

Adapting the re-ID challenge for static sensors

Avirath Sundaresan¹ Jason Parham² Jonathan Cralf² Rosemary Warungu⁵ Timothy Mutham⁵ Jackson Miliiko⁵ Margaret Mwangi⁵ Jason Holmberg⁴ Tanya Berger-Wolf^{6,4} Daniel Rubenstein⁷ Charles Stewart^{3,4} Sara Beery⁸

¹ Computing and Mathematical Sciences, California Institute of Technology, Pasadena, CA, USA

² Kitware, Clifton Park, NY, USA

³ Rensselaer Polytechnic Institute, Troy, NY, USA

⁴ WildMe, Portland, OR, USA

⁵ Laikipia Zebra Project, Mpala Research Centre, Laikipia, KE

⁶ Computer Science and Engineering, Ohio State University, Columbus, OH, USA

⁷ Ecology and Evolutionary Biology, Princeton University, Princeton, NJ, USA

⁸ Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, USA

* E-mail: asundare@caltech.edu

Abstract: The Grévy's zebra, an endangered species native to Kenya and southern Ethiopia, has been the target of sustained conservation efforts in recent years. Accurately monitoring Grévy's zebra populations is essential for ecologists to evaluate ongoing conservation initiatives. Recently, in both 2016 and 2018, a full census of the Grévy's zebra population has been enabled by the Great Grévy's Rally (GGR), a citizen science event that combines teams of volunteers to capture data with computer vision algorithms that help experts estimate the number of individuals in the population. In this work, we explore complementary, scalable, cost-effective, and long-term Grévy's population monitoring using a deployed network of camera traps at the Mpala Research Centre in Laikipia County, Kenya. Unlike the human-captured images collected by large teams of volunteers at GGR events, camera trap images are characterized by poorer quality, high rates of occlusion, and high spatio-temporal similarity within image bursts. We employ an image filtering pipeline incorporating animal detection, species identification, viewpoint estimation, quality evaluation, and temporal subsampling to compensate for these factors and obtain individual crops from camera trap images of suitable quality for re-ID. We then employ the Local Clusterings and their Alternatives (LCA) algorithm, a hybrid computer vision & graph clustering method for animal re-ID, on the resulting high-quality crops. Our method efficiently processed 8.9M unlabeled camera trap images from 70 camera traps over two years into 685 encounters of 173 unique individuals, requiring only 331 contrastive same-vs-different-individual decisions from a human reviewer.

1 Introduction

The population of Grévy's zebras experienced a dramatic decline beginning in the 1970s largely due to hunting and competition for food and water resources with local pastoral communities. Estimates have placed the number of Grévy's zebras remaining in the wild at under 2,000, with the vast majority in the Samburu region of central Kenya. Due to extensive conservation efforts by the Kenyan and Ethiopian governments as well as environmental NGOs, the population of Grévy's zebras have stabilized in recent years [1].

Accurately censusing Grévy's zebra populations is critical for ecologists to evaluate these existing conservation efforts. Developing and maintaining a census of known individuals requires effective animal re-identification methods to ensure that only unique individuals are included in population counts [2]. A popular method for population size estimation is "capture-mark-recapture" [3]. By this method, a set of animals in the target population is first captured and marked, then a second set of animals is independently recaptured, and finally a population estimate formed from the number of animals captured twice. This method, however, proves difficult to scale to large populations and territories, and may lead to inaccurate population estimates when animals are not confined to the study area and are capable of evading tagging [4, 5]. Further, the tagging methods used in population estimation studies (such as ear tags and radio collars [6–8]) can be excessively expensive and time-consuming to implement in the field, and overly harmful to the animal [9].

A modern alternative to manual mark-recapture studies that does not have these limitations is exemplified in the Great Grévy's Rally (GGR) events of 2016 and 2018 where volunteers spread over the range of Grévy's zebra to photograph as many animals as possible

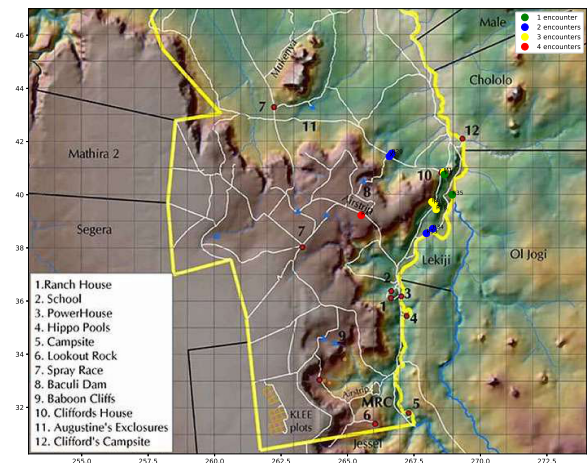


Fig. 1: Map of several encounters of a single individual within Mpala Research Centre identified from camera trap data by our method and requiring minimal human labeling.

over two consecutive days. Relying on the distinctive appearance of an animal itself as a means for identification, a combination of algorithmic and human curation efforts produced population estimates that have been accepted as definitive by the Kenyan Wildlife Service and the IUCN [10].

Complementary to the focused burst of human effort for the GGR photographic events, we explore the possibility of using a fleet of camera traps to collect the data needed for population monitoring. For animals that allow for sight-based ID, images taken by camera

traps are a potentially cost-effective and non-invasive method to re-identify individuals and obtain robust population estimates. Unlike alternative methods such as tagging or even the field photography of the GGR events, camera traps do not require the physical presence of field researchers for animal re-identification; this ensures that the natural behaviors of animals are not disrupted and saves considerable time and resources [11, 12].

However, analysis of camera trap images by human observation alone is not practical. In recent years, computer vision algorithms have been shown to present a highly accurate and standardized method for camera trap image analysis. Computer vision techniques have seen significant success in automated species detection & identification from camera trap images, especially with advances in deep learning [13–17]. Accurate species ID from camera trap images has laid the foundation for the next step in an end-to-end photographic censusing pipeline from camera trap images: automated individual re-ID (for reviews, see: [12, 18, 19]). However, the challenging nature of the automatically-captured data (motion blur, occlusion, poor lighting, far-away animals) often leads to flawed animal re-identification even by human experts [20]. Additionally, images captured in a burst from the same motion trigger are often highly repetitive, leading to potential bias in automatic re-identification if not handled carefully. In this work, we seek to adapt existing techniques for animal re-identification in human-captured data — particularly those developed in response to the effort required for the GGR events — to the challenging static camera trap paradigm.

Unlike many other classification tasks with numerous and highly similar classes, animal re-identification is an open-set classification problem [21], with the need to assign every unique individual to its own class and to recognize novel individuals unseen during training. There are two key categories of animal re-ID algorithms. *Ranking* algorithms for re-ID query an image of the target individual against an existing database to obtain a ranking of the most confident matches. Hotspotter [22] is one such texture-based ranking algorithm that uses the SIFT [23] algorithm to extract salient features from the query image and subsequently a nearest neighbor search to match the query image against the database. The algorithm is specialized for striped and spotted animals, and it has been used for re-ID of Grévy’s zebra [22, 24] and several other animal species from camera trap data [25–27]. *Verification* algorithms for re-ID, on the other hand, do not require querying an existing database; instead, a verification algorithm simply decides whether two images contain the same individual. An example of a verification algorithm is the Verification Algorithm for Match Probabilities (VAMP) [28], a random forest classifier [29, 30] that receives two images and decides if they contain the same animal, different animals, or are incomparable. Lastly, contrastive deep learning algorithms, such as the Pose Invariant Embedding (PIE) network [31, 32], can learn a global feature embedding (instead of handcrafted features with Hotspotter & VAMP) for a particular image, allowing for distance-based comparisons with the feature embeddings of other images in a database; this allows PIE to serve as both a ranking and verification algorithm simultaneously. However, unlike Hotspotter and VAMP, PIE (and deep learning algorithms in general) requires a significant amount of training data. Instead, when considering a population without individual-level ground truth labels readily available, it is preferable to use classical computer vision for ranking and verification that can be easily bootstrapped with minimal human supervision.

In this paper, we use Hotspotter and VAMP in conjunction with the Local Clusterings and their Alternatives (LCA) decision management algorithm [33, 34] for Grévy’s zebra re-ID. Like other animal and human re-ID systems [35–37], the LCA algorithm offers a human-in-the-loop approach to dynamically cluster annotations by individual and request human reviews for verification hard cases. To the best of our knowledge, this marks the first use of interactive / error-driven clustering algorithms like LCA for animal re-ID from exclusively camera trap data for any species.

2 Methods

2.1 Camera Trap Dataset

We use images collected from a network of 70 camera traps distributed around the Mpala Research Centre in Kenya’s Laikipia Plateau. The network has collected 8.9 million images over the past two years of deployment. There are four types of camera trap placement schemes in the network: systematically in a grid, at “magnet” sites (e.g. salt licks), as well as expert-targeted & random placement along roads (see Fig. 2).

2.2 Automated Species Identification

Prior to re-ID, the raw camera trap images were first passed through a YOLO v2 species detection model [38] to localize all zebra with a bounding box (both Grévy’s and plains zebra species), and crop the region of the image within each bounding box for downstream use. Localized bounding boxes crop out irrelevant and potentially distracting background information and yield distinct *annotations* of independent individuals from images that feature several animals. YOLO v2 has shown to be more accurate for animal detection than alternative object detection models, such as Faster R-CNN [39]. Next, the cropped regions are classified to zebra species — Grévy’s vs. plains zebra — and viewpoint — left vs. right — by a DenseNet model. Only right, front-right, and back-right viewpoints are considered for identification; differing viewpoints cannot be matched with one another, as the right and left sides of Grévy’s zebra are distinct. The species and viewpoint classification model was trained on human-captured and labeled data from the Great Grévy’s Rally in 2018 (GGR18) [34].

Finally, we wished to ensure that these annotations would be *universally comparable* to make verification decisions easier for both VAMP and the human verifier. For Grévy’s zebras, this entails annotations that have both the distinctive hip and shoulder chevron, which we will call “Census Annotations” (CA) [34]. A DenseNet with a linear classification layer, also trained on GGR18 data, was used to decide whether an annotation was indeed a CA, and produce an associated CA confidence score. Then, a regression network was used to narrow the annotation region to only surround the hip and shoulder chevron, which was then saved as a new “Census Annotation Region”, or CA-R [34].

2.3 Filtering Pipeline

Mark-recapture statistical models are highly sensitive to inaccurate identifications, but are built with the assumption that not all individuals are seen. Thus, it is important to prioritize the precision of our re-identification pipeline over assigning an ID to each annotation. One significant factor in inaccurate identification, by both humans and algorithms, is image quality — blurry, or poorly lit annotations are difficult or impossible to reliably identify. Filtering annotations by quality is therefore a critical component of the re-ID pipeline, and automated quality filtering reduces the amount of human time and work required to rectify matching errors caused by undesirable annotations. Here, we devise an annotation filtering scheme suitable for re-ID from camera traps.

As already described, we only consider census annotation regions (CA-Rs), those that show right-side viewpoints of Grévy’s zebra, including both the shoulder and hip chevron. Beyond this, we further filter by time of day: only annotations from images taken between 6:30 AM and 7:00 PM (sunrise and sunset at Mpala) were kept for re-ID. Due to the camera trap settings, images taken during the daytime are optical (RGB) and are of higher resolution (13 MP). Nighttime images are taken with an infrared flash at lower resolution (9 MP). Qualitatively, nighttime images are more difficult for human reviewers than daytime images; without modification, we believed the VAMP verification algorithm was likely to perform better on higher contrast and quality images taken during the day.

Next, the annotations were filtered by *encounter*. To define the encounters, we used an agglomerative clustering approach: for each

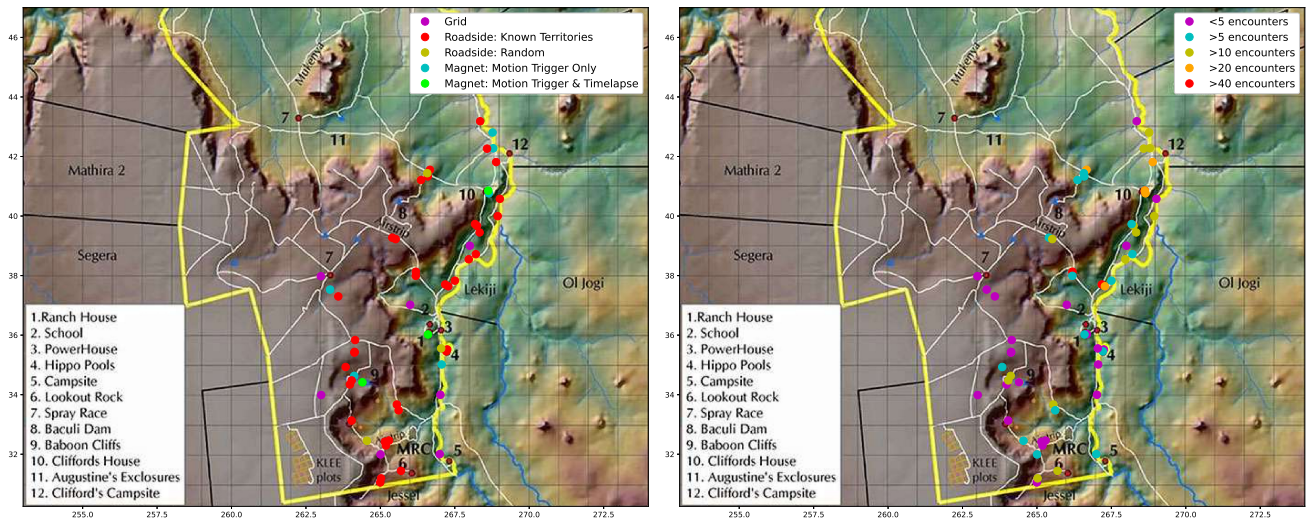


Fig. 2: (Left) Location of our camera traps across the Mpala Research Centre. These cameras are placed according to four distinct strategies: random grid, randomly along roads, at known Grévy's territories along roads, and at magnets such as salt licks and watering holes. **(Right)** Map of all Grévy's encounters across Mpala.

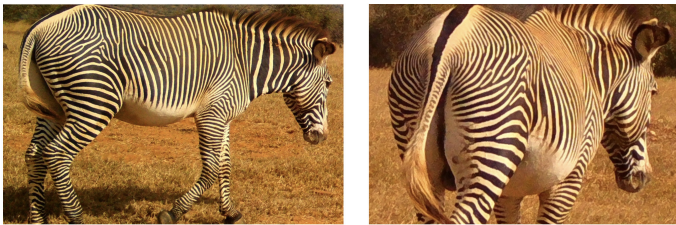


Fig. 3: Example of a good and poor quality CA of an individual within an encounter. **(Left)** Annotation with CA score of 0.9997; the hip and shoulder chevrons are very clearly visible. **(Right)** Annotation with CA score of 0.0032; viewpoint is correct but the hip and shoulder chevrons are not clearly visible.

camera, annotations from images taken within the same minute and in consecutive minutes were grouped together in the same encounter. Next, the annotation with the highest Census Annotation confidence score was selected from each encounter, and the rest were discarded. As the camera traps take images in bursts, the images in a encounter are nearly identical, and hence the corresponding annotations very likely must feature the same individuals. This step encourages matching across the best representatives from every encounter, reducing the amount of data required but preserving the relevant information. Experiments in [34] show that filtering in this manner, in particular filtering for CA-Rs, does not have a significant impact on the individual count, but dramatically reduces the human effort required to produce it. (rephrase - later).

Lastly, the annotations were filtered for CA confidence score. Annotations above a 0.31 confidence score threshold were kept for re-ID. In previous work [34], a lower threshold was shown to add incomparable annotations to the re-ID database and encourage spurious matches, and a higher threshold was shown to eliminate relevant sightings without a considerable increase in matching speed and accuracy. The filtering pipeline yielded 685 right-view Grévy's zebra annotations across the reserve, with each representing a single encounter (see Fig. 2).

2.4 Zebra ID Curation

The goal of the LCA decision management algorithm is to build an *identification graph* with the filtered annotations, using automated ranking & verification and delaying human intervention as much as possible. In the identification graph, annotations are represented as vertices, and relationships between annotations (whether positive or negative matches) are represented as edges. The LCA algorithm

seeks to group annotations into clusters corresponding to individuals by maximizing positive edge weights within clusters and negative weights between clusters. This simplifies the identity labeling process significantly: instead of being asked to identify an annotation as one of up to 150 known individuals, reviewers are asked to compare pairs of annotations and determine whether they belong to the same individual or not. This contrastive task can be accomplished even by non-experts due to innate human capacity for pattern matching — it is similar to a game of spot the difference.

The ID graph is initialized with no edges, with every annotation constituting its own cluster. For every annotation, the Hotspotter ranking algorithm returns a list of its most confident matched annotations, and LCA forms edges between the annotation and its potential matches. Next, the VAMP verification algorithm evaluates every pair of vertices connected by an edge and assigns an edge weight (positive or negative) based on its confidence in the match. The LCA algorithm has two main phases. In the *scoring* phase, it keeps the initial edges and weights intact. LCA iterates through every local clustering (a single cluster or a pair of clusters) and decides if there exists an alternative clustering that increases the sum of intra-cluster edge weights and decreases the sum of inter-cluster edge weights. If such an alternative clustering exists, the new arrangement is adopted. Once there are no longer any better local alternative clusterings, LCA proceeds to the *stability* phase. In this phase, LCA considers local clusterings for which the difference between it and its next best alternative are sufficiently small that changes in edge weights can potentially introduce a superior alternative clustering. For such local clusterings, LCA requests additional reviews to VAMP and the human (if VAMP cannot litigate the review on its own) to determine if certain edge weights must change. Once all local clusterings are significantly better than their next best alternative, the algorithm is considered to have converged.

3 Results

8.9 million camera trap images from the initial dataset were passed through the species ID pipeline, yielding 84,383 zebra images (including both plains and Grévy's zebra). Following species and viewpoint filtering, 23,512 right-view Grévy's zebra annotations remained across 3,338 distinct encounters. After excluding night-time encounters, 1,138 daytime encounters remained. We then sampled the single highest-quality annotation from each of the 1,138 encounters, as determined by our CA model, to avoid matching on near-identical images from the same burst. We further used a score quality threshold on CA that reduced the number of annotations to be

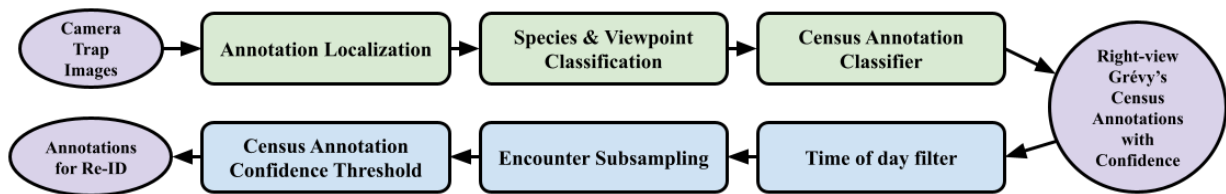


Fig. 4: Data curation pipeline from unlabeled camera trap images to high-quality individually identifiable census identifications. Data is represented in purple, models in green, and quality filtering processes in blue. The annotations remaining after this curation pipeline next have individual features extracted using VAMP which are used as input into the LCA algorithm.

identified to 734. As a last step, we used a blurriness filter to remove any additional challenging annotations, resulting in a final set of 685.

In order to associate these 685 annotations to clusters of individuals, the LCA algorithm requested 5,403 automated reviews by VAMP and 331 additional human reviews; pairwise comparisons of annotations were performed largely automatically using VAMP, with a automation rate of 93.9%. The converged ID graph had 173 clusters, each corresponding to an individual zebra ID within the 685 annotations. On average, each individual was sighted across 1.9 static camera traps and 3.96 encounters, indicating significant matching across space and time. See Supplementary Fig. 5 for the distribution of encounters for each individual, and Supplementary Fig. 6 for the distribution of static camera traps that sighted each individual. The number of individuals sighted by each camera also differed based on its placement strategy (random grid, known Grévy’s territories along roads, randomly along roads, and both timelapse & motion-trigger traps at magnet sites). On average, camera traps at a magnet site (e.g. salt lick, dam, etc.) sighted the most individuals on average (8.5 per trap) and also produced the most *first* individual sightings on average (5.85 per trap). See Table 1 and Supplementary Fig. 7 for the distribution of total individuals sighted by cameras of each placement strategy. Particularly, the encounters for large individual clusters were spread across many months, with some spanning more than a year (see Supplementary Fig. 8). Based on the distinct encounters and cameras associated with each individual, we can generate spatial maps of encounters for any individual across the reserve: see Fig. 1 for a map of encounters for a particular zebra (Individual 32).

4 Discussion

The proposed combined method of subsampling and curation has enabled us to efficiently combine the 685 high-quality encounters in our static camera dataset to form 173 clusters, each representing an individual Grévy’s zebra. The use of LCA for semi-automatic, interactive decision-making resulted in fewer than 0.5 human reviews per annotation. By contrast, when working only with ranking algorithms, typically several — five to ten — potential matches must be examined, and LCA includes consistency checking implicitly.

Each annotation in the cluster is representative of an encounter with that individual at a known position and time, and enables us to convert a set of 8.9 million unlabeled camera trap images into identified encounters that can be used as the input into a spatially-explicit mark recapture model [40] to estimate the total Grévy’s population. Because camera traps are monitoring long-term, in this case over two years unlike the contained 2-day period captured by each Great Grévy’s Rally, assumptions about no births and no deaths and no transients in the population are invalid. It is important to note that the ecologists at the Laikipia Zebra Project have confirmed that the Grévy’s population at Mpala has been an open set over the period in question: of the estimated 150 Grévy’s individuals currently seen over repeat seasons at Mpala, only 4 are “stable” and reside on the reserve long-term, while the vast majority (130 expert re-identified individuals) have been seen to migrate and return periodically, driven by water scarcity and threat of predation [41]. It is thus likely that the camera network captured some number of transient individuals over the two year sampling span.

Table 1 Number of individuals sighted by camera placement strategy

	Total	Avg.	New
Random grid	16	2.29	1.43
Roadside, known territories	105	2.76	2.26
Roadside, random	9	3	0.33
Magnet	107	8.5	5.85

Limitations of the method may include imperfect CA labeling, where annotations with seemingly high CA scores are in actuality difficult to match, leading to two (or more) clusters in place of a single cluster. Additionally, we observe that errors could also occur via potential failures in any one of the algorithmic components whether at the species ID or at the individual re-ID level. Human decision failures in particular may inflate the number of individual clusters, with ambiguous matches mislabeled by the human reviewer resulting in more clusters than necessary. A significant focus of our ongoing work is to further analyze these results in order to tease these factors apart.

Lastly, while the method is currently limited to images taken during daylight hours, an approximately equal-sized set of images and annotations is collected at night. In recent preliminary tests, we selected the twenty-five night time annotations having the best combination of census annotation score and image contrast, and matched them against the clustered daytime annotations produced as a result of the work described here. For twenty of them (80%) Hotspotter produced correct matches in the top few. This shows potential for extending the overall method to the night-time, but many details must be revisited due to the overall lower quality of nocturnal images before they can be reliably included in generating enhanced census results.

5 Conclusion

In this paper, we perform efficient semi-automated Grévy’s zebra re-ID from camera trap data, with an algorithmic tool chain using Hotspotter for ranking, VAMP for verification, and LCA for decision management. Our method can be used even by non-zebra experts [34], as it only requires contrastive comparisons between pairs of individuals instead of matching each individual into the previously identified population. Ultimately, this curational re-ID process found 173 distinct individuals in camera trap data collected across a two year period at the Mpala Research Centre.

Going forward, we aim to further refine the filtering pipeline to improve matching with fewer human reviews and reduce attrition of data through the pipeline. We also are excited to further explore the results of the method, for example by adjusting the CA score threshold and observing how this affects LCA convergence behavior & results. We also hope to further adapt LCA to enable temporal sequences and spatial relationships to be taken into account, to enable us to better make use of the additional images of the individuals from each camera trap image burst and the spatial structure of the habits of individuals over time. Finally, and along similar lines, we hope to introduce “short-circuiting” to reduce human reviews when spatio-temporal constraints imply that there is no possibility of a match.

6 References

- 1 D. Rubenstein, B. Low Mackey, ZD Davidson, F. Kebede, and S.R.B King. Equus grevyi. August 2016. IUCN Red List of Threatened Species.
- 2 Devys Tuia, Benjamin Kellenberger, Sara Beery, Blair R Costelloe, Silvia Zuffi, Benjamin Risse, Alexander Mathis, Mackenzie W Mathis, Frank van Langevelde, Tilo Burghardt, Roland Kays, Holger Klinck, Martin Wikelski, Iain D Couzin, Grant van Horn, Margaret C Crofoot, Charles V Stewart, and Tanya Berger-Wolf. Perspectives in machine learning for wildlife conservation. *Nat. Commun.*, 13(1):792, February 2022.
- 3 R Pradel. Utilization of Capture-Mark-Recapture for the study of recruitment and population growth rate. *Biometrics*, 52(2):703–709, 1996.
- 4 K A Lee, C Huvencors, O Gimenez, V Peddemors, and R G Harcourt. To catch or to sight? a comparison of demographic parameter estimates obtained from mark-recapture and mark-resight models. *Biodivers. Conserv.*, 23(11):2781–2800, October 2014.
- 5 K Tilling and J A Sterne. Capture-recapture models including covariate effects. *Am. J. Epidemiol.*, 149(4):392–400, February 1999.
- 6 K A Alexander and M J Appel. African wild dogs (*lycaon pictus*) endangered by a canine distemper epizootic among domestic dogs near the masai mara national reserve, kenya. *J. Wildl. Dis.*, 30(4):481–485, October 1994.
- 7 C R Thouless. Long distance movements of elephants in northern kenya. *Afr. J. Ecol.*, 33(4):321–334, December 1995.
- 8 Gail H Collins, Steven L Petersen, Craig A Carr, and Leon Pielstick. Testing VHF/GPS collar design and safety in the study of free-roaming horses. *PLoS One*, 9(9):e103189, September 2014.
- 9 Gary W Witmer. Wildlife population monitoring: some practical considerations. *Wildl. Res.*, 32(3):259–263, June 2005.
- 10 Joy Omulupi Ontita, Jennifer Weston, George Anyona I I, Geoffrey Chege, David Kimiti, Kaia Tombak, Andrew Gersick, and Nancy Rubenstein. The state of kenya’s grevy’s zebras and reticulated giraffes: Results of the great grevy’s rally 2018.
- 11 Zackary J Delisle, Elizabeth A Flaherty, Mackenzie R Nobbe, Cole M Wzientek, and Robert K Swihart. Next-generation camera trapping: Systematic review of historic trends suggests keys to expanded research applications in ecology and conservation. *Front. Ecol. Evol.*, 9, February 2021.
- 12 Stefan Schneider, Graham W Taylor, Stefan Linquist, and Stefan C Kremer. Past, present and future approaches using computer vision for animal re-identification from camera trap data. *Methods Ecol. Evol.*, 10(4):461–470, April 2019.
- 13 Mohammad Sadeqh Norouzzadeh, Dan Morris, Sara Beery, Neel Joshi, Nebojsa Jojic, and Jeff Clune. A deep active learning system for species identification and counting in camera trap images. October 2019.
- 14 Sara Beery, Yang Liu, Dan Morris, Jim Piavis, Ashish Kapoor, Markus Meister, Neel Joshi, and Pietro Perona. Synthetic examples improve generalization for rare classes. April 2019.
- 15 Sara Beery, Dan Morris, and Siyu Yang. Efficient pipeline for camera trap image review. July 2019.
- 16 Sara Beery, Guanhang Wu, Vivek Rathod, Ronny Votel, and Jonathan Huang. Context R-CNN: Long term temporal context for Per-Camera object detection. December 2019.
- 17 Michael A Tabak, Mohammad S Norouzzadeh, David W Wolfson, Erica J Newton, Raoul K Boughton, Jacob S Ivan, Eric A Odell, Eric S Newkirk, Reesa Y Conrey, Jennifer Stenglein, Fabiola Iannarilli, John Erb, Ryan K Brook, Amy J Davis, Jesse Lewis, Daniel P Walsh, James C Beasley, Kurt C VerCauteren, Jeff Clune, and Ryan S Miller. Improving the accessibility and transferability of machine learning algorithms for identification of animals in camera trap images: MLWIC2. *Ecol. Evol.*, 10(19):10374–10383, October 2020.
- 18 Prashanth C Ravoor and Sudarshan T. S. B. Deep learning methods for Multi-Species animal re-identification and tracking – a survey. *Computer Science Review*, 38:100289, November 2020.
- 19 Maxime Vidal, Nathan Wolf, Beth Rosenberg, Bradley P Harris, and Alexander Mathis. Perspectives on individual animal identification from biology and computer vision. *Integr. Comp. Biol.*, 61(3):900–916, October 2021.
- 20 Paul Meek, Karl A Vernes, and Gregory Falzon. On the reliability of expert identification of Small-Medium sized mammals from camera trap photos. 2013.
- 21 Chuanxing Geng, Sheng-jun Huang, and Songcan Chen. Recent advances in open set recognition: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 43(10):3614–3631, 2020.
- 22 Jonathan P. Crall, Charles V. Stewart, Tanya Y. Berger-Wolf, Daniel I. Rubenstein, and Siva R. Sundaresan. Hotspotter — patterned species instance recognition. In *2013 IEEE Workshop on Applications of Computer Vision (WACV)*, pages 230–237, 2013. doi: 10.1109/WACV.2013.6475023.
- 23 David G Lowe. Distinctive image features from Scale-Invariant keypoints. *Int. J. Comput. Vis.*, 60(2):91–110, November 2004.
- 24 D. Rubenstein, C. Stewart J. Parham, T. Y. Berger-Wolf, J. Holmberg, J. Crall, B. L. Mackey, S. Funnel, K. Cockerill, Z. Davidson, L. Mate, C. Nzomo, R. Warungu, D. Martins, V. Ontita, J. Omulupi, J. Weston, G. Anyona, G. Chege, D. Kimiti, K. Tombak, A. Gersick, and N. Rubenstein. The state of kenya’s grevy’s zebras and reticulated giraffes: Results of the great grevy’s rally 2018. *Wildlife Direct*, 2018.
- 25 Heebok Park, Anya Lim, Tae-Young Choi, Seung-Yoon Baek, Eui-Geun Song, and Yung Chul Park. Where to spot: individual identification of leopard cats (*prionailurus bengalensis euptilurus*) in south korea. *Hangug hwangyeong saengtae haghoeji*, 43(1):1–5, December 2019.
- 26 Robert B Nipko, Brogan E Holcombe, and Marcella J Kelly. Identifying individual jaguars and ocelots via pattern-recognition software: Comparing HotSpotter and wild-ID. *Wildl. Soc. Bull.*, 44(2):424–433, June 2020.
- 27 Drew Blount, Eve Bohnett, Jason Holmberg, Jason Parham, Sorosh Poya Faryabi, Orjan Johansson, Li An, Bilal Ahmad, Wajid Khan, and Stephane Ostrowski. Comparison of two individual identification algorithms for snow leopards after automated detection. January 2022.
- 28 J. P. Crall. *Identifying individual animals using ranking, verification, and connectivity*. PhD thesis, Dept. Comput. Sci., Rensselaer Polytechnic Inst., Troy, NY, USA, 2017.
- 29 Leo Breiman. Random forests. *Mach. Learn.*, 45(1):5–32, October 2001.
- 30 M Pal. Random forest classifier for remote sensing classification. *Int. J. Remote Sens.*, 26(1):217–222, January 2005.
- 31 Liang Zheng, Yujia Huang, Huchuan Lu, and Yi Yang. Pose invariant embedding for deep person re-identification. January 2017.
- 32 Olga Moskvayak, Frederic Maire, Feras Dayoub, Asia O Armstrong, and Mahsa Baktashmotlagh. Robust re-identification of manta rays from natural markings by learning pose invariant embeddings. In *2021 Digital Image Computing: Techniques and Applications (DICTA)*, pages 1–8, November 2021.
- 33 C. Stewart, T. Y. Berger-Wolf, J. Parham, A. Sundaresan, S. Beery, and J. Crall. Local clusterings and their alternatives. *In preparation*.
- 34 J. R. Parham. *Animal detection for photographic censusing*. PhD thesis, Dept. Comput. Sci., Rensselaer Polytechnic Inst., Troy, NY, USA, 2021.
- 35 Peter Kulits, Jake Wall, Anka Bedetti, Michelle Henley, and Sara Beery. ElephantBook: A Semi-Automated Human-in-the-Loop system for elephant Re-Identification. June 2021.
- 36 Paul Bodesheim, Jan Blunk, Matthias Körschens, Clemens-Alexander Brust, Christoph Käding, and Joachim Denzler. Pre-trained models are not enough: active and lifelong learning is important for long-term visual monitoring of mammals in biodiversity research—individual identification and attribute prediction with image features from deep neural networks and decoupled decision models applied to elephants and great apes. *Mamm. Biol.*, 102(3):853–875, June 2022.
- 37 Hanxiao Wang, Shaogang Gong, Xiatian Zhu, and Tao Xiang. Human-In-The-Loop person Re-Identification. December 2016.
- 38 Jason Parham, Charles Stewart, Jonathan Crall, Daniel Rubenstein, Jason Holmberg, and Tanya Berger-Wolf. An animal detection pipeline for identification. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1075–1083, March 2018.
- 39 Jason Parham and Charles Stewart. Detecting plains and grevy’s zebras in the realworld. In *2016 IEEE Winter Applications of Computer Vision Workshops (WACVW)*, pages 1–9, March 2016.
- 40 Austin M Green, Mark W Chynoweth, and Çağan Hakkı Şekercioğlu. Spatially explicit capture-recapture through camera trapping: a review of benchmark analyses for wildlife density estimation. *Frontiers in Ecology and Evolution*, 8:563477, 2020.
- 41 Kenya Wildlife Service. Recovery and action plan for grevy’s zebra (*equus grevyi*) in kenya (2017-2026). 2017.

Supplementary Figures

Additional data visualization and analysis figures have been included in this section to give additional context to the reader.

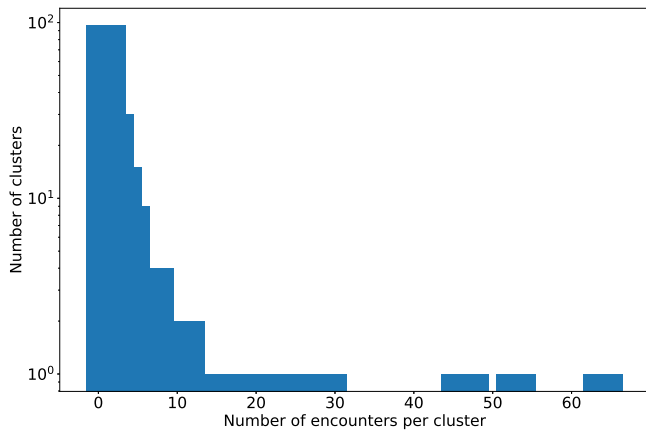


Fig. 5: Distribution of number of encounters per individual cluster.

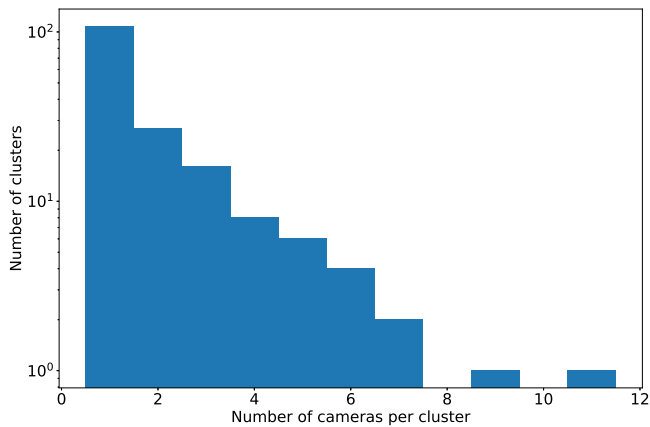


Fig. 6: Distribution per individual of distinct camera locations where the individual was seen, emphasizing that our method is able to overcome the static background bias and identify individuals in different contexts.

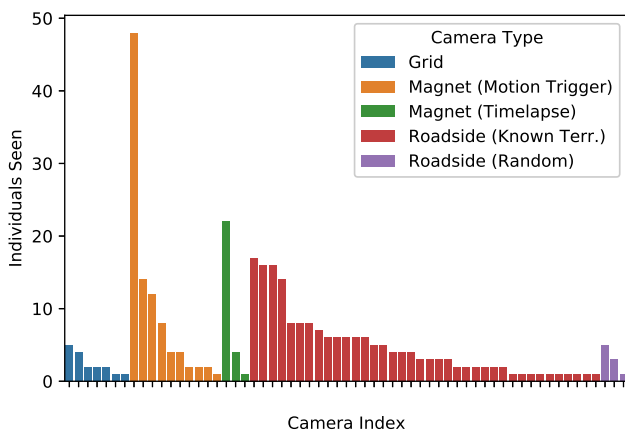


Fig. 7: Distribution of individuals sighted per camera, sorted by placement strategy. A distinction is made here between motion-trigger and timelapse cameras at magnet sites.

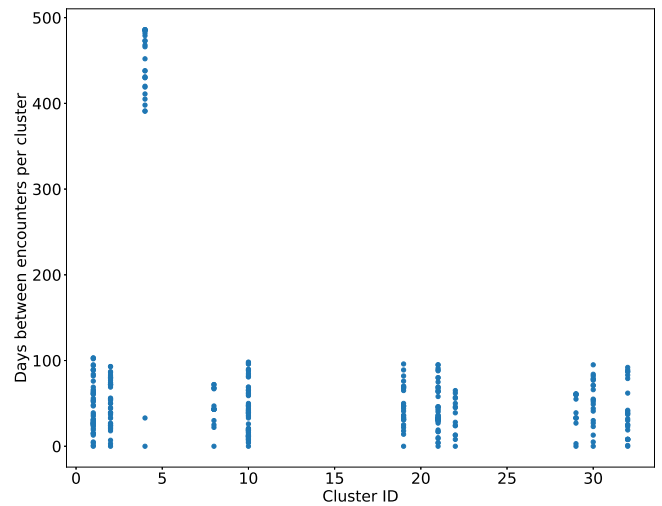


Fig. 8: Time in days between encounters for the largest clusters in the ID graph. Note that our method re-identified one of the individuals ~400 days later, after it had migrated out of Mpala and then returned.