

Relevance Feedback in Biometric Retrieval of Animal Photographs

Chelsea Finn¹, James Duyck¹, Andy Hutcheon², Pablo Vera³,
Joaquin Salas³, and Sai Ravela¹

¹ Massachusetts Institute of Technology

² Department of Conservation, New Zealand

³ Instituto Politecnico Nacional, Mexico

Abstract. The characterization of individual animal life history is crucial for conservation efforts. In this paper, Sloop, an operational pattern retrieval engine for animal identification, is extended by coupling crowdsourcing with image retrieval. The coupled system delivers scalable performance by using aggregated computational inference to effectively deliver precision and by using human feedback to efficiently improve recall. To the best of our knowledge, this is the first coupled human-machine animal biometrics system, and results on multiple species indicate that it is a promising approach for large-scale use.

1 Introduction

Tracking an individual animal's life history can substantially benefit conservation efforts [1]. Current strategies based on Capture-Mark-Recapture (CMR) are intrusive to varying degrees. Therefore, when an animal pattern is recognizable, a pattern recognition approach is beneficial. Contemporary systems [1–10] cannot automatically adjudicate identity without significant misjudgment. To deliver the high recall needed in most applications, Ravela *et al.* [8, 11] advance a pattern retrieval approach wherein images ranked by similarity of image features extracted from photographs of animals are subject to human judgment in cohort discovery, *i.e.*, finding the set of images with shared identity. Sloop [1, 8–10] is a system and framework for animal identification that is deployed operationally and delivers high performance. This paper explores the scalability achieved by incorporating dynamic feedback from human relevance judgments, including the role of crowds in providing large-scale relevance feedback. Relevance feedback has received some attention in image indexing but not animal biometrics, and the algorithms presented here follow from earlier work [11].

One of the earliest approaches for individual animal identification involved the use of 3D deformable patterns [2, 4, 12]. Generic feature methods promise extensibility and thus have become popular, including multi-scale differential feature histograms and multiscale-PCA [1]. For example, SIFT and variants have been used for marbled salamanders [9] and manta rays [13]. Shape contexts have been used in African penguins (*Spheniscus demersus*) [14]. Some deployed

systems include ECOCEAN [15], the recognition of chimpanzees and gorillas [16], and African Penguins (*Spheniscus demersus*) [14].

In contrast, Sloop contains a set of reusable tools that continue to be deployed for multiple new species. It contains implementations that are extensible to new applications. Sloop is in operational use for the Otago skink and grand skink [17] at the Department of Conservation in New Zealand. In this paper, we primarily explore the value of relevance feedback in the context of several recognition techniques available within Sloop. These include relevance feedback with SIFT on grand skinks, Otago skinks, and jewelled geckos, with hybrid context aggregating feature positions and feature densities on whale sharks, and with hierarchical deformable matching on marbled salamanders. The results obtained on these species are highly promising. Additionally, we propose hierarchical and aggregated matching methods that also appear to be promising as generic recognition methods that ultimately can reduce human effort while relevance feedback can improve system performance, thus producing an extensible, scalable system for biological users.

The remainder of this paper is organized as follows. In Section 2, the Sloop architecture is described. The application of the framework to different species is presented in Section 3, followed by a summary and conclusion.

2 The Sloop System Architecture and Process

Sloop is designed to bring three stakeholders together: biologists who provide data and use results for conservation efforts, the computer vision community that provides algorithms, and the “crowd” that provides data and relevance feedback for remuneration, entertainment, or citizen science.

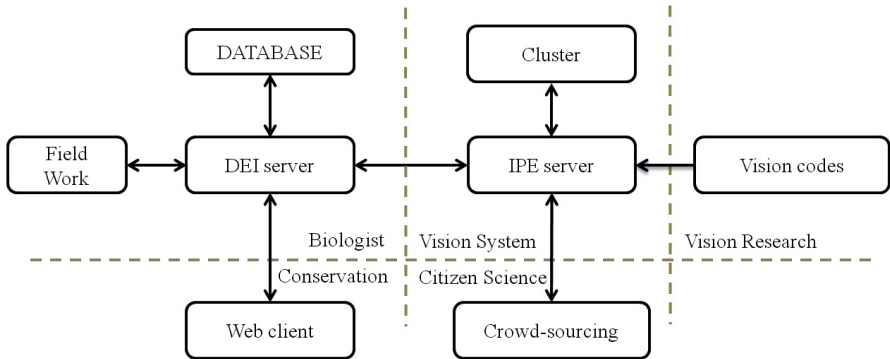


Fig. 1. The Sloop system is an interactive distributed image retrieval system with relevance feedback using crowdsourcing. It comprises of a Data Exchange and Interaction (DEI) server and an Image Processing Engine (IPE) that mediate the interaction between users, crowds, vision algorithms and computation.

Sloop comprises of a Data Exchange and Interaction (DEI) server that implements the user interface/database, and an Image Processing Engine (IPE) that provides tools for components of a work flow (see Figure 1). Typically, the inputs to Sloop include images and metadata, and the output is a table of identities associated with each image. A typical workflow includes interaction between the user and the system through stages of preprocessing, feature extraction, matching and relevance feedback. The stages of a workflow operate asynchronously and Sloop synchronizes the system (for example, closures for identities) during a “Nightly Build” process. The elements of a Sloop system’s toolkit (version 3.0) include:

- Segmentation using mean-shift, SVM, and graph-cut methods.
- Illumination correction using global contrast correction techniques, and an interactive specular removal algorithm [18] guided by the user (Figure 3).
- Rectification, including interactive spline methods (see Figure 2).
- Feature extraction in regions of interest (ROIs), including randomized representation of ROIs that perturb image position and scale to account for uncertainty of features and fiducials.
- Matching, including patch-based (Multi-scale PCA [1], Scale-Cascaded Alignment [9]), local feature-based (Histogram [11], SIFT [19] and affine invariant variations [8]), and hybrid shape context-based matching for point features. Additional features including HOG, SURF, and LBP and variants are also scheduled to be incorporated.
- Hierarchical matching, including coarse-scale matching with invariant features followed by deformable matching with Scale Cascaded Alignment (SCA).
- Aggregation, including rank and score aggregation.
- Relevance feedback, including metric modification and aggregation.
- Crowdsourcing, including Facebook and Mechanical Turk.

For any new species, the toolbox is typically used to develop workflows in a Sloop sandbox. A production Sloop is deployed after successful tests on the sandbox (see `sloop.mit.edu` for examples). For space limitations, and because several techniques are already published, here we will focus on two aspects. The first is the role of hybrid contexts in aggregated matching and the second is the role of relevance feedback in matching including hierarchical and aggregated matching.

2.1 Hybrid Context

The hybrid context algorithm uses Eulerian and Lagrangian shape contexts to match point sets. The method uses two iterations of correspondence and alignment. In the first iteration, one of two shape context approaches—Eulerian [20] or Lagrangian—is used to characterize and correspond the points, and in the end, the two approaches are aggregated using the most-successful matching. After one of the shape context methods is used to find a correspondence, the median distance is used to form an initial alignment. In the second iteration, points are matched

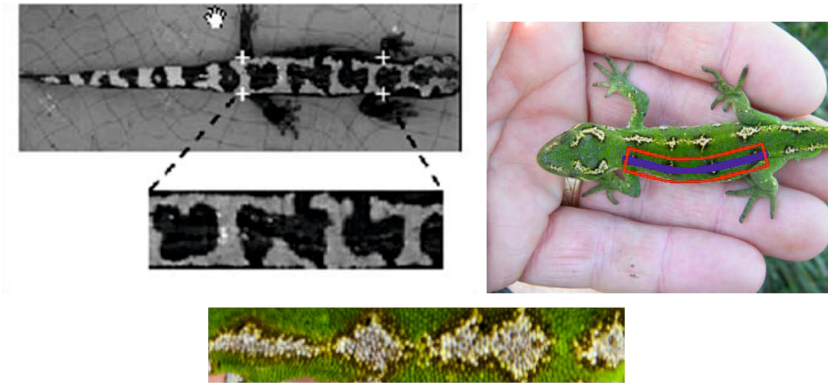


Fig. 2. Either segmented images or user marked key points are used to nonlinearly rectify images for feature extraction. Shown here are the rectified image of a marbled salamander (note distorted grid), a jewelled gecko with a spline for rectification, and the corresponding patches for both.

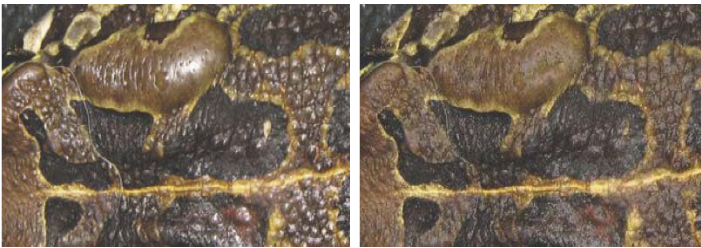


Fig. 3. Interactive example-based specular removal

using a doubly-stochastic formulation [21] and realigned with an affine transform calculated through the RANSAC algorithm [22]. Aligned point sets are finally scored by considering the distribution of Euclidean distances between corresponding points. We note that neither context method, Eulerian or Lagrangian individually perform especially well, but when aggregated, the performance substantially improves, with even more gains achieved through relevance feedback. The results are presented in the next section.

2.2 Relevance Feedback

As the images are ranked, the user verifies by browsing the top few ranks (typically 20) and this information is used as feedback for subsequent rounds of matching to improve similarity matching, in two ways: a) the best score of the cohort group is used to re-rank the images and b) the population-based prior is replaced by a cohort-based posterior [11] estimate in a Bayesian sense. In the

first approach, we estimate the probability $P(C = c_i|Q = q)$ to find the class c_i that matches the query q by maximum likelihood. Thus, as in [11, 23], for an ensemble of images $I_{1,c_i}, I_2, \dots, I_{e_i,c_i}$ belonging to class c_i and obtained through relevance feedback, the likelihood $P(Q = q|C = c_i) = \max_e P(Q = q|I = I_{e,c_i})$. In the second approach [11], feature vectors of the cohort are used to calculate new individual centered representations. The new bases are used for comparing a query to a cohort group, whilst returning the best match over the feature ensemble. In a Bayesian setting, $P(C = c_i|Q = q) \propto P(Q = q|C = c_i)P(C = c_i)$. This expression is evaluated comparing the query to each member in the relevance ensemble (likelihood) and by evaluating the probability of the ensemble member using the population statistics (prior).

Relevance feedback improves identification efficiency by incorporating information from all images that have been matched to a query image (see Figure 4). In particular, many image sets include left and right images that cannot be directly compared. Relevance feedback allows these images to be linked via intermediary left/right pairs that are matched to both lone images.

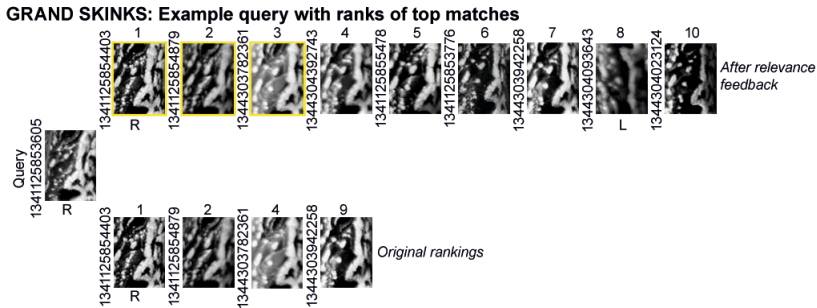


Fig. 4. Images matching a query with the top 20 ranked matches, before and after relevance feedback. Relevance feedback is performed using matching images within the initial top 5 ranked matches. Numbers indicate the rankings, and L/R indicate a session where only one side of the animal was photographed.

Crowd sourcing can accelerate relevance feedback. Akin to ReCAPTCHA, in our approach, feedback is accepted when performance on control images (for which the results are known) is perfect. This self-normalization results in excellent candidate selection, a key issue if crowds are to be useful. On average, after a person performs well on 40 known-pairs (twenty tests) their recall is about 95%. The number of people passing this barrier is about a third of the total population (see Figure 5). While it is difficult to measure the accuracy of people who do not complete many tasks, the people who persist are overwhelmingly also very good at identification.

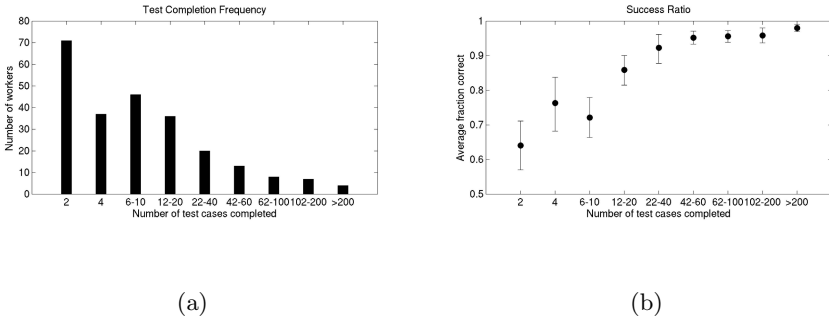


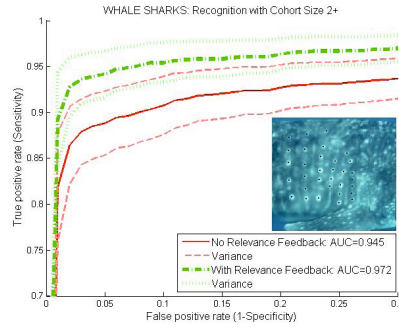
Fig. 5. Persistent test takers are somewhat infrequent (left). They are also the most skillful (right). The success rate is strongly correlated to the number of tests that a candidate undertakes, suggesting that persistent candidates are also the most skillful at identification.

3 Individual Identification Examples: Sloop Systems

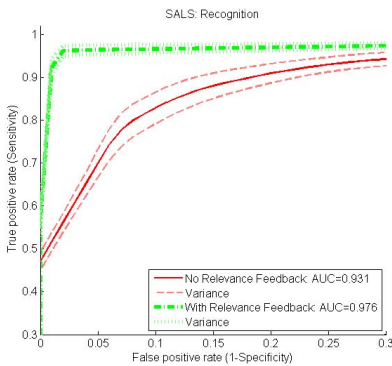
Sloop algorithms are demonstrated on *Rhincodon typus* (whale shark, 35,000 photos/3,000 individuals), *Ambystoma opacum* (marbled salamander, 10,000 photos/2,000), *Naultinus gemmeus* (jewelled gecko, 10,000 photos/1,600 individuals), *Oligosoma otagense* (Otago skink, 8,900 photos/900 individuals), and *Oligosoma grande* (grand skink, 21,700 photos/2,500 individuals) (see Figure 6).

For whale sharks we use the spot patterning behind the gills of the whale shark on one or both sides [6]. Coordinates of the spots have been specified by users or extracted, and the identification algorithm matches pairs of coordinate sets rather than the images themselves. The scale-normalized spot distributions are compared using hybrid contexts. The Eulerian context is based on 16 overlapping bins at two ranges of distances. In the Lagrangian context, a spot's context consists of its five nearest neighboring point coordinates; subsequently, the top three highest matching contexts are used for alignment. After a final round of RANSAC-based affine alignment, the score of the aligned point sets is determined by the area between two CDF curves: the cumulative distribution of Euclidean distances between corresponding points and that of an ideal perfect match. Thus, lower scores indicate closer matches. The results of the two algorithms are aggregated by taking the minimum score. Relevance feedback is also applied to this method, with improved performance. To test our algorithm, we queried 147 images of 18 total individual whale sharks for which the cohorts were known against 5,200 or 9,300 photos from the database of left or right side photos respectively. The results of the individual, weak matching algorithms, compared to the aggregated version are shown in ROC curves (see Figure 7).

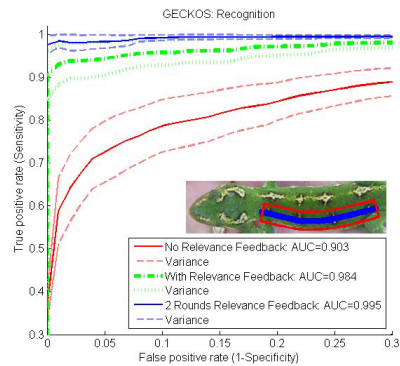
For marbled salamanders a hierarchical approach is used in which SIFT/MS-PCA is followed by SCA refinement. The performance, measured as the area



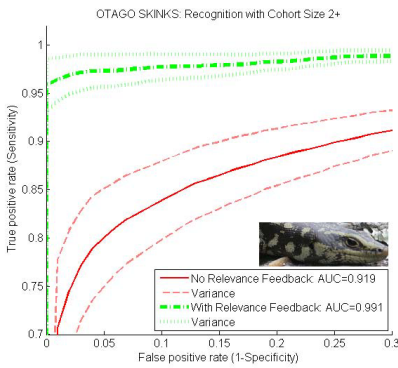
(a) whale shark



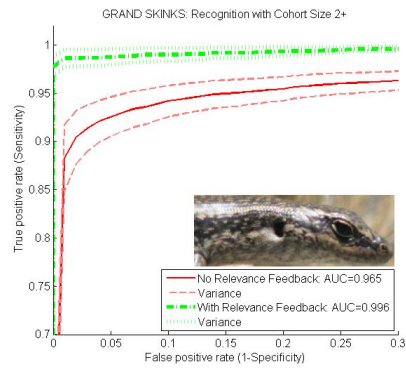
(b) marbled salamander



(c) gecko



(d) Otago skink



(e) grand skink

Fig. 6. These graphs indicate that the baseline performance is quite good and improves with coupled human-machine relevance feedback cycles. After relevance feedback was applied the ROC improvement is statistically significant and nonlinear.

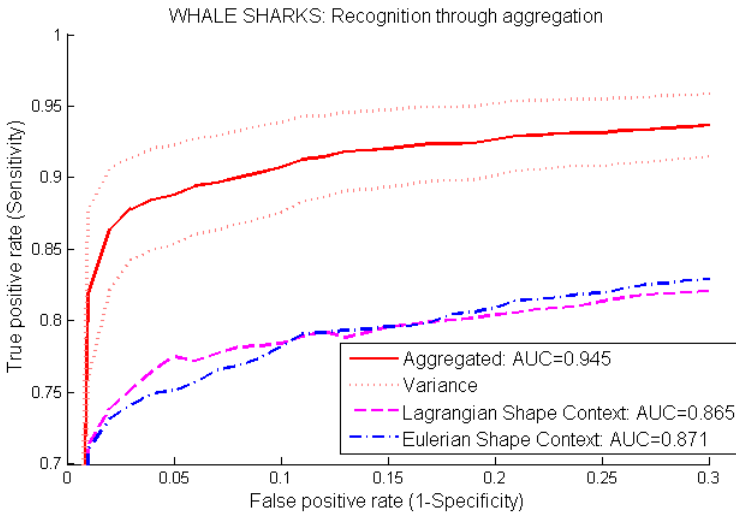


Fig. 7. ROC curves for whale shark of two weak matching algorithms, combined through aggregation to achieve much higher performance

under an ROC curve (AUC), improves from 90% with the use of SIFT/MS-PCA to 93.1%. The addition of relevance feedback on the top 5 ranked retrievals results in an improvement to 97.8%. For jewelled geckos, a randomized SIFT with bidirectional matching is used. With relevance feedback the AUC is 99.5%.

For Otago and grand skinks a worker defines three patches on each image, between the nostril and eye, between the eye and ear, and between the ear and the shoulder. A capture may contain images of the left or right side of the animal or both, for up to six patches. The patches are normalized in orientation and scale. SIFT features [19] are then extracted from the patches. Next, corresponding patches on the same side of the animal are matched in the non-parametric Bayesian approach. The maximum score over the patch ensemble is used for ranking, removing the effects of low-scoring patches and features. Relevance feedback is done with the top ten results, producing excellent performance. On test sets of 1002 Otago and 1008 grand skinks, AUC values are above 90% with SIFT alone and above 99% after relevance feedback on the top 10 ranked retrievals.

The Otago and grand skink Sloop systems are in operational use at the Grand and Otago Skink Recovery Programme at the New Zealand Department of Conservation. Since a major protection operation has reversed the decline of the skink species, large datasets of skink photographs are produced yearly. These images need to be matched to identify individuals and to track population dynamics. This was formerly a labor intensive task, with up to half a million manual image comparisons taking place in 2011. In comparison, since Sloop presents ranked matches, matches are found in the top 100 comparisons (and most of these are in the top ten) 96% of the time for Otago skinks and 99% of the time for grand

skins. This reduces manual matching to 4% and 1%, respectively, of the labor that would otherwise be required.

Conclusion

The MIT Sloop system is an individual animal identification framework where hybrid systems utilize computer resources and human skills resulting in high performance recognition systems for large-scale conservation efforts. In this document, we described how this framework is applied and gives state-of-the-art results in five species.

Acknowledgment. The Sloop system and algorithms used for experiments here are developed by Ravela (version 2.0), revised by Yang and Ravela (version 2.5) and then by Duyck and Ravela (2.7-3.0) at the Earth Signals and Systems Group, MIT. This work is supported in part by AFOSR(FA9550-12-1-0313), NSF DBI-1146747, IPN-SIP 20140325, and MIT-Mexico Seed Fund (MISTI). Any opinion, findings, and conclusions in this material are those of the authors. We thank Jason Holmberg for whale shark data.

References

1. Gamble, L., Ravela, S., McGarigal, K.: Multi-scale Features for Identifying Individuals in Large Biological Databases: An Application of Pattern Recognition Technology to the Marbled Salamander *Ambystoma opacum*. *Journal of Applied Ecology* 45(1), 170–180 (2008)
2. Kelly, M.: Computer-Aided Photograph Matching in Studies using Individual Identification: An Example from Serengeti Cheetahs. *Journal of Mammalogy* 82, 440–449 (2001)
3. Araabi, B., Kehtarnavaz, N., McKinney, T., Hillman, G., Wursig, B.: A String Matching Computer-Assisted System for Dolphin Photoidentification. *Annals of Biomedical Engineering* 28, 1269–1279 (2000)
4. Hiby, L., Lovell, P.: A Note on an Automated System for Matching the Callosity Patterns on Aerial Photographs of Southern Right Whales. *Journal of Cetacean Research and Management* 2, 291–295 (2001)
5. Mizroch, S., Beard, J., Lynde, M.: Individual Recognition of Cetaceans: Use of Photo-Identification and Other Techniques to Estimate Population Parameters. In: Hammond, P., Mizroch, S., Donovan, G. (eds.) *Computer Assisted Photo-Identification of Humpback Whales*, Cambridge, UK. International Whaling Commission, vol. 12, pp. 63–70 (1990)
6. Arzoumanian, Z., Holmberg, J., Norman, B.: An Astronomical Pattern-matching Algorithm for Computer-Aided Identification of Whale Sharks *Rhincodon typus*. *Journal of Applied Ecology* 42(999-1011) (2005)
7. Kumar, N., Belhumeur, P.N., Biswas, A., Jacobs, D.W., Kress, W.J., Lopez, I.C., Soares, J.V.B.: Leafsnap: A Computer Vision System for Automatic Plant Species Identification. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) *ECCV 2012, Part II. LNCS*, vol. 7573, pp. 502–516. Springer, Heidelberg (2012)

8. Ravela, S., Gamble, L.: On Recognizing Individual Salamanders. In: Asian Conference on Computer Vision, vol. 2, pp. 741–747 (2004)
9. Yang, C., Ravela, S.: Spectral Control of Viscous Alignment for Deformation Invariant Image Matching. In: International Conference on Computer Vision, vol. 1, pp. 1303–1310 (2009)
10. Ravela, S., Duyck, J., Finn, C.: Vision-Based Biometrics for Conservation. In: Carrasco-Ochoa, J.A., Martínez-Trinidad, J.F., Rodríguez, J.S., di Baja, G.S. (eds.) MCPR 2012. LNCS, vol. 7914, pp. 10–19. Springer, Heidelberg (2013)
11. Ravela, S.: On Multi-scale Differential Features and their Representations for Recognition and Retrieval. PhD thesis, University of Massachusetts at Amherst (2002)
12. Hiby, L., Lovell, P., Patil, N., Kumar, S., Gopalaswamy, A., Karanth, U.: A Tiger cannot Change its Stripes: Using a Three-dimensional Model to Match Images of Living Tigers and Tiger Skins. *Biology Letters* 5(3), 383–386 (2009)
13. Town, C., Marshall, A., Sethasathien, N.: Manta Matcher: Automated Photographic Identification of Manta Rays using Keypoint Features. *Ecology and Evolution* 3(7), 1902–1914 (2013)
14. Sherley, R., Burghardt, T., Barham, P., Campbell, N., Cuthill, I.: Spotting the Difference: Towards Fully-automated Population Monitoring of African Penguins *Spheniscus demersus*. *Endangered Species Research* 11(2), 101–111 (2010)
15. Holmberg, J., Norman, B., Arzoumanian, Z.: Estimating Population Size, Structure, and Residency Time for Whale Sharks *Rhincodon typus* through Collaborative Photo-Identification. *Endangered Species Research* 7, 39–53 (2009)
16. Loos, A., Ernst, A.: An Automated Chimpanzee Identification System using Face Detection and Recognition. *EURASIP Journal on Image and Video Processing* 12(1), 1–17 (2013)
17. Duyck, J., Finn, C., Hutcheon, A., Vera, P., Salas, J., Ravela, S.: Sloop: A Pattern Retrieval Engine for Individual Animal Identification. *Pattern Recognition* (2014)
18. Runge, J.: Reducing Spectral Reflections through Image Inpainting. Master's thesis, Massachusetts Institute of Technology (2009)
19. Lowe, D.: Distinctive Image Features from Scale-invariant Keypoints. *International Journal of Computer Vision* 60(2), 91–110 (2004)
20. Belongie, S., Malik, J., Puzicha, J.: Shape Matching and Object Recognition using Shape Contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24(4), 509–522 (2002)
21. Rangarajan, A., Chui, H., Mjolsness, E., Pappu, S., Davachi, L., Goldman-Rakic, P., Duncan, J.: A Robust Point Matching Algorithm for Autoradiograph Alignment. *Medical Image Analysis* 1(4), 379–398 (1997)
22. Hartley, R., Zisserman, A.: *Multiple View Geometry in Computer Vision*, 2nd edn. Cambridge University Press (2004)
23. Ravela, S., Yang, C., Runge, J., Gamble, L., McGarigal, K., Chesser, M.: Visual Recapture for Movement Ecology at Interannual Timescales. In: Workshop on Visual Observation and Analysis of Animal and Insect Behavior (2008)