learning. *Proc. IEEE/IAPR Int. Conference VAIB*.
- Torney, C.J., Lamont, M., Debell, L., Angohiatok, R.J., Leclerc, L.-M., Berdahl, A.M., 2018. Inferring the rules of social interaction in migrating caribou. *Philosophical Transactions of the Royal Society B: Biological Sciences* 373, 20170385.
- Tuia, D., Kellenberger, B., Beery, S., Costelloe, B.R., Zuffi, S., Risse, B., Mathis, A., Mathis, M.W., van Langevelde, F., Burghardt, T., Kays, R., Klinck, H., Wikelski, M., Couzin, I.D., van Horn, G., Crofoot, M.C., Stewart, C.V., Berger-Wolf, T., 2022. Perspectives in machine learning for wildlife conservation. *Nat Commun* 13, 792.
- Walker, S.E., Sheaves, M., Waltham, N.J., 2023. Barriers to Using UAVs in Conservation and Environmental Management: A Systematic Review. *Environmental Management* 71, 1052–1064.
- Walter, T., Couzin, I.D., 2021. TRex, a fast multi-animal tracking system with markerless identification, and 2D estimation of posture and visual fields. *eLife* 10, e64000.
- Sun, K., Xiao, B., Liu, D., Wang, J., 2019. Deep High-Resolution Representation Learning for Human Pose Estimation. *Proc. CVPR*, pp. 5686–5696.
- Wang, R., Que, G., Chen, S., Li, X., Li, J., Yang, J., 2023. Creative Birds: Self-Supervised Single-View 3D Style Transfer. *Proc. ICCV*, pp. 8741–8750.
- Wang, Y., Kolotouros, N., Daniilidis, K., Badger, M., 2021. Birds of a Feather: Capturing Avian Shape Models from Images. *Proc. CVPR*, pp. 14734–14744.
- Wirsing, A.J., Johnston, A.N., Kiszka, J.J., 2022. Foreword to the Special Issue on 'The rapidly expanding role of drones as a tool for wildlife research.' *Wildl. Res.* 49, i–v.
- Wu, S., Li, R., Jakab, T., Rupprecht, C., Vedaldi, A., 2023. MagicPony: Learning Articulated 3D Animals in the Wild. *Proc. CVPR*, pp. 8792–8802.

- Wu, S., Jakob, T., Rupprecht, C., Vedaldi, A., 2023. DOVE: Learning Deformable 3D Objects by Watching Videos. *International Journal of Computer Vision* 131.
- Xu, J., Zhang, Y., Peng, J., Ma, W., Jesslen, A., Ji, P., Hu, Q., Zhang, J., Liu, Q., Wang, J., Ji, W., Wang, C., Yuan, X., Kaushik, P., Zhang, G., Liu, J., Xie, Y., Cui, Y., Yuille, A., Kortylewski, A., 2023. Animal3D: A Comprehensive Dataset of 3D Animal Pose and Shape. Proc. *ICCV*, pp. 9065–9075.
- Yang, G., Sun, D., Jampani, V., Vlasic, D., Cole, F., Chang, H., Ramanan, D., Freeman, W.T., Liu, C., 2021. LASR: Learning Articulated Shape Reconstruction from a Monocular Video. Proc. *CVPR*, pp. 15975–15984.
- Yang, G., Vo, M., Neverova, N., Ramanan, D., Vedaldi, A., Joo, H., 2022. BANMo: Building Animatable 3D Neural Models From Many Casual Videos. Proc. *CVPR*, pp. 2863–2873.
- Yao, C.-H., Hung, W.-C., Li, Y., Rubinstein, M., Yang, M.-H., Jampani, V., 2022. LASSIE: Learning Articulated Shapes from Sparse Image Ensemble via 3D Part Discovery. Presented at the *Advances in Neural Information Processing Systems*.
- Zuffi, S., Kanazawa, A., Berger-Wolf, T., Black, M., 2019. Three-D Safari: Learning to Estimate Zebra Pose, Shape, and Texture From Images "In the Wild". Proc. *ICCV*, pp. 5358–5367.
- Zuffi, S., Kanazawa, A., Black, M.J., 2018. Lions and Tigers and Bears: Capturing Non-rigid, 3D, Articulated Shape from Images. Proc. *CVPR*, pp. 3955–3963.
- Zuffi, S., Kanazawa, A., Jacobs, D.W., Black, M.J., 2017. 3D Menagerie: Modeling the 3D Shape and Pose of Animals. Proc. *CVPR*, pp. 5524–5532.