Wildlife Society Bulletin 1–10; 2020; DOI: 10.1002/wsb.1086

Tools and Technology

Identifying Individual Jaguars and Ocelots via Pattern-Recognition Software: Comparing HotSpotter and Wild-ID

ROBERT B. NIPKO,[1] *Department of Fish and Wildlife Conservation—Virginia Tech, 310 W Campus Drive, Blacksburg, VA 24061, USA*
BROGAN E. HOLCOMBE, *Department of Fish and Wildlife Conservation—Virginia Tech, 310 W Campus Drive, Blacksburg, VA 24061, USA*
MARCELLA J. KELLY, *Department of Fish and Wildlife Conservation—Virginia Tech, 310 W Campus Drive, Blacksburg, VA 24061, USA*

ABSTRACT Camera-trapping is widespread in wildlife studies, especially for species with individually unique markings to which capture–recapture analytical techniques can be applied. The large volume of data such studies produce have encouraged researchers to increasingly look to computer-assisted pattern-recognition software to expedite individual identifications, but little work has been done to formally assess such software for camera-trap data. We used 2 sets of camera-trap images—359 images of jaguars (*Panthera onca*) and 332 images of ocelots (*Leopardus pardalis*) collected from camera traps deployed in 4 study sites in Orange Walk District, Belize, in 2015 and 2016—to compare the accuracy of 2 such programs, HotSpotter and Wild-ID, and assess the effect of image quality on matching success. Overall, HotSpotter selected a correct match as its top rank 71–82% of the time, whereas the rate for Wild-ID was 58–73%. Positive matching rates for both programs were highest for high-quality images (85–99%) and lowest for low-quality images (28–52%). False match rates were very low for HotSpotter (0–2%) but these were greater in Wild-ID (6–28%). When lower ranks were also considered, both programs performed similarly (overall 22–24% nonmatches for HotSpotter, 17–26% nonmatches for Wild-ID). We found that in both programs, images more often matched to other images of the same quality; therefore, including multiple reference images of an individual, of different qualities, improves matching success. These programs do not provide fully automatic identification of individuals and human involvement is still required to confirm matches, but we found that they are effective tools to expedite processing of camera-trap data. We also offer usage recommendations for researchers to maximize the benefits of these tools.

KEY WORDS camera trap, HotSpotter, jaguar, *Leopardus pardalis*, ocelot, *Panthera onca*, pattern-recognition, photographic identification, Wild-ID.

Recognizing and monitoring individual animals is essential to obtain insights for wildlife management, including tracking individual fitness and reproductive success (Dinsmore and Johnson 2012), determining local activity patterns and migrations (Millspaugh et al. 2012), and monitoring demographic parameters like abundance and survival (Pierce et al. 2012; Satter et al. 2019*a*, *b*). Many wildlife species possess individually distinct natural markings that can be used to noninvasively track individuals. The rise of high-quality, low-cost digital cameras—especially digital trail cameras—has made this approach particularly viable for studying individually marked species like jaguars (*Panthera onca*), and thus camera-trapping has become a widespread technique in wildlife studies (Kelly et al. 2012).

Received: 12 March 2019; Accepted: 15 November 2019
Published:

[1]*E-mail: rnipko@vt.edu*

Studies employing digital photography, however, present new challenges because they usually generate vast amounts of data. Manually reviewing images and matching individual animals by eye according to their spot, stripe, or blotch patterns is time-consuming, tedious, and error-prone. Many camera-trapping studies use capture–recapture analytical techniques (e.g., Otis et al. 1978, White and Burnham 1999, Williams et al. 2002, Royle et al. 2013) and several studies have found that misidentifications can severely bias the resulting estimates (Creel et al. 2003, Lukacs and Burnham 2005, Yoshizaki et al. 2009, Morrison et al. 2011).

Computer-aided photo-identification or pattern-recognition is an emerging technology with the potential to address the challenges of identifying individuals by natural patterns. Early implementations involved complex interfaces and required substantial input from the end-user (e.g., Hiby and Lovell 1990, Kelly 2001). More recently, several open-source, freely downloadable software packages have been developed that are more sophisticated and are increasingly being used in wildlife studies. Such studies involve diverse taxa including ungulates (e.g., Strauss et al. 2015, Lea et al. 2016, Morrison et al. 2016*b*), felids (Lingaraja et al. 2017, Miller et al. 2018), cetaceans (Jablons 2016), herpetofauna (Bendik et al. 2013, Cross et al. 2014, Tumulty et al. 2018), and even mollusks (Barord et al. 2014). Despite their proliferation, little has been published formally comparing software packages. Of the comparisons that have been done (e.g., Morrison et al. 2016*a*, Cruickshank and Schmidt 2017, Matthé et al. 2017, Suriyamongkol and Mali 2018), all used data sets consisting of photographs of herpetofauna. In all cases, this involved capturing the animals and photographing them with a handheld camera with standardized distance and positioning. These methods produce images of much higher quality than the blurry, partial, or oddly oriented images often obtained in camera-trapping studies, yet these programs are also being widely used with such mixed-quality images. To our knowledge, there is no published comparison of the effectiveness of these programs when applied specifically to camera-trapping data.

Pattern-recognition programs can broadly be classified into 2 different groups: pixel-based and feature-based (Matthé et al. 2017). Programs employing pixel-based algorithms examine the entire image on a pixel-by-pixel basis. Images that have a greater number of differences at specific pixel locations are less likely to be a match than images that have a large number of pixel positions with the same (or very similar) values. Examples include Amphident (Matthé et al. 2008) and APHIS (Moya et al. 2015). Feature-based algorithms instead identify several distinct characteristics (e.g., spots, stripes, warts, etc.) within each image, then compare the shapes and positions of these characteristics between images. Examples of software using a feature-based approach include Wild-ID (Bolger et al. 2012), HotSpotter (Crall et al. 2013), and I^{3}S Pattern+ (Van Tienhoven et al. 2007). Matthé et al. (2017) found that pixel-based algorithms generally performed better (i.e., had a higher recognition rate—which is equivalent to a lower false rejection rate—for images with known matches) across several salamander data sets, but they also found that these algorithms were sensitive to differences in the way an image was oriented or cropped. For example, an image of a salamander that was cropped somewhat higher on the body than another image of the same salamander would be much less likely to be matched. Therefore, pixel-based algorithms would likely be inappropriate for most camera-trap data because of the great variation in image quality. Thus, we focused our comparison on programs using feature-based algorithms because these were much more robust to quality variations, and are likely to perform better for camera-trap data. I^{3}S Pattern+ has been primarily cited in studies of aquatic organisms and herpetofauna (e.g., Cochran et al. 2016, Calmanovici et al. 2018, Davis et al. 2018), whereas Wild-ID and HotSpotter are the programs most widely cited in studies of large terrestrial wildlife, including camera-trap studies (e.g., Lee et al. 2017, Lingaraja et al. 2017, Miller et al. 2018). Therefore, our comparison focused on the latter 2 programs.

Wild-ID (Bolger et al. 2012) implements the Scale Invariant Feature Transform (SIFT) of Lowe (2004) to identify, extract, and describe distinctive image features. The SIFT feature descriptors can identify analogous features in different images despite differences in scale, illumination, and orientation of the subject to the camera. This makes it potentially well-suited to photographs taken in uncontrolled field settings, especially camera-trap photos. For each combination of images in an analysis, Wild-ID then performs a pairwise comparison of the shape and relative geometry of SIFT features between the images. Each pair is assigned a numerical score based on these comparisons, with a larger score indicating a greater degree of similarity. These pairs are ranked accordingly, and up to 20 top potential matches are presented for the end-user to accept or reject as valid. Wild-ID is implemented in Java, so it can be run on Windows, MacOS, and Linux platforms.

HotSpotter (Crall et al. 2013) builds on this method, extending it to a 2-algorithm approach. The first algorithm extracts SIFT features and performs one-versus-one image comparisons similar—with minor improvements—to the algorithm used by Wild-ID. The second algorithm uses Local Naïve Bayes Nearest Neighbor methods (McCann and Lowe 2012) to perform quick one-versus-many comparisons with an entire database of images to identify images with similar groupings of features. The program assigns a similarity score to each pairing, based on the results of both algorithms, with a larger score indicating greater similarity. HotSpotter initially presents the end user with thumbnails of the 6 highest scores for review, but lower ranked scores are also viewable. As with Wild-ID, the end-user makes the final determination to accept a match or not. In addition to the standalone package evaluated here, HotSpotter's algorithms and methods have also been incorporated into the citizen science platform, Wildbook (Berger-Wolf et al. 2017). Versions of HotSpotter are available for both Windows and MacOS, and source code is available for Linux.

The goal of our study was to evaluate the effectiveness of using pattern-recognition software to identify individual animals in camera-trap photos. To accomplish this, we had several objectives: 1) use 2 distinct data sets—1 of jaguars and 1 of ocelots (*Leopardus pardalis*)—to test both HotSpotter and Wild-ID and quantify their accuracy for large versus small species with photos of varying quality; 2) compare results between programs and data sets, including statistical evaluation of any observed differences; given the refined, multi-algorithm method HotSpotter uses, we predicted that HotSpotter would prove more accurate than Wild-ID; and 3) provide usage recommendations for both programs specifically and pattern-recognition software in general.

STUDY AREA

Both of our photo data sets of jaguars and ocelots were assembled from images collected during the same long-term, noninvasive, camera-trapping study in 4 sites in Orange Walk District, Belize, Central America: Hill Bank, La Milpa, Gallon Jug, and Yalbac. All 4 sites primarily consisted of lowland broadleaf tropical forest, with some relatively small areas of pine savannah. Hill Bank and La Milpa together formed the 1,052-km^2 Rio Bravo Conservation and Management Area (RBCMA), managed by Programme for Belize. Gallon Jug Estate and the Yalbac Ranch and Cattle Company border the RBCMA to the south and west. Yalbac Ranch and Cattle Company, managed by the Forestland Group, occupied 849 km^2, including 423 km^2 purchased from Gallon Jug in 2012. These lands now completely surround Gallon Jug Estate, which now comprises 113 km^2 after the sale. See Satter et al. (2019*b*) for additional details about the study area.

METHODS

We collected photo-captures across the 4 sites from April to October 2015 and May to September 2016. Each study site consisted of 10–36 camera stations with paired trail cameras facing each other across trails, old logging roads, or main roads. Camera stations were placed at 2–3-km intervals following previously established methods for camera-trap studies of felids in Belize (Kelly 2003, Silver et al. 2004, Harmsen 2009). Cameras operated 24 hours/day for 2–3 months/camera station.

We tested both HotSpotter and Wild-ID using 2 distinct groups of these images, 1 set of jaguars and 1 set of ocelots, to determine whether the program worked equally well on large versus small species. For each species, we assembled a group of test images and a group of reference images (Table 1). Initially, both reference images and test images were manually identified by eye to individual animals. We explicitly entered the identities of the reference images in the software so the test images could be compared with them, while we had the programs treat the test images as unknown. We knew the true identities in advance, so we were able to evaluate the accuracy of the matches suggested by the programs for each test image.

The distinction between reference and test images is not intrinsic to either program. Both compared the entire set of images (reference and test) and ranked the best matches regardless of the image set to which they belonged. In many cases, a given side of a given individual was represented multiple times in the test image set, and sometimes the top match would be to another test image, rather than a reference image. As long as both were in fact the same side of the same individual, we considered this a positive match. HotSpotter includes an internal mechanism for cross-referencing matched images, so that even such test–test matches were ultimately linked to a reference image and ID. In Wild-ID, this cross-referencing had to be done separately (e.g., with an external spreadsheet). Inclusion of a reference set was a convention that we adopted to ease identification—as opposed to merely matching—and to ensure that every image would have ≥1 other image of the same side of the same individual to which it could potentially be matched.

Table 1. The quantity of reference images and test images of jaguars (A) and ocelots (B)—totaled by image quality—used to assess HotSpotter and Wild-ID. The identities of reference images were entered into the programs in advance, whereas test images were those that the programs attempted to match. Jaguar images were collected from camera traps deployed in Belize, in Gallon Jug, Yalbac, Hill Bank, and La Milpa during the summers of 2015 and 2016. Ocelot images were collected from camera traps deployed in Gallon Jug, Belize, in the summer of 2016.

Species	Image quality	Reference images	Test images
A) Jaguars	Total	321	359
	High	143	136
	Medium	95	112
	Low	83	111
B) Ocelots	Total	171	332
	High	56	110
	Medium	58	128
	Low	57	94

For ocelots, we selected test images only from data collected in Gallon Jug in 2016. To obtain roughly the same number of test images of jaguars, and because photo-capture rates for jaguars were much lower than ocelots, we selected jaguar test images from data collected in 2015 from Hill Bank, La Milpa, Gallon Jug, and Yalbac, and in 2016 from Hill Bank, Gallon Jug, and Yalbac. We ranked the quality of all images on a 3-point scale: 1 for blurred, low-quality images, 2 for medium-quality images, and 3 for crisp, high-quality images (Fig. 1; Table 1).

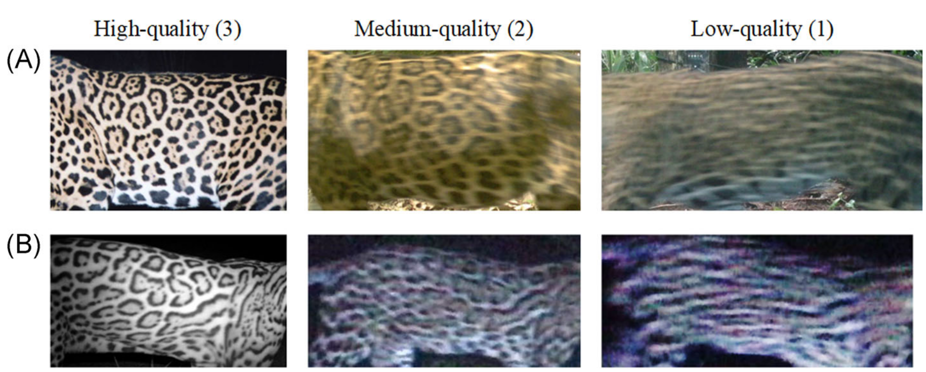

Figure 1. Representative examples of high- (3), medium- (2), and low- (1) quality images of jaguars (A) and ocelots (B) collected from camera traps deployed in 4 study sites in Orange Walk District, Belize, in 2015 and 2016. Images were classified as high-quality when images were crisp and fine details of spots were clear. Medium-quality images were moderately blurry, but the general shape of most spots was still recognizable. Low-quality images were heavily blurred and individual spots were not clearly discernable. The images shown are all of the same individual jaguar and ocelot, respectively.

We assembled ocelot reference images for all individuals that had ever been photographed in Gallon Jug since we began surveying there in 2013, and jaguar reference images for all individuals ever photographed in any of the 4 sites since we began surveying them (since 2008 in La Milpa, 2009 in Hill Bank, 2013 in Gallon Jug, and 2014 in Yalbac). For both species, we included—to the extent possible—reference images of each quality level of each side of each individual. In some cases, though, images were only available for one side of an individual, or there were no images of some quality levels. In the few cases for which reference images of one side of an individual were unavailable, this was due to us never having captured images of that side of that individual and hence our test images also did not include examples of the missing side of those individuals. Thus, in terms of assessing the matching success of these programs, the lack of both sides did not impact our results.

To minimize background objects while maximizing the visibility of distinct pelage patterns, we cropped all images to a rectangular area just around each animal's flank. HotSpotter includes a built-in feature that effectively accomplished this by allowing a user to define a rectangular "chip" (Fig. 2) specifying a region of interest (ROI) to which the algorithm would be limited. For Wild-ID, we externally cropped images using third-party image-editing software (e.g., Microsoft Paint [Microsoft Corporation, Redmond, WA, USA] or Adobe Photoshop [Adobe, San Jose, CA, USA]) before loading the images for analysis. Image processing software is widely and—for some packages—freely available, but this did make preprocessing images for Wild-ID more cumbersome than for HotSpotter.

We classified the result for each test image in each program as either a positive match, a false match, or a nonmatch. We considered the outcome a positive match when the top match suggested by the program was in fact another image of the same side of the same individual. A false match occurred when an accurate match was included somewhere in the top 10 suggested matches, but other images that were not of the same side of the same individual were ranked higher. We considered the result a nonmatch when the program did not include an image of the same side of the same individual anywhere in its top 10 suggested matches.

We summarized results as proportions, grouped by species, program, and image quality. We used the plus 4 method of Agresti and Coull (1998) to calculate 95% confidence intervals for these proportions and tested for significant differences ($\alpha = 0.05$) between the proportions at each quality level (Baldi and Moore 2014). For positive matches, we also recorded the photo quality of the matching images and calculated percentages grouped by species, program, and test image quality. For jaguar images only, we also recorded the total time to complete the matching process. This did not include the time needed to assemble the database of reference images, nor did it include any preprocessing time (i.e., cropping images for Wild-ID or defining ROIs in HotSpotter), only the time for each program to calculate and rank matches and the user to review them.

RESULTS

In HotSpotter, the proportion of positive matches for all image qualities combined was 0.77 (95% CI = 0.73–0.82) for jaguars and 0.76 (95% CI = 0.71–0.80) for ocelots, while in Wild-ID it was 0.68 (95% CI = 0.63–0.73) for jaguars and 0.63 (95% CI = 0.58–0.68) for ocelots (Fig. 3; Appendix A, available online in Supporting Information). High-quality test images had the greatest rate of positive matches for both image sets in both programs (observed proportions = 0.85–0.99). Medium-quality test images had lower positive match rates (observed proportions = 0.70–0.88). Low-quality images had the lowest positive match rates (observed proportions = 0.28–0.52).

HotSpotter consistently had greater positive match rates than Wild-ID for both data sets and all quality levels (Fig. 3). These differences were significant (Appendix B, available online in Supporting Information) for jaguar images of high-quality (0.10 greater than Wild-ID, 95% CI = 0.02–0.17) and for all jaguar images overall (0.09 greater than Wild-ID, 95% CI = 0.03–0.16), and for ocelot images that were high-quality (0.13 greater than Wild-ID, 95% CI = 0.06–0.19), medium-quality (0.19

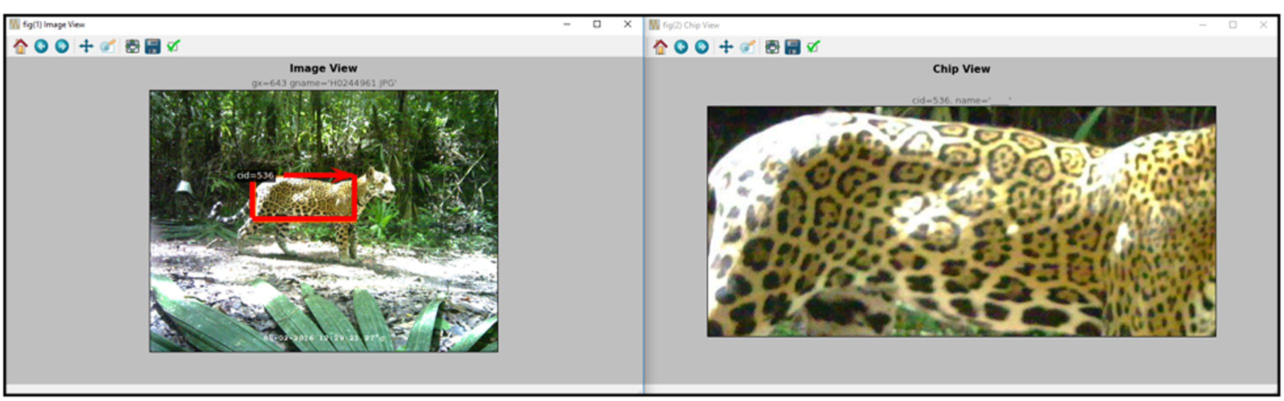

Figure 2. An example of defining a region of interest in HotSpotter software for identifying jaguars and ocelots from photos collected from camera traps deployed in 4 study sites in Orange Walk District, Belize, in 2015 and 2016. Both HotSpotter and Wild-ID required that the original image be reduced to minimize the possibility of their algorithms attempting to match patterns in the background. HotSpotter includes a built-in feature to do so, whereas Wild-ID required the use of third-party image-editing software to crop images ahead of time.

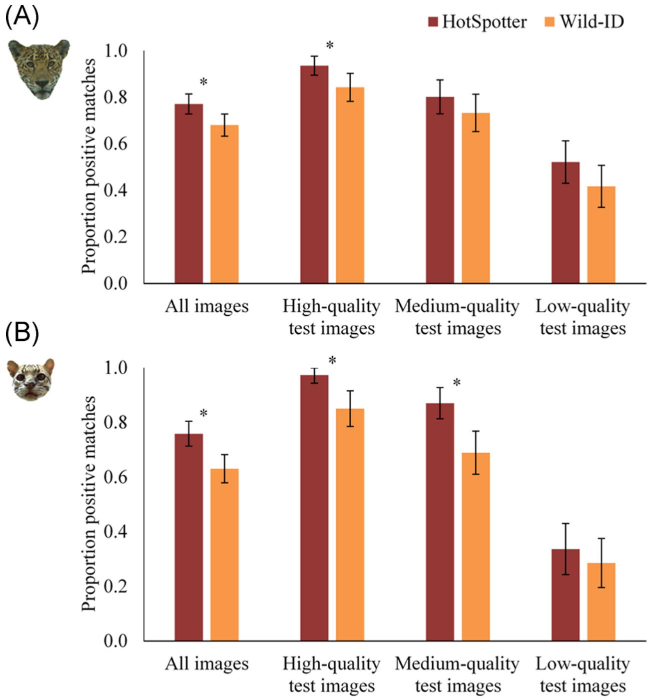

Figure 3. Proportions with 95% confidence intervals of positive matches for images of jaguars (A) and ocelots (B) collected from camera traps deployed in 4 study sites in Orange Walk District, Belize, in 2015 and 2016, grouped by image quality and tests in HotSpotter or Wild-ID. Positive matches were defined as cases in which the top-ranked match for a test image was in fact another image of the same side of the same individual. * Indicates a significant difference ($\alpha = 0.05$).

greater than Wild-ID, 95% CI = 0.09–0.28), and for all ocelot images overall (0.13 greater than Wild-ID, 95% CI = 0.06–0.20). There were slight differences in a given program between species (e.g., a lower positive match rate for high-quality jaguar images in HotSpotter compared with high-quality ocelot images in HotSpotter), but most of these differences were not significant (Appendices A, C, available online in Supporting Information). The exceptions were the low-quality image sets; in both programs, positive match rates for low-quality ocelots were significantly lower than for low-quality jaguars (0.19 lower in HotSpotter, 95% CI = 0.06–0.32; and 0.14 lower in Wild-ID, 95% CI = 0.01–0.26).

False match rates (Fig. 4; Appendix A) for HotSpotter were very low (observed proportions = 0.00–0.02) for both jaguars and ocelots, while Wild-ID's false match rates were significantly greater (observed proportions = 0.06–0.28) across all categories. False match rates in Wild-ID were greatest for low-quality images at 0.29 (95% CI = 0.20–0.37) for jaguars and 0.17 (95% CI = 0.10–0.25) for ocelots. False matches were reduced for medium-quality images at 0.13 (95% CI = 0.07–0.19) for jaguars and 0.14 (95% CI = 0.08–0.20) for ocelots, and further reduced for high-quality images at 0.08 (95% CI = 0.03–0.12) for jaguars and 0.08 (95% CI = 0.03–0.13) for ocelots.

There was no consistent trend with respect to nonmatches (Fig. 5; Appendix A). For jaguar images, HotSpotter had a slightly greater nonmatch rate than Wild-ID for most

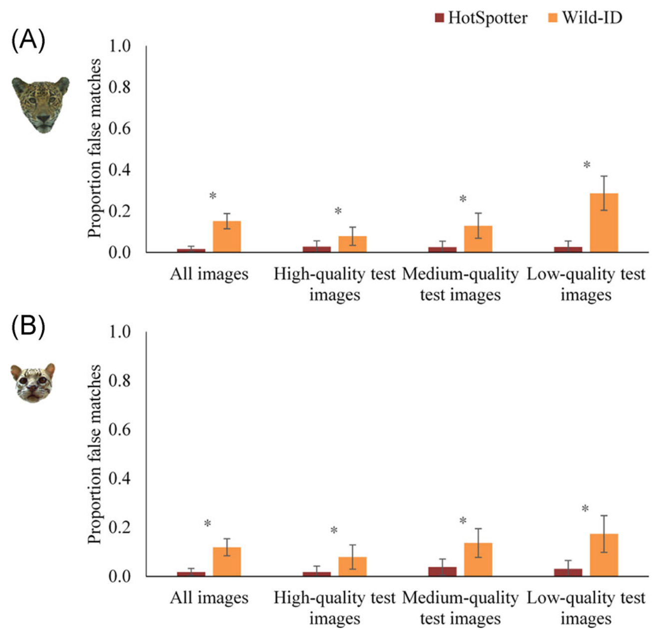

Figure 4. Proportions with 95% confidence intervals of false matches for images of jaguars (A) and ocelots (B) collected from camera traps deployed in 4 study sites in Orange Walk District, Belize, in 2015 and 2016, grouped by image quality and tests in HotSpotter or Wild-ID. False matches were defined as cases in which another image of the same side of the same individual was somewhere among the top 10 matches suggested by the program, but ≥1 incorrect match was ranked higher. * Indicates a significant difference ($\alpha = 0.05$).

categories, but this difference was only significant for low-quality test images ($z = 2.48$, $P = 0.013$; 0.16 greater than Wild-ID, 95% CI = 0.03–0.28). The nonmatch rate for high-quality jaguar images, though, was 0.04 (95% CI = 0.01–0.09)

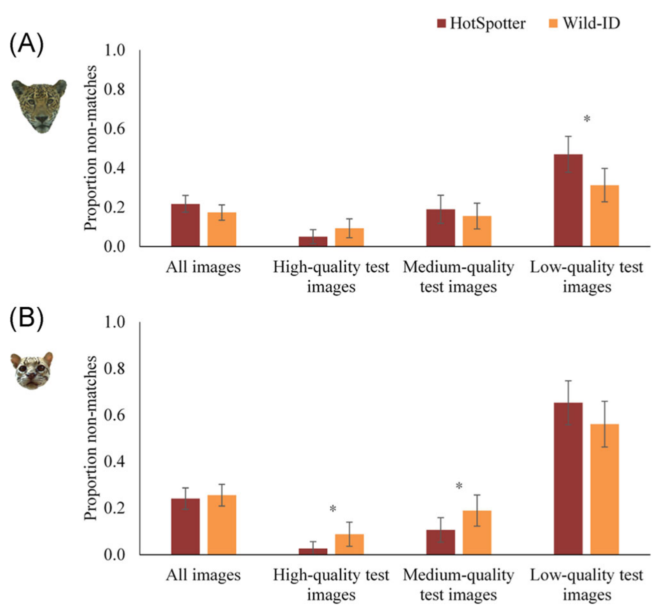

Figure 5. Proportions with 95% confidence intervals of nonmatches for images of jaguars (A) and ocelots (B) collected from camera traps deployed in 4 study sites in Orange Walk District, Belize, in 2015 and 2016, grouped by image quality and tests in HotSpotter or Wild-ID. Nonmatches were defined as cases in which another image of the same side of the same individual was not among any of the top 10 matches suggested by the program. * Indicates a significant difference ($\alpha = 0.05$).

in HotSpotter, compared with 0.08 (95% CI = 0.05–0.14) in Wild-ID. Conversely, the nonmatch rate for ocelot images was generally greater in Wild-ID. For high-quality test images, the nonmatch rate was 0.06 (95% CI = 0.01–0.12) lower for HotSpotter than Wild-ID ($z = -2.38$, $P = 0.017$), and for medium-quality test images, HotSpotter's nonmatch rate was 0.09 (95% CI = 0.00–0.17) lower than Wild-ID ($z = -2.00$, $P = 0.045$). HotSpotter's nonmatch rate for low-quality ocelot images (0.65, 95% CI = 0.56–0.75) was greater than Wild-ID's (0.56, 95% CI = 0.46–0.66), but this difference was not significant. Nonmatch rates were greatest among low-quality images for both species in both programs (observed proportions = 0.31–0.66) and lowest among high-quality images (observed proportions = 0.01–0.08).

We found that across most quality categories in both species and both programs, the majority of the top-ranked positive matches were matched to images of the same quality as the test images (Fig. 6). This was especially pronounced in all high-quality test images and low-quality jaguar test images, in which 63–90% of the top matches were images of the same quality. For medium-quality test images, there was a larger proportion of top matches of different quality (primarily matches to high-quality images), but for 3 of the 4 groups (all except ocelot images in HotSpotter) there were still more top matches to medium images than to both other categories combined. Among medium-quality ocelot images in HotSpotter, other medium-quality images still account for the single greatest category of positive matches (47.8%) but matches to low- and high-quality images together accounted for 52.2% of positive matches. Low-quality ocelot images in Wild-ID were similar in that the majority of matches (53.9%) were to images of the same quality, but there were a larger proportion of matches to other qualities as well (34.6% to medium-quality images and 11.5% to high-quality images). Low-quality ocelot images in HotSpotter were the major exception to these trends, with only 35.5% of positive matches to other low-quality images while 45.2% of the matches were to medium-quality images and 19.4% were to high-quality images.

Measurements of the time to complete the matches were based on each program processing 706 jaguar images (reference and test images). We subsequently removed 26 of those test images from the rest of our analysis because either they were individuals we had never detected before so there were no other images to which they could possibly be matched, or because we deemed the images unmatchable (e.g., only the tip of a tail was visible). In either case, including those images would have negatively biased our results and would not have been an accurate reflection of the fundamental matching abilities of the programs, so those 26 test images are not included in any of the other results above. In HotSpotter, it took 4 hours and 44 minutes to

Figure 6. Percent and 95% confidence intervals of top positive matches by quality for each quality of jaguar test image in HotSpotter (A) and Wild-ID (C) and each quality of ocelot test image in HotSpotter (B) and Wild-ID (D). Images collected from camera traps deployed in 4 study sites in Orange Walk District, Belize, in 2015 and 2016.

complete matching the 706 images. In Wild-ID, it took 17 hours and 50 minutes to complete matching the same images. It should be noted that we conducted tests in Wild-ID first and we became more efficient by the time we conducted the tests in HotSpotter. So, although Wild-ID's computing time was slower, the difference noted here is likely inflated.

DISCUSSION

Our positive match rates are lower than those reported by Bolger et al. (2012) or Crall et al. (2013) when presenting Wild-ID and HotSpotter, respectively, but are broadly consistent with the findings of other literature assessing the programs. For example, Mettouris et al. (2016) used Wild-ID to match high-quality images of 2 different newt species and found that the top-ranked match was correct 22–99% of the time, and a correct match was somewhere in the top 20 candidates 100% of the time. Cruickshank and Schmidt (2017), analyzing high-quality images of yellow-bellied toads (*Bombina variegata*), found that Wild-ID presented a correct, top-ranked match 91.6% of the time and that a correct match was within the top 2 candidates 95% of the time. Suriyamongkol and Mali (2018) studied Rio Grande cooter (*Pseudemys gorzugi*) and found that Wild-ID offered a correct match as the top choice 61–66% of the time, and that a correct match was somewhere in the top 20 ranks 77–84% of the time.

Our positive match rates only considered the top match and not top 10–20; therefore, our results may understate the true benefits of the programs. In an applied context, even when the top image is not a positive match, if a correct match is somewhere among the top choices, then the program has effectively reduced the number of images that a researcher must review by eye, thereby still saving substantial time and effort. However, Matthé et al. (2017) found the performance of different programs may vary dramatically between species, and Morrison et al. (2016*a*), found that neither Wild-ID nor HotSpotter performed adequately for their data set. These differences in performance may be partially due to differences in markings across species. Both jaguars and ocelots bear prominent, distinct markings; however, even among other felid species—let alone other taxa—there is substantial variation in the presence, shape, and intensity of markings. These programs may not perform as well for species with less distinct markings, and certainly would not be appropriate for species lacking natural markings.

Our findings also suggest that images are more likely to match other images of the same quality, with the exception of low-quality ocelot images in HotSpotter. It is likely that the low-quality ocelot images we selected were exceptionally poor, even compared with the low-quality jaguar images. Sometimes we used multiple images taken in a photo-burst to identify an individual but may have used the lowest quality image in that burst in the matching program. Such extremely low-quality images were also present among the jaguar images but were likely more prevalent among the ocelot images, possibly because of their smaller size or differences in their markings or movement patterns. This likely also accounts for the significantly lower positive match rate in both programs for low-quality ocelot images versus low-quality jaguars.

Notwithstanding this inconsistency, the broad takeaway remains the same: both programs were generally more successful matching images of similar quality, so including different quality images of the same individual should improve matching rates. Matthé et al. (2017) also concluded that the performance of multiple programs improved when more potential matches were available, though they did not specifically examine the effects of image quality. This may be less important for data sets containing only high-quality photos—such as those collected with a handheld camera under relatively controlled conditions—but it has important implications about how to use software most effectively with camera-trap data (or any other data set with substantial variation in image quality).

This observation is specific to image quality, but the concept could be extended to apply to the selection of reference images in other ways. In camera-trapping data, it is fairly common to obtain photographs of only part of an animal (e.g., only the head, or tail). Similarly, images of a moving animal might be blurry and low-quality around the flank, but the legs are often relatively clear. Increasing the size of a reference set to include other parts of individuals could improve accuracy but also would increase both the preprocessing time and the actual computing time. It might be possible to mitigate this issue by creating separate reference databases for different parts of the animal (e.g., a reference set just for heads or just for medial, right forelegs). Test images would only need to be processed with the applicable database.

We found both programs to be beneficial tools that improved the efficiency of identifying individual animals, both large and small species, in camera-trap data. Considering only positive match rates, HotSpotter consistently performed better; it had greater positive match rates than Wild-ID across all quality categories we examined, and those differences were significant in 5 out of the 8 categories. HotSpotter's negligible rate of false matches indicates that most of the time, if it is able to make a match at all, that match will be the top-ranked suggestion. Wild-ID's higher false match rate requires researchers to spend additional time manually comparing the top images to find some matches, but this is still substantially more expedient than reviewing all possible matches by eye. In this sense, a lower nonmatch rate is a more comprehensive metric of overall matching success. Rates of nonmatches were more similar between the programs, with only 3 of the 8 differences being significant. In some cases (e.g., low-quality jaguar images), Wild-ID actually had lower nonmatch rates than HotSpotter.

Neither program is fully automated; even though the positive match rates are good—especially for higher quality images—they are not foolproof. Human involvement is still necessary to account for circumstances when the algorithms fail—especially for lower quality images. Nonetheless, HotSpotter is closer to an automatic system in that often only the top suggested match needs to be reviewed. Thus, it may be better suited to users with little experience visually matching individuals, or to studies in which matching speed is a priority. However, for users with more experience at visual identification, Wild-ID may in some cases result in more overall matches (especially for data sets with a large number of low-quality images). The tradeoff would be more time to conduct the analysis because some of those matches would require reviewing more candidate images within Wild-ID. Sometimes it may even be beneficial to use Wild-ID to reanalyze images that HotSpotter could not match. If some of those images at least resulted in false matches in Wild-ID (so that a researcher could identify a match from a reduced subset of images), this may still save time compared with matching by eye alone all of the images that HotSpotter failed to identify. However, the increased time to process some images in both programs means that this would likely only be beneficial in very large data sets.

Beyond considerations of matching performance, HotSpotter includes several additional features that enhance its utility. The ability to define chips within the program eliminates the need to switch between third-party programs and expedites image preprocessing. Its interface is also more interactive, allows users to view the image features being extracted, and provides several options for users to visualize how those features are being compared between images (Appendix D; available online in Supporting Information). Wild-ID simply imports a directory full of images and compares them all against each other; the user may choose to include a set of reference images in this directory, but the algorithm reprocesses these (i.e., re-extracts features and compares them between images—including other reference images) every time another analysis is run. For a long-term, constantly growing data set, this inefficiency adds up over time. By comparison, once a user has identified a match in HotSpotter, an image can be assigned a name, which then adds it to a database of known individuals. Images with the same name are clearly cross-referenced to each other. New images can then be imported and compared with the existing database without the need to reprocess old images. This also allows the database to grow and evolve over time by adding new individuals or new images of existing individuals. In HotSpotter, it is easier to go back and review a previous set of matches for an image whereas Wild-ID only records the accepted match for an image (and does not cross-reference this with any other images), so the only way to review other previous matches is to rerun the entire analysis.

The time it took for each program to match all jaguar images may provide one rough point of comparison between the programs, but we offer several words of caution when considering this. The time to match a given image varies substantially depending on image quality (i.e., how long the user has to spend looking at potential matches before deciding there is or is not a match), so a different set of images could produce different results even if the total number of images were the same. In application, considerations of the time needed to use the programs should also take into account the substantial time needed to assemble a reference database and preprocess images, which we do not include here. Finally, any fundamental difference between the time it takes each program to match images probably does not scale linearly; this is due to differences in the ways the programs calculate matches—for small sets of images, the times may actually be fairly close, but as the number of images grows, any difference likely becomes increasingly pronounced.

Usage Recommendations

As camera-trapping studies continue to proliferate, computer-aided systems to process the resultant data will also continue to grow. Wild-ID has already been successfully used in a relatively large number of studies, and our results suggest that HotSpotter may offer even greater benefits in the future. Our results suggest several best-practices that apply regardless of specific objectives or the particular pattern-recognition program used. To account for potential variation in performance for different study species, we echo the recommendation of Matthé et al. (2017) that researchers conduct a trial analysis with a small subset of their own data before fully committing to any particular algorithm or attempting a large-scale analysis. Both programs require the user to make the final determination confirming or denying a match, so neither is able to fully automate the individual identification process and researchers should budget time accordingly. Finally, building a reference database that includes multiple images of a given individual should improve overall matching rates.

Additionally, our results suggest a number of situational recommendations, depending on researchers' particular objectives. First, users should consider the nature of their photo-data. For images of fairly uniform quality taken under controlled conditions, pixel-based software packages such as those reviewed by Matthé et al. (2008, 2017) and Moya et al. (2015) may provide better performance than either of the feature-based programs reviewed here. For images with substantial variation in quality or alignment—including camera-trapping images—these feature-based programs are likely more appropriate. In this latter case, it becomes especially important to include in the reference database not only multiple images of each individual, but a range of image qualities for each. For short-term studies in which a single set of images is to be analyzed a single time, it may not be necessary to build a true reference database and Wild-ID may be adequate for simply matching all of the images against each other. However, for longer term or ongoing studies in which multiple sets of images will be analyzed and individuals in newer images compared with previous data, HotSpotter's ability to create and easily add individuals to an ongoing database of images is especially powerful.

Researchers also should consider several tradeoffs between time and accuracy. For cases in which minimizing time is more important than maximizing the programs' matching accuracy, HotSpotter seems to be faster in raw processing time, though this may be highly variable. In such circumstances—especially for researchers experienced at matching images by eye, who can be relatively efficient at matching images the program cannot—it may also be beneficial to minimize the repeat images for each individual in the database in order to reduce preprocessing and computing times. Conversely, if maximizing program accuracy is more important than saving time, or if users are less experienced at matching images by eye, expanding the database could offer several improvements. In this case, HotSpotter's ability to create and maintain true databases would again be invaluable. Users could expand a database—or create separate databases—to include images of individuals at different orientations, or other parts of individuals like tails, legs, or heads that may still allow identification even if a flank is heavily blurred or not visible in some idiosyncratic images. In some circumstances, accuracy might be maximized further by employing both programs. Images could initially be processed in HotSpotter, then images that it failed to match could subsequently be processed in Wild-ID. Any of these strategies would entail increased preprocessing and computing time but should also maximize the accuracy of the programs. This evaluation provides guidance to researchers—especially those with camera-trap data—about how to use such technology effectively as it continues to evolve.

ACKNOWLEDGMENTS

We thank the Forest Department of Belize for permission to conduct this work, Programme for Belize (especially E. Romero and R. Pacheco), Gallon Jug Estate (especially A. Jeal), Yalbac Ranch and Cattle Corporation Limited (especially J. Roberson), and the Forestland Group for assistance and access to work on their lands. We thank the University of Belize for connecting interns with this project. We thank M. Aguilar, N. Quetzal, J. Alvarado, and the Vasquez family for local knowledge, expertise, and assistance, and the many volunteers and assistants (especially D. McNitt, A. Bellows, and D. McPhail) who conducted field work and data processing.

LITERATURE CITED

Agresti, A., and B. A. Coull. 1998. Approximate is better than "exact" for interval estimation of binomial proportions. The American Statistician 52:119–126.

Baldi, B., and D. S. Moore. 2014. The practice of statistics in the life sciences. Third edition. W. H. Freeman and Company, New York, New York, USA.

Barord, G. J., F. Dooley, A. Dunstan, A. Ilano, K. N. Keister, H. Neumeister, T. Preuss, S. Schoepfer, and P. D. Ward. 2014. Comparative population assessments of *Nautilus* sp. in the Philippines, Australia, Fiji, and American Samoa using baited remote underwater video systems. PLoS ONE 9(6):e100799.

Bendik, N. F., T. A. Morrison, A. G. Gluesenkamp, M. S. Sanders, and L. J. O'Donnell. 2013. Computer-assisted photo identification outperforms visible implant elastomers in an endangered salamander, *Eurycea tonkawae*. PLoS ONE 8(3):e59424.

Berger-Wolf, T. Y., D. I. Rubenstein, C. V. Stewart, J. A. Holmberg, J. Parham, S. Menon, J. Crall, J. Van Oast, E. Kiciman, and L. Joppa. 2017. Wildbook: crowdsourcing, computer vision, and data science for conservation. Presented at the Data For Good Exchange 2017 meeting, 24 September 2017, New York, New York, USA. <https://arxiv.org/abs/1710.08880>. Accessed 1 Mar 2019.

Bolger, D. T., T. A. Morrison, B. Vance, D. Lee, and H. Farid. 2012. A computer-assisted system for photographic mark–recapture analysis. Methods in Ecology and Evolution 3:813–822.

Calmanovici, B., D. Waayers, J. Reisser, J. Clifton, and M. Proietti. 2018. I^3S Pattern as a mark–recapture tool to identify captured and free-swimming sea turtles: an assessment. Marine Ecology Progress Series 589:263–268.

Cochran, J. E. M., R. S. Hardenstine, C. D. Braun, G. B. Skomal, S. R. Thorrold, K. Xu, M. G. Genton, and M. L. Berumen. 2016. Population structure of a whale shark *Rhincodon typus* aggregation in the Red Sea. Journal of Fish Biology 89:1570–1582.

Crall, J. P., C. V. Stewart, T. Y. Berger-Wolf, D. I. Rubenstein, and S. R. Sundaresan. 2013. HotSpotter—patterned species instance recognition. Pages 230–237 *in* 2013 IEEE Workshop on Applications of Computer Vision (WACV). Institute of Electrical and Electronics Engineers, 15–17 January 2013, Tampa, Florida, USA.

Creel, S., G. Spong, J. L. Sands, J. Rotella, J. Zeigle, L. Joe, K. M. Murphys, and D. Smith. 2003. Population size estimation in Yellowstone wolves with error-prone noninvasive microsatellite genotypes. Molecular Ecology 12:2003–2009.

Cross, M. D., G. J. Lipps, Jr., J. M. Sapak, E. J. Tobin, and K. V. Root. 2014. Pattern-recognition software as a supplemental method of identifying individual eastern box turtles (*Terrapene c. carolina*). Herpetological Review 45:584–586.

Cruickshank, S. S., and B. R. Schmidt. 2017. Error rates and variation between observers are reduced with the use of photographic matching software for capture–recapture studies. Amphibia-Reptilia 38:315–325.

Davis, C. L., E. W. Teitsworth, and D. A. W. Miller. 2018. Combining data sources to understand drivers of spotted salamander (*Ambystoma maculatum*) population abundance. Journal of Herpetology 52:116–126.

Dinsmore, S. J., and D. H. Johnson. 2012. Population analysis in wildlife biology. Pages 349–380 *in* N. J. Silvy, editor. The wildlife techniques manual: research. Volume 1. Seventh edition. Johns Hopkins University Press, Baltimore, Maryland, USA.

Harmsen, B. J. 2009. The use of camera traps for estimating abundance and studying the ecology of jaguars (*Panthera onca*). Dissertation, University of Southampton, Southampton, England, United Kingdom.

Hiby, L., and P. Lovell. 1990. Computer aided matching of natural markings: a prototype system for grey seals. Reports of the International Whaling Commission, Special Issue 12:3–17.

Jablons, Z. 2016. Identifying humpback whale flukes by sequence matching of trailing edge curvature. Thesis, Rensselaer Polytechnic Institute, Troy, New York, USA.

Kelly, M. J. 2001. Computer-aided photograph matching in studies using individual identification: an example from Serengeti cheetahs. Journal of Mammalogy 82:440–449.

Kelly, M. J. 2003. Jaguar monitoring in the Chiquibul Forest, Belize. Caribbean Geography 13:19–32.

Kelly, M. J., J. Betsch, C. Wultsch, B. Mesa, and L. S. Mills. 2012. Noninvasive sampling for carnivores. Pages 47–69 *in* L. Boitani and R. A. Powell, editors. Carnivore ecology and conservation: a handbook of techniques. Oxford University Press, Oxford, England, United Kingdom.

Lea, J. M. D., G. I. H. Kerley, H. Hrabar, T. J. Barry, and S. Shultz. 2016. Recognition and management of ecological refugees: a case study of the Cape mountain zebra. Biological Conservation 203:207–215.

Lee, D. E., M. L. Bond, and D. T. Bolger. 2017. Season of birth affects juvenile survival of giraffe. Population Ecology 59:45–54.

Lingaraja, S. S., S. Chowdhary, R. Bhat, and S. Gubbi. 2017. Evaluating a survey landscape for tiger abundance in the confluence of the Western and Eastern Ghats. Current Science 113:1759–1763.

Lowe, D. G. 2004. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60:91–110.

Lukacs, P. M., and K. P. Burnham. 2005. Review of capture–recapture methods applicable to noninvasive genetic sampling. Molecular Ecology 14:3909–3919.

Matthé, M., M. Sannolo, K. Winiarski, A. Spitzen-van der Sluijs, D. Goedbloed, S. Steinfartz, and U. Stachow. 2017. Comparison of photo-matching algorithms commonly used for photographic capture–recapture studies. Ecology and Evolution 7:5861–5872.

Matthé, M., T. Schönbrodt, and G. Berger. 2008. Computergestutzte bildanalyse von bauchfleckenmustern des kammmolchs (*Triturus cristatus*). Zeitschrift für Feldherpetologie 15:89–94. [In German.]

McCann, S., and D. G. Lowe. 2012. Local naïve Bayes nearest neighbor for image classification. Pages 3650–3656 *in* 2012 IEEE Conference on Computer Vision and Pattern Recognition. Institute of Electrical and Electronics Engineers, 16–21 June 2012, Providence, Rhode Island, USA.
Mettouris, O., G. Megremis, and S. Giokas. 2016. A newt does not change its spots: using pattern mapping for the identification of individuals in large populations of newt species. Ecological Research 31:483–489.
Miller, J. R. B., R. T. Pitman, G. K. H. Mann, A. K. Fuller, and G. A. Balme. 2018. Lions and leopards coexist without spatial, temporal or demographic effects of interspecific competition. Journal of Animal Ecology 87:1709–1726.
Millspaugh, J. J., R. A. Gitzen, J. L. Belant, R. W. Kays, B. J. Keller, D. C. Kesler, C. T. Rota, J. H. Schultz, and C. M. Bodinof. 2012. Analysis of radiotelemetry data. Pages 480–501 *in* N. J. Silvy, editor. The wildlife techniques manual: research. Volume 1. Seventh edition. Johns Hopkins University Press, Baltimore, Maryland, USA.
Morrison, T. A., D. Keinath, W. Estes-Zumpf, J. P. Crall, and C. V. Stewart. 2016*a*. Individual identification of the endangered Wyoming toad *Anaxyrus baxteri* and implications for monitoring species recovery. Journal of Herpetology 50:44–49.
Morrison, T. A., W. A. Link, W. D. Newmark, C. A. H. Foley, and D. T. Bolger. 2016*b*. Tarangire revisited: consequences of declining connectivity in a tropical ungulate population. Biological Conservation 197:53–60.
Morrison, T. A., J. Yoshizaki, J. D. Nichols, and D. T. Bolger. 2011. Estimating survival in photographic capture–recapture studies: overcoming misidentification error. Methods in Ecology and Evolution 2:454–463.
Moya, Ó., P. Mansilla, S. Madrazo, J. Igual, A. Rotger, A. Romano, and G. Tavecchia. 2015. APHIS: a new software for photo-matching in ecological studies. Ecological Informatics 27:64–70.
Otis, D. L., K. P. Burnham, G. C. White, and D. R. Anderson. 1978. Statistical inference from capture data on closed animal populations. Wildlife Monographs 62.
Pierce, B. L., R. R. Lopez, and N. J. Silvy. 2012. Estimating animal abundance. Pages 284–310 *in* N. J. Silvy, editor. The wildlife techniques manual: research. Volume 1. Seventh edition. Johns Hopkins University Press, Baltimore, Maryland, USA.
Royle, J. A., R. B. Chandler, B. Gardner, and R. Sollmann. 2013. Spatial capture–recapture. Elsevier Science and Technology, San Diego, California, USA.
Satter, C. B., B. C. Augustine, B. J. Harmsen, R. J. Foster, and M. J. Kelly. 2019*a*. Sex-specific population dynamics of ocelots in Belize using open population capture–recapture. Ecosphere 10(7):e02792.
Satter, C. B., B. C. Augustine, B. J. Harmsen, R. J. Foster, E. E. Sanchez, C. Wultsch, M. L. Davis, and M. J. Kelly. 2019*b*. Long-term monitoring of ocelot densities in Belize. Journal of Wildlife Management 83:283–294.
Silver, S. C., L. E. T. Ostro, L. K. Marsh, L. Maffei, A. J. Noss, M. J. Kelly, R. B. Wallace, H. Gomez, and G. Ayala. 2004. The use of camera traps for estimating jaguar *Panthera onca* abundance and density using capture/recapture analysis. Oryx 38:148–154.
Strauss, M. K. L., M. Kilewo, D. Rentsch, and C. Packer. 2015. Food supply and poaching limit giraffe abundance in the Serengeti. Population Ecology 57:505–516.
Suriyamongkol, T., and I. Mali. 2018. Feasibility of using computer-assisted software for recognizing individual Rio Grande cooter (*Pseudemys gorzugi*). Copeia 106:646–651.
Tumulty, J. P., A. Pašukonis, M. Ringler, J. D. Forester, W. Hödl, and M. A. Bee. 2018. Brilliant-thighed poison frogs do not use acoustic identity information to treat territorial neighbors as dear enemies. Animal Behavior 141:203–220.
Van Tienhoven, A. M., J. E. Den Hartog, R. A. Reijns, and V. M. Peddemors. 2007. A computer-aided program for pattern-matching of natural marks on the spotted raggedtooth shark *Carcharias taurus*. Journal of Applied Ecology 44:273–280.
White, G. C., and K. P. Burnham. 1999. Program MARK: survival estimation from populations of marked animals. Bird Study 46:120–139.
Williams, B. K., J. D. Nichols, and M. J. Conroy. 2002. Analysis and management of animal populations: modeling, estimation and decision-making. Academic Press, San Diego, California, USA.
Yoshizaki, J., K. H. Pollock, C. Brownie, and R. Webster. 2009. Modeling misidentification errors in capture–recapture studies using photographic identification of evolving marks. Ecology 90:3–9.

Associate Editor: Blackwell.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article. Appendix A. Proportional results of program tests. Appendix B. Significance tests comparing programs. Appendix C. Significance tests comparing data sets. Appendix D. Hotspotter interface.