

Integrating Biological Data into Autonomous Remote Sensing Systems for In Situ Imageomics: A Case Study for Kenyan Animal Behavior Sensing with Unmanned Aerial Vehicles (UAVs)

Jenna M. Kline¹, Maksim Kholiavchenko², Otto Brookes³, Tanya Berger-Wolf¹, Charles V. Stewart², Christopher Stewart¹,

¹The Ohio State University

²Rensselaer Polytechnic Institute

³University of Bristol

kline.377@buckeyemail.osu.edu

Abstract

In situ imageomics leverages machine learning techniques to infer biological traits from images collected in the field, or in situ, to study individuals organisms, groups of wildlife, and whole ecosystems. Such datasets provide real-time social and environmental context to inferred biological traits, which can enable new, data-driven conservation and ecosystem management. The development of machine learning techniques to extract biological traits from images are impeded by the volume and quality data required to train these models. Autonomous, unmanned aerial vehicles (UAVs), are well suited to collect in situ imageomics data as they can traverse remote terrain quickly to collect large volumes of data with greater consistency and reliability compared to manually piloted UAV missions. However, little guidance exists on optimizing autonomous UAV missions for the purposes of remote sensing for conservation and biodiversity monitoring. The UAV video dataset curated by *KABR: In-Situ Dataset for Kenyan Animal Behavior Recognition from Drone Videos* (Kholiavchenko et al. 2024) required three weeks to collect, a time-consuming and expensive endeavor. Our analysis of *KABR* revealed that a third of the videos gathered were unusable for the purposes of inferring wildlife behavior. We analyzed the flight telemetry data from portions of UAV videos that were usable for inferring wildlife behavior, and demonstrate how these insights can be integrated into an autonomous remote sensing system to track wildlife in real time. Our autonomous remote sensing system optimizes the UAV's actions to increase the yield of usable data, and matches the flight path of an expert pilot with an 87% accuracy rate, representing an 18.2% improvement in accuracy over previously proposed methods.

Introduction

Imageomics is a new scientific field that aims to understand and analyze biological traits of organisms, species, populations, and ecosystems from images (Imageomics 2023). In situ imageomics focuses on studying biological organisms and systems within the context of their natural environment. Sources of such datasets include photos and videos captured by experts using camera traps (Brookes et al. 2023; Zualkeran et al. 2022), UAVs (Koger et al. 2023; Andrew, Greenwood, and Burghardt 2019), and citizen scientists, such as

Copyright © 2024, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

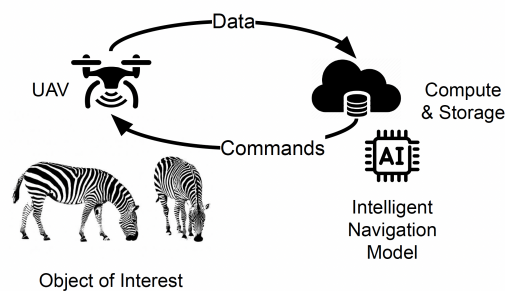


Figure 1: Autonomous Remote Sensing System for In Situ Imageomics

iNaturalist (Horn et al. 2018). As wildlife respond to changing environmental conditions due to climate change, new approaches to remote sensing data collection and analysis are required to develop data-driven approaches to conservation and ecosystem management. Technological advancements in remote sensing data collection technology, namely camera traps (Brookes et al. 2023; Tabak et al. 2019) and UAVs (Kholiavchenko et al. 2024; Koger et al. 2023), have made it easier to gather large volumes of biological data more quickly than traditional manual techniques. With an imageomics approach, biological insights can be extracted from these datasets at scale using machine learning techniques. However, these models require massive volumes of quality data far exceeding what it typically captured for biological studies. This is particularly vital for models built to accomplish specific tasks, such as inferring behavior with the environmental and social context, which is difficult to infer from camera traps or data gathered by citizen scientists.

UAVs are well suited to gather the volume of data required to train such machine learning models for specific tasks. UAVs can dynamically track wildlife to capture fine-grained details, such as behavior, while traversing remote terrain quickly with little disturbance to the surrounding landscape. Experts pilots are required to conduct UAV missions tailored to the specific species and region. The ability to conduct such specialized missions with autonomous UAVs would significantly reduce the barriers to conduct this re-

search. Previous works have demonstrated that autonomous UAV missions are 3x cheaper, safer, more consistent, and reliable than manual flights (Boubin et al. 2019; Kline et al. 2023).

Existing autonomous UAVs methods are not optimized for in situ imageomics data collection. Autonomous UAVs may lose track of individuals over large distances, or fail to capture fine-grain details such as social dynamics. For both autonomous and manual approaches, crucial time may be wasted on collecting and storing images with poor lighting and viewing angles. Such datasets require more storage and transport costs, and require more manual pre-processing before they can be analyzed. Our analysis of a three-day subset of the *KABR* dataset (Kholiavchenko et al. 2024) revealed that only 2.2 hours of video was captured, and of that, only 65% was usable for the mission objective of inferring wildlife behavior. UAVs have only recently become inexpensive enough for widespread use in biology applications, so little guidance exists on optimal strategies for maximizing the usable data yielded from such missions.

In this work, we analyze the flight telemetry from portions of videos successfully used to extract behavior data in *KABR*. We demonstrate how these insights can be integrated into a remote sensing system to improve the performance of autonomous UAVs to gather videos of wildlife behavior, with Kenyan animal behavior as a use-case. Figure 1 illustrates such an autonomous remote sensing system for in situ imageomics. We show that dynamically adjusting the altitude and minimizing the movement of the UAV while ensuring that the wildlife are in view leads to higher usability rate of the collected videos.

Methodology

We analyzed the flight telemetry data from portions of the UAV videos successfully used to extract behavior data from *KABR: In-Situ Dataset for Kenyan Animal Behavior Recognition from Drone Videos* from Kholiavchenko et al. (2024). The videos were collected at the Mpala Research Centre in Laikipia, Kenya using a DJI Air 2S in 4K resolution over a period of three weeks in January 2023. The dataset includes three species: Plains zebra (*Equus quagga*), Grevy’s zebra (*Equus grevyi*), and reticulated giraffes (*Giraffa reticulata*). The *KABR* dataset includes the bounding box coordinates for each animal in the frame, the behavior annotations for each bounding box in each frame of the video, and the frame number and timestamp information. Since we were only interested in frames that contained usable behavior data, we excluded the subset of frames containing behavior annotations of “Occluded” “Out of Frame” or “Out of Focus” from our count of usable frames. Next, we obtained the telemetry data produced by the manually flown UAV missions. The telemetry data for each mission began at the time the UAV was powered on and launched and ended once the UAV was returned to the pilot and powered off. The telemetry data contains data gathered during the UAV flight which describes the status of the system, including altitude, speed, direction, and battery level. We used the frame number and timestamp information to reconcile the behavioral annotations with the corresponding telemetry data, as shown in Table 1.

Distribution of Speed, Altitude, Length, and Width Values from Telemetry Analysis

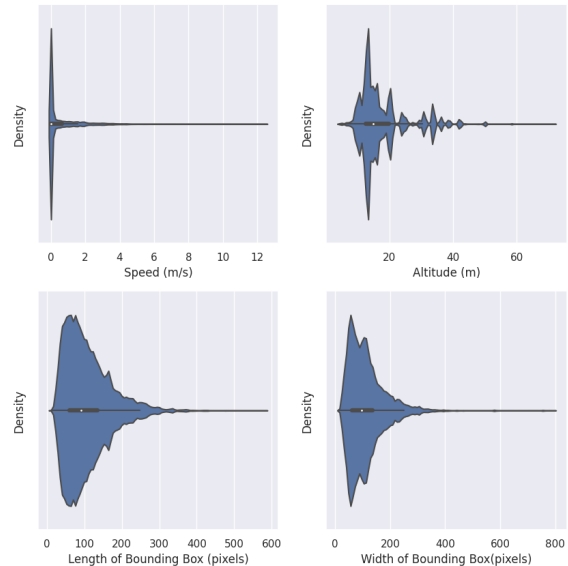


Figure 2: Distribution of altitude, velocity, and bounding box values associated with usable video data for wildlife behavior inference

After reconciling the available telemetry and behavioral annotations, our final dataset consisted of 133 minutes of raw videos from 10 missions over three days. Of this final dataset, only 87 minutes contained video frames of suitable quality to both detect the animals of interest and determine their behavior, which is a 65% usability rate. In other words, a third of the video data collected in the field could not be used to infer animal behavior using computer vision methods. We analyzed the telemetry data to identify what range of altitudes and bounding box values produced video that was suitable to be analyzed with machine learning computer vision methods. Specifically, we focused on computer vision methods for object detection with YOLO (Jocher, Chaurasia, and Qiu 2023) to automatically localize the animals in the scene and automatic behavior detection proposed in *KABR* to automatically extract behavior. Using these insights, we improved the autonomous navigation UAV model proposed in Kline et al. (2023) to increase the yield of usable video data. We implemented the autonomous navigation model from Kline et al. (2023) using the Yolov8n model (Jocher, Chaurasia, and Qiu 2023) as the object detection model to localize the animals.

Results and Discussion

We used our telemetry analysis shown in Table 2 and Figure 2 to improve the autonomous navigation model for tracking wildlife proposed in Kline et al. (2023). This autonomous navigation model samples the video stream once every second, and utilizes a YOLO model (Jocher, Chaurasia, and Qiu 2023) to automatically detect the animals in the frame. Once a zebra or giraffe is detected, the navigation model

Data	Description	Source
Date & Time	Date and time of video	Telemetry & KABR
Frame	Video frame (30 frames per second)	Telemetry & KABR
Behavior	Behavior of the animal	KABR
Bounding Box Dimensions	Dimensions of the bounding boxes surrounding the animal in pixels	KABR
Speed	Velocity of the UAV in x,y,z direction (meters per second)	Telemetry
Altitude	Altitude of UAV above ground (meters)	Telemetry

Table 1: Telemetry and KABR Dataset Kholiavchenko et al. (2024)

	Altitude (m)	Velocity (m/s)	Bounding Box Width (pixels)	Bounding Box Length (pixels)
Mean	17.55	0.62	106.22	110.40
Standard Deviation	7.80	1.31	61.65	69.01
Minimum	4.40	0.00	11.00	14.00
25%	12.70	0.00	61.00	61.67
50%	14.90	0.00	92.00	96.67
75%	19.80	0.63	136.00	137.00
Maximum	71.60	12.47	583.00	767.00

Table 2: Telemetry Data Analysis: Ranges of Optimal Values for UAV altitude, velocity, and animal bounding box sizes in pixels



Figure 3: Example Autonomous Navigation Model Output

uses the bounding box dimensions generated by YOLO to determine the commands to send to the UAV. For situations when multiple animals are in the frame, this approach tracks the group by using the bounding box values to calculate the herd centroid. If this value strays sufficiently far from the centroid of the UAV’s camera, the navigating model sends commands to the UAV to adjust its position to return the herd centroid to the center of the camera’s view. One weakness of this autonomous navigation approach is that it is constrained to generating commands in x-y plane only (i.e. ‘fly left/right’ and ‘fly forward/backward’), since this approach assumes a constant altitude.

Our approach adds the ability to generate commands in z-direction (altitude), as well as the x and y directions, by integrating the insights from the telemetry analysis. This new

and improved autonomous navigation model includes a range of preferred altitudes to tune the UAV’s flight path. We also improve the autonomous navigation model by minimizing the movement of the UAV while navigating the UAV close enough for the bounding boxes to measure at approximately 100×100 pixels. This produces a 18.2% increase in accuracy and a 8.3% increase in the F1 score, as shown in Table 3. Figure 3 shows the output of the autonomous navigation model proposed here. We report the F1 value, which combines the precision and recall scores of the model’s predicted commands to the UAV. For this application, ‘positive’ values means that the UAV is directed to move, while ‘negative’ values means that the UAV is directed to hover in place. An example of a ‘true positive’ result would be the model correctly predicting that the next command sent to the UAV should ‘fly left 3 meters’. An example of a ‘false positive’ result would be the model incorrectly predicting that the next command sent to the UAV should ‘fly left 3 meters’, when the correct decision was directing the UAV to hover in place, as determined by the expert pilot executing the manual mission. The accuracy values reported in Table 3 simply measures the number of predicted commands that match the original flight path conducted by the expert pilot. The telemetry dataset is imbalanced in favor of ‘positive’ values, so the F1 score is a better measure of performance compared to accuracy.

As shown in Table 2, the mean altitude of the usable behavior video frames is 17.55 meters with a standard deviation of 7.8 meters. The maximum altitude is 71.60 meters, however the 75th percentile is 19.8 meters, so this maximum is an outlier. Figure 4 shows the relationship between behavior and altitude, where behaviors are ordered by how often they occur in the KABR dataset. The majority of the behavior samples were collected below 30 meters, with rare behaviors such as ‘urinating’ and ‘defecating’ captured below 20 meters. These results suggest that behavior is best captured between 10 and

Metric	Original Navigation Model	New & Improved Navigation Model	Improvement
% of actions matching original flight	68.8	87.0	+18.2
F1 Score	82.1	90.4	+8.3

Table 3: Autonomous Navigation Model Performance Compared to Approach from Kline et al. (2023)

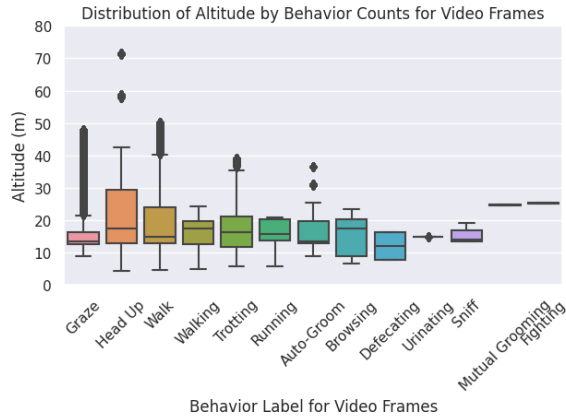


Figure 4: Behavior Distribution by Altitude (behavior labels ordered from most to least common, left to right)

30 meters.

The mean velocity of usable video frames is 0.62 meters per second, with a standard deviation of 1.31 meters per second. The distribution of velocity values is shown in Table 2 and Figure 2. For context, 0.62 meters per second is approximately 1.4 miles per hour. The 75th percentile is only 0.63 meters per second, while the maximum velocity is 12.47 meters per second and the lower percentiles are 0 meters per second (i.e. hovering), suggesting the maximum value may be skewed by the UAV approaching the wildlife. These results suggest that behavior is best captured by minimizing the UAV’s movement and hovering once it reaches a suitable altitude, enabling the generation of bounding boxes of the appropriate size for behavioral analysis.

The mean bounding box dimensions is approximately 106×110 pixels, with a standard deviation of 62×69 pixels. The distribution of bounding box dimensions is shown in Figure 5. The bounding box size is related to the distance between the UAV and the wildlife. Using the bounding box size as an approximation for distance is an effective approach to tracking groups of free-ranging wildlife autonomously, as shown in Kline et al. (2023). This approach eliminates the need for costly distance calculations, and can leverage the bounding box values produced by light-weight object detectors such as the YOLO nano model (Jocher, Chaurasia, and Qiu 2023).

Conclusion & Future Work

We present guidelines for gathering in situ imageomics wildlife behavior video data with UAVs and demonstrate how these guidelines can be integrated into an autonomous

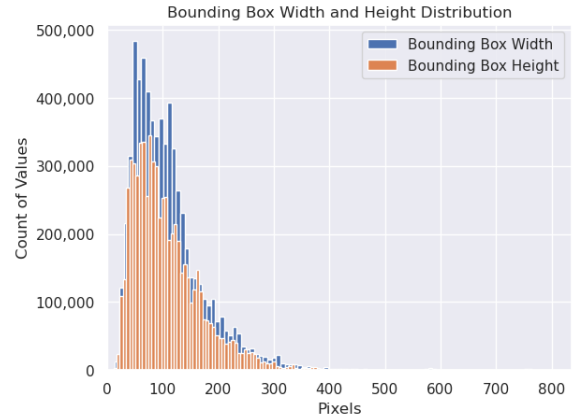


Figure 5: Bounding Box Dimension Distribution

UAV navigation model. The techniques proposed in this work can be applied to others’ UAV wildlife video datasets to determine the optimal guidelines for additional species. These techniques can also be adapted to increase of the yield of usable UAV data beyond wildlife behavior studies; for example, individual identification of wildlife, such as those implemented by Wildbook (Berger-Wolf et al. 2017). Future work will explore integrating behavior-adaptive flight data into autonomous UAV navigation models. This approach can minimize the disturbance from UAV data collection by using a behavior recognition model to determine if animals are exhibiting signs of reacting to the UAV’s presence, such as vigilance behaviors or running. Reinforcement learning approaches can also be used to balance competing metrics, such as the range of altitudes and minimum desired bounding box size. These approaches should be validated on free ranging wildlife to confirm the proposed approaches produce data suitable for analysis with computer vision techniques for imageomics applications.

Acknowledgments

This work was supported by NSF Awards No. 2118240 and 2112606. Data was collected in accordance with Kenya’s National Commission for Science, Technology & Innovation research license NACOSTI/P/22/18214 and Princeton University’s Institutional Animal Care and Use Committee 1835F.

References

Andrew, W.; Greatwood, C.; and Burghardt, T. 2019. Aerial Animal Biometrics: Individual Friesian Cattle Recovery and

Visual Identification via an Autonomous UAV with Onboard Deep Inference. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 237–243.

Berger-Wolf, T. Y.; Rubenstein, D. I.; Stewart, C. V.; Holmberg, J. A.; Parham, J.; Menon, S.; Crall, J.; Oast, J. V.; Kiciman, E.; and Joppa, L. 2017. Wildbook: Crowdsourcing, computer vision, and data science for conservation. arXiv:1710.08880.

Boubin, J.; Chumley, J.; Stewart, C.; and Khanal, S. 2019. Autonomic Computing Challenges in Fully Autonomous Precision Agriculture. In *2019 IEEE International Conference on Autonomic Computing (ICAC)*, 11–17.

Brookes, O.; Mirmehdi, M.; Kühn, H.; and Burghardt, T. 2023. Triple-stream Deep Metric Learning of Great Ape Behavioural Actions. (arXiv:2301.02642). ArXiv:2301.02642 [cs].

Horn, G. V.; Aodha, O. M.; Song, Y.; Cui, Y.; Sun, C.; Shepard, A.; Adam, H.; Perona, P.; and Belongie, S. 2018. The iNaturalist Species Classification and Detection Dataset. In *CVPR*.

Imageomics. 2023. About the Imageomics Institute (NSF OAC-2118240).

Jocher, G.; Chaurasia, A.; and Qiu, J. 2023. YOLO by Ultralytics.

Kholiavchenko, M.; Kline, J.; Ramirez, M.; Stevens, S.; Sheets, A.; Babu, R. R.; Banerji, N.; Campolongo, E.; Thompson, M.; Tiel, N. V.; Miliko, J.; Bessa, E.; Duporge, I.; Berger-Wolf, T.; Rubenstein, D.; and Stewart, C. 2024. KABR: In-Situ Dataset for Kenyan Animal Behavior Recognition from Drone Videos. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*.

Kline, J. M.; Stewart, C.; Berger-Wolf, T.; Stewart, C. V.; and Rubenstein, D. I. 2023. A Framework for Autonomic Computing for In Situ Imageomics. In *4th IEEE International Conference on Autonomic Computing and Self-Organizing Systems - ACSOS 2023*.

Koger, B.; Deshpande, A.; Kerby, J.; Graving, J.; Costelloe, B.; and Couzin, I. 2023. Quantifying the movement, behaviour and environmental context of group-living animals using drones and computer vision. <https://onlinelibrary.wiley.com/doi/abs/10.1111/1365-2656.13904>.

Tabak, M. A.; Norouzzadeh, M. S.; Wolfson, D. W.; Sweeney, S. J.; Vercauteren, K. C.; Snow, N. P.; Halseth, J. M.; Di Salvo, P. A.; Lewis, J. S.; White, M. D.; Teton, B.; Beasley, J. C.; Schlichting, P. E.; Boughton, R. K.; Wight, B.; Newkirk, E. S.; Ivan, J. S.; Odell, E. A.; Brook, R. K.; Lukacs, P. M.; Moeller, A. K.; Mandeville, E. G.; Clune, J.; and Miller, R. S. 2019. Machine learning to classify animal species in camera trap images: Applications in ecology. *Methods in Ecology and Evolution*, 10(4): 585–590.

Zualkernan, I.; Dhou, S.; Judas, J.; Sajun, A. R.; Gomez, B. R.; and Hussain, L. A. 2022. An IoT System Using Deep Learning to Classify Camera Trap Images on the Edge. *Computers*, 11(11): 13.