Mammal Research
https://doi.org/10.1007/s13364-022-00657-z

SHORT COMMUNICATION

Reducing identification errors of African carnivores from photographs through computer-assisted workflow

Stijn Verschueren[1,2] · Ezequiel Chimbioputo Fabiano[3] · Mikael Kakove[1] · Bogdan Cristescu[1,4] · Laurie Marker[1]

Received: 14 June 2022 / Accepted: 19 October 2022

Abstract
Precise and accurate estimates of population size are fundamental to the study and conservation of wildlife. Identification of individual animals is often required to obtain such estimates, yet manual classifications by human observers induce bias, which can propagate across long-term datasets or large spatial scales. Pattern recognition algorithms have been developed to aid identification efforts and here, we demonstrate the efficacy of this technology to reduce misidentifications of African large carnivores in historic data. We used a 7-year camera-trapping dataset from north-central Namibia and revised cheetah and leopard individuals identified by human observers through a pattern recognition algorithm, HotSpotter, implemented in a web-based and open-source application, the African Carnivore Wildbook (ACW). Verification of individuals with ACW resulted in a reduction from 43 to 40 cheetah individuals and from 59 to 46 leopard individuals. This is equivalent to a difference of 7% and 22% of individuals identified for cheetahs and leopards, respectively. Additionally, this revision increased the proportion of individuals that were detected over multiple years and at multiple locations. Our findings may have implications for population and trend estimates of these and other species, given that current estimates often rely on manual identification that could overestimate population size.

Keywords African Carnivore Wildbook · Camera trap · Cheetah · HotSpotter · Leopard · Pattern recognition

Introduction

The need to correctly assess species status and population trends is well-established for wildlife conservation (IUCN 2012). This is particular for keystone species like large carnivores that are threatened by various anthropogenic pressures (Ripple et al. 2014). Large carnivores are often difficult to detect because they occur at low densities across large spatial scales (Carbone and Gittleman 2002). Non-invasive, cost-effective survey techniques increase the detectability of these species (Kelly et al. 2012), and current analytical methods largely rely on a capture-recapture framework to estimate population abundances and infer population densities (Obbard et al. 2010; Gilbert et al. 2021). One critical assumption of this framework is the correct identification of individuals as erroneous identifications can lead to significant bias in models used to derive population parameter estimates, with implications for conservation management and decision-making (Stevick et al. 2001; Tucker et al. 2019; Johansson et al. 2020; Rakhimberdiev et al. 2022). Modeling misidentification errors explicitly may overcome this bias, but this requires a minimal detection rate that is often higher than those achieved for large carnivores (Yoshizaki et al. 2009; Morrison et al. 2011). Therefore, technical solutions to reduce misidentifications by human observers (hereafter, manual identification) before the data are used in a modeling framework may hold better promise.

Camera traps are widely used as a non-invasive survey technique to detect elusive species, while allowing for the individual identification of species with unique body markers (i.e., spots, rosettes, stripes) (Burton et al. 2015).

Communicated by Dries Kuijper

Stijn Verschueren
stijn@cheetah.org

1 Cheetah Conservation Fund, Otjiwarongo, Namibia

2 Evolutionary Ecology Group, University of Antwerp, Antwerp, Belgium

3 University of Namibia, Katima, Namibia

4 Namibia University of Science and Technology, Windhoek, Namibia

Camera-trap surveys have gained broad interest for photographic mark-recapture (Agha et al. 2018; Green et al. 2020), yet data processing and individual identification are time-consuming and prone to human error, in particular for large datasets (Foster and Harmsen 2012; Tucker et al. 2019). Technological advances facilitate individual identifications by implementing pattern recognition algorithms that search for and suggest matching images in a database to a query image (Bolger et al. 2012; Crall et al. 2013). This approach has been widely used across various taxa (e.g., Shorrocks and Croft 2009; Sherley et al. 2010; Hughes and Burghardt 2017; Balme et al. 2019), which underlines the importance and broad applications of this technology for wildlife conservation.

Several research efforts have focused on the performance of pattern recognition algorithms using test datasets (Bolger et al. 2012; Crall et al. 2013; Matthé et al. 2017; Schneider et al. 2019; Nipko et al. 2020). The extent to which computer-assisted identification software can reduce misidentifications in historic databases is however lacking, while identification errors may propagate systematically across long-term datasets and large spatial scales. This may have implications for inferring the status of many threatened species given that current estimates often rely on manual identification, which may over or under-estimate population size. We subjected a photographic dataset of cheetah (*Acinonyx jubatus*) and leopard (*Panthera pardus*) individuals previously identified by human observers over a 7-year camera-trapping period, to a pattern recognition algorithm, HotSpotter, implemented in a web-based application and open-source platform, the African Carnivore Wildbook (ACW; https://africancarnivore.wildbook.org). After computer-assisted identification, we refined the number of individuals in the dataset and discussed the implications of our findings for the conservation of these and other threatened species.

Methods

Between July 2007 and February 2014, we conducted eight repeated camera-trapping surveys on privately-owned farmland in north-central Namibia (S20.4835, E17.0312). Camera-trap surveys lasted 3–5 months, and the trap array covered a mean area of 381 km^2 per survey. For more information on the study design, we refer to Fabiano et al. (2020). We included data from 12 locations that were consistently monitored throughout the survey period. Camera-trap images were classified manually to species level. Cheetah and leopard individuals were identified manually during the research period. For each species, one observer led the identification effort with assistance from other observers. Identifications were based on a set of one to three unique spot patterns on any clearly visible part of either side of an animal. We compared all left flanks among each other and all right flanks among each other. Other viewpoints were used to complement the identification process when available. If a photo did not match any other photo, it was considered to depict a new individual. An ID catalogue was created with a selection of high-quality images from all available viewpoints of the individual to assist the manual identification process.

We revised the ID catalogue using the HotSpotter algorithm implemented in ACW to detect cases of splitting (i.e., a splitting error splits the captures from one individual into two and creates a "ghost" individual (Johansson et al. 2020)) (Crall et al. 2013, 2021; Van Oast et al. 2022). The HotSpotter algorithm searches for pair-wise similarities in key patterns and textures between images in the database to a query image and produces a list of match candidates ranked on relative similarity scores. The HotSpotter algorithm produces correct top-ranked labels for > 95% of queries (Crall et al. 2013) and selects a correct match as its top rank 71–82% of the time (Nipko et al. 2020). Final confirmation of match candidates to the query image relies on careful human evaluation and decision. We only revised high-quality images compiled in the ID catalogue for which HotSpotter provided 85–99% positive matching rates (Nipko et al. 2020).

We compared the outcome of both identification methods to each other and calculated the difference in minimum population size (i.e., minimum number of individuals detected) relative to the total number of individuals identified through manual identification. In addition, we compiled frequency distributions of the number of individuals detected at multiple locations and over multiple years. We used the Pearson chi-squared test to detect differences in space-use (i.e. number of locations an individual was detected) and naïve survival (i.e., number of years an individual was detected: year [last detection]–year [first detection]). These metrics are indicative for potential bias induced by misidentifications as individual recapture probability across space and time may be higher than human observers presume. Therefore, we expected that a larger proportion of individuals is captured over multiple locations and across multiple years after verification through ACW.

Results and discussion

The camera-trapping effort resulted in 17,599 cheetah images and 2115 leopard images. Human observers identified 43 cheetah and 59 leopard individuals. Verification of the ID catalogue (457 cheetah images and 200 leopard images) through ACW reduced these numbers to 40 cheetah and 46 leopard individuals. This is equivalent to a difference of 7% and 22% of individual animals identified for cheetahs and leopards, respectively (Fig. 1). Verification through ACW increased the proportion of individuals detected at multiple locations and over multiple years, though this was not significant for either species (Table 1).

The effect of identification errors tends to result in systematic overestimation of population size as well as underestimation of space-use and naïve survival. In a controlled experiment on snow leopards (*Panthera uncia*), misidentifications resulted in an overestimation of 35% of population abundance (Johansson et al. 2020), while in a simulation of a red knot (*Calidris canutus*) population, survival was underestimated up to 6% (Tucker et al. 2019). Our results add to this growing body of literature using a real-world example, and we demonstrate how computer-assisted identification can reduce misidentifications. This is critical given the well-established need for robust population assessments, in particular of large carnivores.

We recognize that our research only accounted for splitting errors; hence, spatial capture-recapture was not performed. This would require the revision of the full dataset through computer-assisted identification to additionally detect combination errors (i.e., captures of two individuals are combined into one, so that an animal not captured previously is erroneously believed to be a recapture), shifting errors (i.e., a photographic capture is shifted from one individual's capture history to another), and exclusion errors (i.e., a photographic capture is not assigned to any capture history and instead is excluded from classification despite it containing enough information for it to be reliably classified) (Johansson et al. 2020). We therefore rely on previous studies demonstrating the effect of misidentifications on estimates of population density and associated parameters (Stevick et al. 2001; Tucker et al. 2019; Johansson et al. 2020; Rakhimberdiev et al. 2022), while we also recommend future studies to investigate the effect of identification errors on other statistical assumptions in population modeling, such as the random distribution of individuals, population closure, and sufficient detection rates. We are concerned

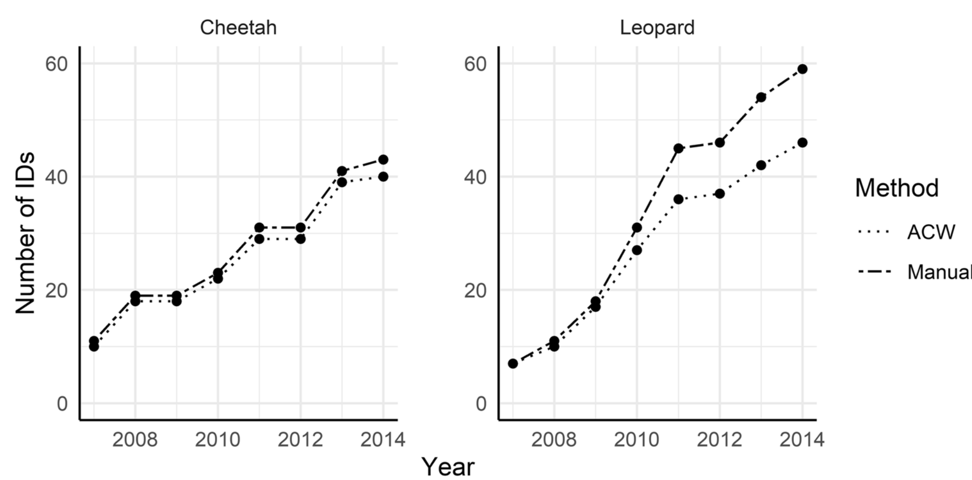

Fig. 1 Cumulative sum of new individuals (IDs) per year identified manually and with the African Carnivore Wildbook (ACW)

Table 1 The proportion of cheetah and leopard individuals captured at multiple locations and over multiple years identified manually and with the African Carnivore Wildbook (ACW). Pearson's chi-squared tests were used to determine differences in the frequency distribution of individuals grouped by "Space-use" and "Naïve survival" according to the different identification methods

Cheetah								Test statistic
Space-use (# locations)	1	2	3	4	5	6		
Manual	0.47	0.26	0.02	0.12	0.05	0.09		$\chi^2=0.13$, df = 5, p = 0.99
ACW	0.43	0.28	0.03	0.13	0.05	0.10		
Naïve survival (# years)	1	2	3	4	5	6	7	
Manual	0.70	0.19	0.05	0.00	0.07	0.00	0.00	$\chi^2=$NaN, df = 6, p = NA*
ACW	0.65	0.23	0.05	0.00	0.05	0.00	0.03	
Leopard								Test statistic
Space-use (# locations)	1	2	3	4	5			
Manual	0.66	0.22	0.07	0.00	0.03			$\chi^2=1.95$, df = 4, p = 0.75
ACW	0.61	0.22	0.11	0.02	0.04			
Naïve survival (# years)	1	2	3	4	5	6	7	
Manual	0.64	0.08	0.08	0.10	0.00	0.07	0.00	$\chi^2=4.95$, df = 6, p = 0.55
ACW	0.54	0.07	0.13	0.11	0.02	0.07	0.04	

*Data was too sparse; hence, no statistical testing was performed about the interpretation of current estimates when based on manual identification (Johansson et al. 2020), especially for long-term datasets across large spatial scales. Re-evaluation of such datasets through computer-assisted workflow may be needed and could refine conservation assessments of threatened species and populations pending its outcome.

The difference in identification errors between cheetah and leopard was notable and suggests that species-specific traits may facilitate or complicate individual identifications, causing variable bias in identification success among species. This idea is straightforward at the species level due to differences in body markers (such as spots or rosettes), whereas for conspecifics which naturally have comparable coat patterns, differences in behavior may affect identification success. Nonetheless, behavioral differences could also affect identification success among species. Cheetahs are typically more day-active, return more frequently to, and stay longer at marking sites compared to leopards (Verschueren et al. 2021), which may yield better quality images and a greater amount of different viewpoints that allow easier identification. Identification errors may also increase with the addition of individuals, and in our study system, leopards tend to occur at higher densities compared to cheetahs (Marker and Dickman 2005; Fabiano et al. 2020). Alternatively, inter-observer variability may account for differences among species, as the level of expertise influences identification success (Van Horn et al. 2014; Johansson et al. 2020). Irrespective of the underlying cause, we demonstrate herein how the use of computer-assisted identification allows for improved population metrics across species.

We recommend the use of pattern recognition algorithms to assist individual identifications, both for future identification efforts and for the revision of historic data and current estimates. We acknowledge some of the established shortcomings of approaches using pattern recognition algorithms, including but not limited to a dearth of training data, controlling for image quality, and environmental, positional, and timing-related challenges (Schneider et al. 2019). We therefore highlight the importance of human involvement in the decision process to obtain the best outcomes.

Identification errors may increase with surveys across large spatial scales, spanning multiple years, and with population density. Large carnivores are typically wide-ranging and long-living; thus, identification errors from manual methods may propagate systematically through the photographic dataset. Computer-assisted workflow streamlines data management and facilitates image comparisons that are often overlooked, seem unlikely, or stem from different data sources. As per the Wildbook fundamental model, ACW additionally provides a space for data management, sharing, and citizen science and facilitates pre-processing tasks such as bulk imports and animal detection and post-processing tasks such as compilation of capture histories and exploratory data analysis (Berger-Wolf et al. 2017; Blount et al. 2022). This results in reduced identification errors, saves time and resources spent on data processing, and promotes collaborative efforts. Centralizing datasets enables new and broader avenues of research, as well as collaboration, data sharing, and best practices to inform wildlife conservation efforts (Ahumada et al. 2020).

Acknowledgements We thank Cheetah Conservation Fund staff and volunteers who contributed to the data collection. We thank Tech 4 Conservation (https://www.t4c.org), Botswana Predator Conservation (https://www.bpctrust.org), and Wild Me (https://www.wildme.org) for the development of the ACW.

Funding Fieldwork was supported by Cheetah Conservation Fund donors. ACW was funded by Microsoft for Startups grant provided to Wild Me.

Data availability The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Competing interests The authors declare no competing interests.

References

Agha M, Batter T, Bolas EC, Collins AC, Gomes da Rocha D, Monteza-Moreno CM, Preckler-Quisquater S, Sollmann R (2018) A review of wildlife camera trapping trends across Africa. Afr J Ecol 56:694–701. https://doi.org/10.1111/AJE.12565

Ahumada JA, Fegraus E, Birch T, Flores N, Kays R, O'Brien TG, Palmer J, Schuttler S, Zhao JY, Jetz W, Kinnaird M, Kulkarni S, Lyet A, Thau D, Duong M, Oliver R, Dancer A (2020) Wildlife Insights: a platform to maximize the potential of camera trap and other passive sensor wildlife data for the planet. Environ Conserv 47:1–6. https://doi.org/10.1017/S0376892919000298

Balme G, Rogan M, Thomas L, Pitman R, Mann G, Whittington-Jones G, Midlane N, Broodryk M, Broodryk K, Campbell M, Alkema M, Wright D, Hunter L (2019) Big cats at large: density, structure, and spatio-temporal patterns of a leopard population free of anthropogenic mortality. Popul Ecol 61:256–267. https://doi.org/10.1002/1438-390X.1023

Berger-Wolf TY, Rubenstein DI, Stewart C V., Holmberg JA, Parham J, Menon S, Crall J, Van Oast J, Kiciman E, Joppa L (2017) Wildbook: crowdsourcing, computer vision, and data science for conservation. arXiv 1710.08880

Blount D, Gero S, Van Oast J, Parham J, Kingen C, Scheiner B, Stere T, Fisher M, Minton G, Khan C, Dulau V, Thompson J, Moskvyak O, Berger-Wolf T, Stewart CV, Holmberg J, Levenson JJ (2022) Flukebook: an opensource AI platform for cetacean photo identification. Mamm Biol 2022. https://doi.org/10.1007/S42991-021-00221-3

Bolger DT, Morrison TA, Vance B, Lee D, Farid H (2012) A computer-assisted system for photographic mark–recapture analysis. Methods Ecol Evol 3:813–822. https://doi.org/10.1111/J.2041-210X.2012.00212.X

Burton AC, Neilson E, Moreira D, Ladle A, Steenweg R, Fisher JT, Bayne E, Boutin S (2015) Wildlife camera trapping: a review and recommendations for linking surveys to ecological processes. J Appl Ecol 52:675–685. https://doi.org/10.1111/1365-2664.12432

Carbone C, Gittleman JL (2002) A common rule for the scaling of carnivore density. Science 295:2273–2276. https://doi.org/10.1126/science.1067994

Crall J, Parham J, Mulich M, Chan K, Zrutfield, Weideman HJ, Weinstock A, Jablons Z, Blount D, Eylor, Lowe D, Southerland K, Batbouta A, Audrey, Scheiner B, Wrona J, Charles (2021) WildMe-Org/wildbook-ia: Version 4.0.0 (v4.0.0). Zenodo. https://zenodo.org/record/5765701

Crall JP, Stewart C V, Berger-Wolf TY, Rubenstein DI, Sundaresan SR (2013) HotSpotter - Patterned species instance recognition, in: IEEE Workshop on Applications of Computer Vision (WACV). Institute of Electrical and Electronics Engineers, Tampa, Florida, USA, pp. 230–237

Fabiano EC, Sutherland C, Fuller AK, Nghikembua M, Eizirik E, Marker L (2020) Trends in cheetah *Acinonyx jubatus* density in north-central Namibia. Popul Ecol 62:233–243. https://doi.org/10.1002/1438-390X.12045

Foster RJ, Harmsen BJ (2012) A critique of density estimation from camera-trap data. J Wildl Manage 76:224–236. https://doi.org/10.1002/jwmg.275

Gilbert NA, Clare JDJ, Stenglein JL, Zuckerberg B (2021) Abundance estimation of unmarked animals based on camera-trap data. Conserv Biol 35:88–100. https://doi.org/10.1111/COBI.13517

Green AM, Chynoweth MW, Şekercioğlu ÇH (2020) Spatially explicit capture-recapture through camera trapping: a review of benchmark analyses for wildlife density estimation. Front Ecol Evol 8:473. https://doi.org/10.3389/FEVO.2020.563477/BIBTEX

Hughes B, Burghardt T (2017) Automated visual fin identification of individual great white sharks. Int J Comput vis 122:542–557. https://doi.org/10.1007/S11263-016-0961-Y/FIGURES/16

IUCN (2012) IUCN red list categories and criteria. Gland, Switzerland

Johansson Ö, Samelius G, Wikberg E, Chapron G, Mishra C (2020) Low M (2020) Identification errors in camera-trap studies result in systematic population overestimation. Sci Reports 101(10):1–10. https://doi.org/10.1038/s41598-020-63367-z

Kelly MJ, Betsch J, Wultsch C, Mesa B, Mills LS (2012) Noninvasive sampling for carnivores. In: Boitani L, Power RA (eds) Carnivore ecology and conservation: a handbook of techniques. Oxford University Press, UK, pp 47–69

Marker L, Dickman AJ (2005) Factors affecting leopard (Panthera pardus) spatial ecology, with particular reference to Namibian farmlands. South African J Wildl Res 35:105–115

Matthé M, Sannolo M, Winiarski K, van der SpitzenSluijs A, Goedbloed D, Steinfartz S, Stachow U (2017) Comparison of photomatching algorithms commonly used for photographic capture–recapture studies. Ecol Evol 7:5861–5872. https://doi.org/10.1002/ECE3.3140

Morrison TA, Yoshizaki J, Nichols JD, Bolger DT (2011) Estimating survival in photographic capture–recapture studies: overcoming misidentification error. Methods Ecol Evol 2:454–463. https://doi.org/10.1111/J.2041-210X.2011.00106.X

Nipko RB, Holcombe BE, Kelly MJ (2020) Identifying individual jaguars and ocelots via pattern-recognition software: comparing HotSpotter and Wild-ID. Wildl Soc Bull 44:424–433. https://doi.org/10.1002/WSB.1086

Obbard ME, Howe EJ, Kyle CJ (2010) Empirical comparison of density estimators for large carnivores. J Appl Ecol 47:76–84. https://doi.org/10.1111/J.1365-2664.2009.01758.X

Rakhimberdiev E, Karagicheva J, Saveliev A, Loonstra AHJ, Verhoeven MA, Hooijmeijer JCEW, Schaub M, Piersma T (2022) Misidentification errors in reencounters result in biased estimates of survival probability from CJS models: evidence and a solution using the robust design. Methods Ecol Evol. https://doi.org/10.1111/2041-210X.13825

Ripple WJ, Estes JA, Beschta RL, Wilmers CC, Ritchie EG, Hebblewhite M, Berger J, Elmhagen B, Letnic M, Nelson MP, Schmitz OJ, Smith DW, Wallach AD, Wirsing AJ (2014) Status and ecological effects of the world's largest carnivores. Science 343:1241484. https://doi.org/10.1126/science.1241484

Schneider S, Taylor GW, Linquist S, Kremer SC (2019) Past, present and future approaches using computer vision for animal re-identification from camera trap data. Methods Ecol Evol 10:461–470. https://doi.org/10.1111/2041-210X.13133

Sherley R, Burghardt T, Barham P, Campbell N, Cuthill I (2010) Spotting the difference: towards fully-automated population monitoring of African penguins Spheniscus demersus. Endanger Species Res 11:101–111. https://doi.org/10.3354/esr00267

Shorrocks B, Croft DP (2009) Necks and networks: a preliminary study of population structure in the reticulated giraffe (Giraffa camelopardalis reticulata de Winston). Afr J Ecol 47:374–381. https://doi.org/10.1111/J.1365-2028.2008.00984.X

Stevick PT, Palsbøll PJ, Smith TD, Bravington MV, Hammond PS (2001) Errors in identification using natural markings: rates, sources, and effects on capture-recapture estimates of abundance. Can J Fish Aquat Sci 58:1861–1870. https://doi.org/10.1139/F01-131

Tucker AM, McGowan CP, Robinson RA, Clark JA, Lyons JE, Derose-Wilson A, Du Feu R, Austin GE, Atkinson PW, Clark NA (2019) Effects of individual misidentification on estimates of survival in long-term mark–resight studies. Condor 121:1–13. https://doi.org/10.1093/CONDOR/DUY017

Van Horn RC, Zug B, Lacombe C, Velez-Liendo X, Paisley S (2014) Human visual identification of individual Andean bears Tremarctos ornatus. Wildlife Biol 20:291–299. https://doi.org/10.2981/WLB.00023

Van Oast J, Holmberg J, June R, Kingen CW, Blount D, Mark, Me W, gpoautomation, Winstanley G, schibel21, Becker A, Donovan J, Wainner C, marcial21, Overbeck C, Southerland K, mbrecunier, Huayhuaca L, Matt, Shay E (2022) WildMeOrg/Wildbook: 2022-01-31 Release (2022-01-31_release_notes). Zenodo. 10.5281/ZENODO.5999339

Verschueren S, Briers-Louw WD, Cristescu B, Fabiano E, Nghikembua M, Torres-Uribe C, Walker EH, Marker L (2021) Spatiotemporal sharing and partitioning of scent-marking sites by cheetahs and leopards in north-central Namibia. Afr J Ecol 59:605–613. https://doi.org/10.1111/aje.12878

Yoshizaki J, Pollock KH, Brownie C, Webster RA (2009) Modeling misidentification errors in capture–recapture studies using photographic identification of evolving marks. Ecology 90:3–9. https://doi.org/10.1890/08-0304.1