Pattern Recognition 48 (2015) 1059–1073

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Sloop: A pattern retrieval engine for individual animal identification

James Duyck [a], Chelsea Finn [a], Andy Hutcheon [b], Pablo Vera [c], Joaquin Salas [c,*], Sai Ravela [a]

[a] Earth Signals and Systems Group, Earth Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA, USA
[b] Department of Conservation, Dunedin, New Zealand
[c] CICATA-IPN, Cerro-Blanco 141, Colinas del Cimatario, Queretaro 76090, Mexico

ARTICLE INFO

Article history:
Received 2 January 2014
Received in revised form
7 June 2014
Accepted 19 July 2014
Available online 1 August 2014

Keywords:
Photo-identification
Animal biometrics
Individual identification
Relevance feedback
Crowdsourcing
Conservation
Scale-cascaded alignment
Local features
Hybrid shape contexts
Gecko
Skink
Whale shark
Salamander

ABSTRACT

Identifying individuals in photographs of animals collected over time is a non-invasive approach for ecological monitoring and conservation. This paper describes the design and use of Sloop, the first image retrieval system for individual animal identification incorporating crowd-sourced relevance feedback. Sloop's iterative retrieval strategy using hierarchical and aggregated matching and relevance feedback consistently improves deformation and correspondence-based approaches for individual identification across several species. Its crowdsourcing strategy is successful in utilizing relevance feedback on a large scale. Sloop is in operational use. The user experience and results are presented here to facilitate the creation of a community-based individual identification system for conservation planning.

1. Introduction

The development of effective conservation strategies for rare and endangered species requires unbiased and precise information on their life history and population ecology [1]. Capture-Mark-Recapture (CMR) studies enable researchers to track individual animals over time to answer questions related to individual growth, survival, dispersal and reproductive strategies. CMR studies typically use techniques in which animals are physically marked or tagged. These methods are intrusive to varying degrees and inefficient to implement in large numbers. Alternative identification techniques that overcome these limitations are needed.

Numerous efforts exist to identify individual animals using photographs. The simplest one to conceive is manual identification, but its high recall comes at exceptional cost. Manual searches are only feasible for small collections; at 10 s per comparison a 10,000-sized catalog will take approximately 15 person-years to analyze. Manual searches typically also employ *ad hoc* strategies using individual markings. This process is difficult to automate and extend and may lead to imprecise quantitative analysis.

Computer-based pattern recognition approaches [1–10] decidedly benefit the identification problem. They scale to large collections, extend to multiple species and are convenient to use. However, despite the advantages of computational speed and advances in automatic pattern recognition, it is not feasible to automatically deliver the high recall needed for correctly tagging many species. Some degree of human involvement arguably benefits the identification process. A problem that emerges is how to determine the optimal level of human involvement. The earliest approach used deformable template matching [2,3] that could be used for multiple species. However, it demands user inputs for every compared pair of photographs; clearly too much.

Motivated by an image retrieval approach to face recognition [11], Ravela and Gamble [8] advanced the interactive retrieval paradigm for animal biometrics. Their retrieval approach characteristically begins with an unlabeled collection of photographs that are segmented, rectified and illumination corrected. Initially automated, these steps were later performed in a semi-automated

* Corresponding author. Tel.: +52 55 5729 6000x81015.
E-mail addresses: jsalasr@ipn.mx, salas@ieee.org (J. Salas).

http://dx.doi.org/10.1016/j.patcog.2014.07.017
0031-3203/

fashion to limit errors from propagating to subsequent stages. Typically, generic appearance or geometry features are extracted from regions of interest. The spatial representations of features including vectors, histograms or graphs are constructed and compared, usually using Bayesian inference, to rank images.

The key step then follows. The user verifies a limited number of top ranked retrievals and thus makes the final decision for labeling cohorts[1] with shared identities [8,1,9,10]. Once an initial collection is indexed, over time, this linear approach sequentially assimilates new images by comparing them with existing cohorts or singletons to maintain updated capture histories. The system becomes an interactive search engine for identities because the features are generic, appropriate representations of images are precomputed, and comparisons are fast. By ranking images and presenting top matches, the retrieval approach only requires the user to view pairs that are most likely to be matches. This precision enables human recall that is much more efficient than manual matching. With this approach, a 10,000-size collection may typically require about three person-months to analyze; this is the state of the art.

A retrieval approach contrasts with an automated recognition approach for identifying an animal photograph. The latter is preferred when there are fixed identities with labeled exemplars and, most importantly, when indexing is not the objective. This is not typically true for small animal species where new individuals frequently enter the database and demand persistent indexing and high recall. In these cases, full automation is not yet feasible and will decidedly benefit from a retrieval approach.

Sloop, a distributed interactive system for individual animal identification, emerged through several iterations of the image retrieval methodology [8,1,9,10]. Sloop is neither a complete CMR system nor does it provide statistical analysis, but it is a tool with which users can reconstruct capture or encounter histories. It predominantly uses visual features but ongoing work includes additional textual metadata. Sloop utilizes multiple techniques plus human input, and combines them to deliver high performance. It is deployed operationally and finds mention in popular press [12].

In the community, there are now several retrieval-based systems that, as with Sloop [8,1,9], propose using generic visual (largely local) features with attendant claims of extensibility. Clearly, the freedom that generic approaches provide from *ad hoc* individual markings with linear indexing cost is beneficial and delivers reasonably good performance. However, there are at least two limitations.

First, just as real-world face recognition technologies are highly specialized to satisfy recall demands, high-recall animal biometrics also demands trading invariance of the generic algorithms for selectivity by application. As this paper shows, while generic one-method-for-all approaches are abundant and perform comparably, rarely does any deliver intrinsically high recall at least in the several species considered here. As a consequence, not only are refinements towards higher recall of interest, reusable techniques for tuning multiple arbitrary generic methods, including combining (aggregating) them to deliver a performance better than any one source alone, are also relevant.

Second, as collections grow, even linear human effort is substantial for small teams to undertake. Reductions in mouse-click counts and keystrokes are important but methodology to recycle already-performed work to reduce subsequent human effort is also of great interest. For example, as matches start to form, it should be possible to feedback this relevance information to improve subsequent matching. Neither of these issues receives much attention in the Animal Biometrics community which they must as collections transition to Big Data regimes with richer images, feature varieties and collection sizes.

In this paper, the Sloop system architecture, its workflows, and algorithms are described with application to several species including the jewelled gecko (*Naultinus gemmeus*), grand and Otago skink (*Oligosoma grande* and *Oligosoma otagense*), whale shark (*Rhincodon typus*) and marbled salamander (*Ambystoma opacum*). In each of these cases, *baseline* algorithms using deformable models, patch-based appearance models, local features, and/or shape context models are shown to deliver reasonable performance. Then, new algorithms that uniformly outperform baseline algorithms are presented. This includes hierarchical ranking and retrieval (ordered combination of algorithm results), rank and score aggregation (parallel combination of algorithm results), and relevance feedback including the use of crowdsourcing for scalability. Our experiments suggest that the advantages of these steps compound. For example, indexing a 10,000-size collection can, in principle, be accelerated and completed in a few person-days.

Thus, the central advance of this paper is that whilst fully automated recognition systems are not yet within reach for Animal Biometrics, coupled human-machine systems that deliver high performance are achievable. In these approaches, the algorithms can reduce human effort and the human feedback can improve system performance. Together, they can produce extensible, scalable, and effective large-scale deployments. Within the realm of Animal Biometrics, Sloop appears to be the first such operational system. The result of deploying Sloop on the Grand and Otago Skink Recovery Programme, in Dunedin New Zealand, and the first full year of its use is described. The results are extremely encouraging and may serve as a useful model to integrate biologists, computational vision researchers and citizen scientists in a unified framework.

The remainder of this document is organized as follows. In Section 2, closely related systems are reviewed and compared to Sloop. In Section 3, the Sloop system architecture, workflow and methods are discussed. The application of the methods to individual species is presented in Section 4, operational experience in the Grand and Otago Skink Recovery Programme is described in Section 5, and the paper concludes with a discussion in Section 6.

2. Related work

Identifying individuals among a species population is of increasing interest. Some of the earliest approaches [2,3,13] use 3D deformable matching, which is extensible, but more frequently used are new techniques driven by the need for rapid large-scale matching. Sloop contains a 2D deformation invariant matching algorithm [9], which is only used to improve existing rankings. It is relatively fast and demonstrated to be highly effective on marbled salamanders. Here we show additional improvements by relevance feedback which also facilitates comparisons where one or both sides may be imaged [14].

Within the realm of feature-based methods, there are two main classes of recognition methods; specialized methods [6,15,16] for individual species, and generic methods [8,1,9,17–19,10]. The specialized techniques are of limited interest. For example, a generic correspondence-based approach in Sloop [10] can be adapted for whale sharks with markedly improved performance over an earlier specialized approach [6].

Within the realm of generic methods, Ravela and Gamble [8], motivated by face recognition [11], proposed using multi-scale differential feature histograms, and later using randomized multiscale-PCA [1] on marbled salamanders. SIFT features [20] are popular generic features. For example, Yang and Ravela [9]

[1] A cohort is a set of images with the same identity; a singleton is a cohort of size one.

apply it to marbled salamanders showing performance similar to multiscale-PCA, and Town et al. [17] use SIFT features to characterize the ventral surface area of manta rays. They compare different features, including SURF [21] and ORB [22], arguing that SIFT works well on small collections. SIFT features are poor at capturing gist and global summaries, or for highly repetitive textures, or materials such as cheese.[2] Distributions of local features can be used for these cases [11,8], which are available in Sloop. SIFT features can also be improved in a number of ways to deal with nonlinear deformations, for example in Sloop by using them as a precursor to scale-cascaded alignment [9] or by randomized perturbations of the image geometry [1,10]. Here, the role of relevance feedback is explored with other established techniques SIFT, SURF and ORB, extending recent work [10,23]. This paper is the first detailed presentation including a streamlined architecture and comparative and exhaustive experiments on all species data. The improvements are substantial and thus of interest.

Shape contexts [24] also play a role, for example, in African penguins (*Spheniscus demersus*) [25] using detected minutiae [18] whose positions are randomized for robustness. Randomized representations are a standard feature of Sloop [1] but, more importantly, hybrid contexts aggregating Lagrangian (feature positions) and Eulerian (feature densities) information are used in an iterative RANSAC-based correspondence and alignment scheme. This is further improved through relevance feedback [11,10], but both aspects are novel.

There are several notable systems including ECOCEAN [26], face-recognition-based systems for chimpanzees and gorillas [27] and the African Penguin Recognition System (APRS) [18]. The latter two can be applied to different species, particularly as APRS is further motivated by biometrics for Turing patterns [28]. Sloop is also an operational system deployed to multiple species. Its ever expanding toolbox allows for myriad workflows. It places particular emphasis on the interaction between human and system, leveraging relevance feedback on a large scale.

A characteristic of extensible systems is the variety of vision tools available and workflows that can be implemented. Loos and Ernst [27] use AdaBoost [29] detection, alignment via Procrustes analysis, and light-normalization via histogram equalization, before proceeding to face identification using Gabor filters [30] for the global features and SURF [21] descriptors for the local ones. They use a Sparse Representation classifier [31] on the global features, and a Support Vector Machine [32] on the local ones. Sloop contains mean-shift, SVM and graph-cut segmentation using color-texture features [10], several illumination correction techniques, most notably a novel exemplar-based method for specularity removal [33], and it contains spline-based methods for rectification [8]. Sloop includes multiscale differential feature histograms [8], SIFT [20], affine invariant features [34], hybrid shape contexts, multiscale patches and derivatives [1]. Multi-method aggregation, randomized representations, iterated correspondence and RANSAC alignment, and scale-cascaded alignment are all available to implement a variety of workflows.

A notable element of our approach is the use of crowdsourcing. This is suggested by several authors [19,26,17] as a way to gather images. The conceptualization here is different. We propose Sloop as an interactive human–machine system where relevance feedback improves algorithm performance [11,35,10] and improved algorithms reduce the scientist's subsequent work. In this framework, crowdsourcing is used to garner large-scale relevance feedback, which is novel for animal biometrics, and includes engagement through citizen science. The work described here distinguishes itself from the current state of the art, and include some of the best results for the species described. The framework provides a path forward for scalability by using relevance feedback in a crowdsourcing context. Sloop is operational and its methods can be extended to many species.

[2] Ted Adelson, personal communication

3. The sloop system architecture and indexing process

Sloop's overarching design criteria give the biologist primary control over the data and results, give the vision community an easier integration path into the system and automatically leverage crowds for citizen science and relevance feedback. It does this using a system architecture that maintains a separation of IT and Vision components, allowing individual advances to be easily and independently absorbed.

3.1. System architecture

The structure of the Sloop retrieval system (v2.7) is shown in Fig. 1. Sloop is composed of a Data Exchange & Interaction Server (DEI) and an Image Processing Engine (IPE). The DEI implements the user interface/database (see Fig. 2) as a web application running on GlassFish with Postgres binding [36]. It typically is installed and run in the client space but we house several web-based systems for prototyping. IPE contains preprocessing tools including segmentation, illumination correction, rectification, matching, relevance feedback and crowdsourcing. It is run as a native MATLAB/Octave server, testing research codes with relative ease. IPE and DEI interact through a database and, together, implement a workflow that each image undergoes on its journey from being a photograph to becoming an identity.

A workflow (see Fig. 3) typically involves uploading images and metadata, preprocessing images to correct for illumination and geometry, extracting features, producing ranked retrievals, incorporating user judgments and iterating using relevance feedback. The output of the system is a table with identities associated with each image.

IPE and DEI typically operate asynchronously and can operate in batch or interactive modes. Sloop enables multiple users to work on different aspects of the workflow in parallel. Some users may be aiding preprocessing and others verification of matches while the IPE may be rectifying and matching jobs on the cloud.

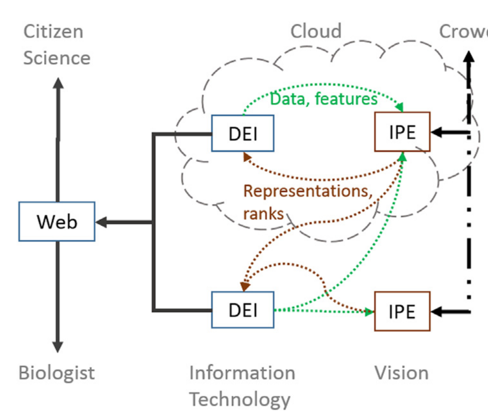

Fig. 1. The Sloop system is an interactive distributed image retrieval system with relevance feedback using crowdsourcing. It comprises of Data Exchange and Interaction (DEI) server and an Image Processing Engine (IPE) that mediate the interaction between users, crowds, vision algorithms and computation. Sloop can be used in standalone mode, the predominant form, or in distributed mode where IPE and DEI are separated to leverage computing resources including cloud computing. Vision codes, including in Matlab, are directly ingested within the IPE.

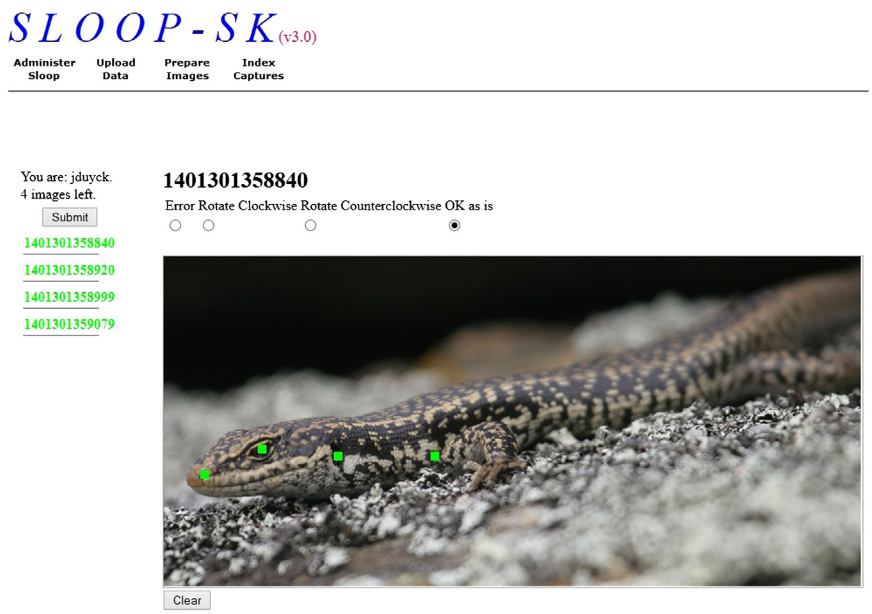

Fig. 2. The user interface for the Otago and grand skink Sloop systems. This figure shows the preprocessing stage of marking four key points which are used to obtain normalized patches for comparisons.

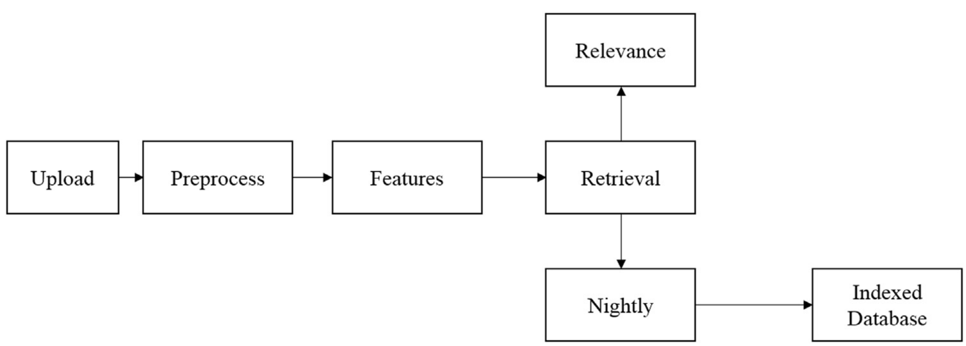

Fig. 3. A Sloop workflow is an interaction between the users and the system and includes preprocessing, feature extraction, retrieval and relevance feedback. "Nighty" processes synchronize databases across distributed use.

This greatly facilitates operational use, in contrast to linear methods.

Naturally, in such a system, synchronization of the system state across multiple steps becomes necessary. For example, as multiple users verify images, the closure of all matching pairs into a clique resolving multiple, possibly contradictory, relevance judgments is necessary. Sloop synchronizes the collection's state through a *Nightly* build process that locks users out before performing closure operations (among others), thus preventing deadlocks or conflicts. At the end of each Nightly build the Sloop system is consistent; all photographs that share the same identity become cohorts with identities being merged or initiated as needed. At any time after the latest Nightly build the user can lock Sloop to control the quality of the identity tables and unlock it with updated information. In the remainder of this section, Sloop's methods along a typical workflow are described.

3.2. Preprocessing methods

Images are typically preprocessed for feature extraction and matching. Although the steps vary between species, segmentation, illumination correction, image rectification and feature extraction are useful tasks. Fig. 4 shows examples of preprocessing for several species. The methods chosen for a particular dataset depend on the characteristics of that dataset and the human resources available. As noted, Sloop contains mean-shift based, SVM and graph-cut segmentation methods on color-texture features, shown in Fig. 4(B). These were used in early work with rectification but the demands of high throughput and accuracy have led to semi-automated methods, discussed next. For these reasons, and because these are standard algorithms, the implementations are not discussed further.

For illumination correction, Sloop includes global contrast correction and our exemplar-based specularity removal method [33]. This technique emphasizes Sloop's interactive approach in which users sparingly seed first guesses or correct intermediate results of the algorithms for high performance. In Fig. 4 (A), the specularity removal algorithm within Sloop [33] is shown for Fowler's toad. In this approach, the user marks a few specular and normal regions. Based on this input, Sloop replaces detected specular spots with information from normal regions to seamlessly in-fill. In contrast to pde-based in-filling, Sloop allows for a realistic and detailed removal of specularities [33].

A common mode of rectification for deformable species with strong bilateral symmetry such as marbled salamanders, skinks and geckos is to mark a preferred axis of symmetry (the *medial axis*) as a spline using a few key points, and then to rectify a bounded region around the spline. An example of this process is shown in Fig. 4 (C,D) for a marbled salamander and gecko. In current use, typically three to four points across the length of the

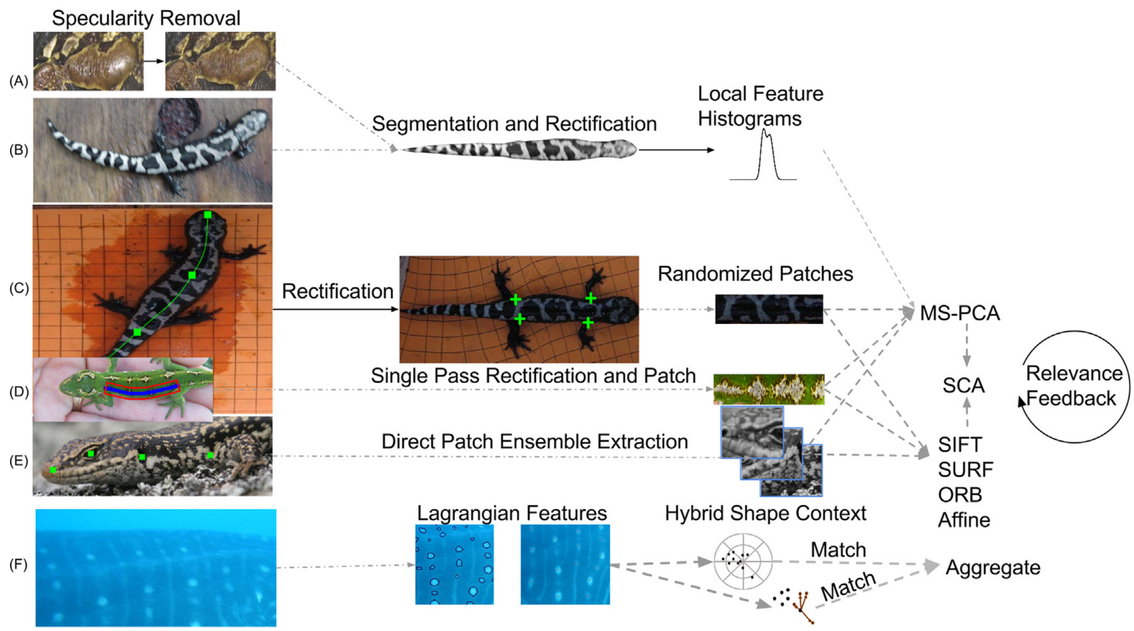

Fig. 4. Images are preprocessed within Sloop, often with some species specific variations. In a typical workflow, segmentation or user marked key points are used to extract patch, histogram or point features at multiple scales, sometimes preceded by specularity removal.

animal are sufficient to produce excellent spline candidates and rectifications. For other species, ROI patch selection around *fiducial* points with rotation, affine or spline based normalization is used. In the skinks, for example, a few points mark the patches and their orientations (see Fig. 4 (E)) which are rectified and used for matching. Finally, Lagrangian features are also used. Spots, both manually marked and automatically detected (in addition to invariant feature detectors for matching), are available and used in whale sharks (see Fig. 4 (F)). The progression of preprocessing approaches (A–F) points to our early rejection of fully automated preprocessing in favor of manual intervention. A steady and careful reintroduction of automation to reduce task-dependent human effort followed so that very few inputs are currently needed per image for delivering high performance. Features are extracted and matched after preprocessing, as discussed in the next section.

3.3. Matching techniques

Multiple matching techniques are embodied within Sloop. They include patch-based (Multi-scale PCA [1], Scale-Cascaded Alignment [9]), local feature-based (Histogram [11], SIFT [20] and affine invariant variations [8]), and context-based matching (hybrid shape-contexts with cascaded correspondence, a novel method). Here we describe the methods that will be used in the subsequent applications to geckos, skinks, whale sharks and salamanders. They are presented from a methodological perspective with parameterization and application discussed subsequently in Section 4.

3.3.1. Patch-based methods: MS-PCA

The first technique contained within Sloop is an MS-PCA (and Fisher's Discriminant variant) technique. Patches are selected from ROIs bounded by fiducials on the rectified image that is filtered at multiple scales, using Gaussians and their derivatives. All the responses are assembled into a large vector and inference in this vector space is used to determine similarity.

A major source of uncertainty in marking the fiducial points is human error. A key element of the MS-PCA approach is that patches are obtained by perturbing the image in position and scale so as to account for this uncertainty. The perturbed-patch vectors form a randomized sample ensemble for each photograph.

When matching principal components, typically the top 100–150, the best match overall of the ensemble vectors associated with an image is reported as the score. Thus, a key aspect is the use of “max (min) measure” as the aggregate measure on the ensemble, in contrast to the mean or some other moment-based statistic. The hypothesis is that the probability of a high score being accidentally generated by a mismatch is small and gets smaller as the ensemble grows. Indeed, it has been observed that increased perturbations increase the probability of finding a very good match. A large-scale implementation of the MS-PCA approach was used for the marbled salamander and is discussed in the next section.

3.3.2. Patch-based methods: SCA

A vast number of animals and plants deform in highly nonlinear ways. Even after rectification during preprocessing, there is typically strong residual deformation that is difficult to ignore during matching. A second technique within Sloop is a novel deformable matching algorithm [9], designed to handle such deformations. Scale-cascaded Alignment [9] is a fast 2D nonparametric deformation invariant matcher. It calculates a deformation vector-field $\mathbf{q}$ that nonlinearly maps a 2D template X into a target Y on a discretized domain Ω, defined as $X \circ \mathbf{q} \equiv X(\underline{p} - \mathbf{q}(\underline{p}))$, using position coordinate $\underline{p} = (x, y)$. This deformation field will typically be the solution of an optimization procedure of the form:

$$J(\mathbf{q}) := \tfrac{1}{2} \parallel Y - X \circ \mathbf{q} \parallel^2_{\underline{\underline{R}}} + L(\mathbf{q}), \tag{1}$$

where $\underline{\underline{R}}$ is a covariance defining the norm and $L(\mathbf{q})$ expresses parametric or non-parametric deformation constraints.

Parametric approaches face difficulties when encountering motion modes for which they are not parameterized. For example, a small degree polynomial may be a good model for smooth large-scale characteristic motions but a poor model for the higher order bends exhibited by another species. Non-parametric methods on the other hand, particularly diffeomorphic ones [9], can easily morph fields and capture a broad range of nonlinear deformations. They are natural candidates for deformation invariance but alas at the cost of selectivity so that with increasing invariance, nuisance deformations become inseparable from relevant ones.

The well-known dilemma, remove obviously is how to maintain control over a broad range of modes of deformation across object categories without losing selectivity within a category. SCA turns viscous alignment algorithm into a pattern recognition tool by finding the most *parsimonious* explanation of deformation [9]. SCA does this without explicit parameterization by representing the instantaneous deformation as a sum of Gabor bases weighted by a power law "turbulence" spectrum, and interpreted as the Green's function of $L(q)$. The deformation solution is obtained in a coarse-to-fine manner; the lowest wave number motion solution (translation) is found first, followed by higher (local) modes. Deformations in the low and very high wave numbers are those for which invariance is typically desired because they tend to correspond to large scale motions or noise. Selectivity is desired for deformations in middle wave numbers because they help distinguish the natural modes from abnormal ones.

As a novel, correspondence free approach to deformation invariant matching, SCA is also useful in other applications, such as weather prediction and storm tracking. The result of applying it to the marbled salamander [1,9] is shown in Fig. 15 and the application is discussed in Section 4.

3.3.3. Local features

Sloop contains several algorithms using local features. The first such algorithms are multiscale differential feature histogram-based methods, namely CO-1 and P.CO-1 variants [11]. These variants are useful for "global" matching and are successful in face recognition and trademark retrieval. Initial experiments with salamanders [8] and, more recently, on identifying species (not individuals) including *Anopheles gambiae*, *Anopheles stephensi*, and *Apis mellifera*[3], show success with these methods. There is also a second set of techniques based on affine invariant features [34] and their histograms.

The third local feature technique consists of "LIFE"-based methods including SIFT [20], SURF [21] and ORB [22]. The implementation used is the VLFeat toolkit [37] and mexopencv library for SURF and ORB. Sloop typically calculates features on patches and compares them using the maximal bidirectional score. Perturbed patches akin to MS-PCA are also used. Typically, when patch ensembles represent an image, either through perturbations or, for example, from different parts of the animal (see Section 4), the maximal score over the ensemble is used as the match score when comparing two images.

3.3.4. Cascaded correspondence and alignment by context

Sloop contains a novel algorithm to match feature coordinates. In the current implementation, coordinate sets are matched in an iterative correspondence and alignment approach. This technique is applied to whale sharks, discussed in Section 4. In the first iteration of alignment, two approaches are applied and aggregated by taking the most successful transformation of the two. In the first approach, an Eulerian shape context is used to characterize and correspond points, and the median translation over all correspondences is used to align the coordinate sets. In the second approach, several very strong correspondences of Lagrangian context seed the alignment. Much of the large scale correspondence and alignment error is thereby reduced. Subsequent iterations, therefore, find correspondences between the points by their position in a doubly stochastic formulation [40]. They realign the points using an affine transform calculated through the RANSAC algorithm [38]. Aligned point sets are finally scored by considering the distribution of Euclidean distances between corresponding points.

The Eulerian shape context [39] is calculated here with overlapping bins in polar coordinates. The Lagrangian context is based on the relative locations of the nearest points. Correspondence is determined using doubly stochastic normalization [40] on either the context match or the locations themselves in subsequent iterations. The Eulerian approach is robust to small amounts of noise because it discretizes the surrounding points into bins and uses all of the points when determining the alignment. However, it may fail from different image croppings, *i.e.*, when a row, column, or region of points is missing. The Lagrangian approach to alignment can accurately align the points when regions are missing by identifying few top-quality matches. However it is more sensitive to noise in the feature locations and is biased towards just a few points. The final algorithm requires only one of the two initial alignments to be successful because it considers the minimum scoring transformed points of the two outcomes, much like the ensemble aggregation techniques in other described methods.

3.4. Relevance feedback

Once matching is complete and ranked hypotheses are produced for photographs in a catalog, the user must verify whether the hypothesized matches are true. Although the number of images a user must review changes from species to species, typically ten to twenty images at a time are shown. The user selects the matching images (non-matching ones are marked implicitly) and this information is incorporated by Sloop.

In addition to logging the verification and naively presenting the next set of ranked retrievals, Sloop uses the cohorts identified by the user to iteratively improve similarity judgments in two ways: (a) the best score from the cohort group is used to re-rank images and, (b) the population-based prior is replaced by a cohort-based posterior [11] estimate in a Bayesian sense. The first method is used for all the applications described in the next section and both are used for the marbled salamander. During run-time, as multiple users are providing relevance feedback, all conflict-free aggregation is immediately made thereby propagating the benefits of verification immediately across users and photographs. During the nightly build, the remaining linkages are established as closure is completed.

In the first feedback approach, we estimate the probability $P(C=c_i|Q=q)$ to find the best class c_i that matches the query q by maximum likelihood. For an ensemble of images $I_{1,c_i}\ldots I_{e_i,c_i}$ belonging to class c_i and obtained through relevance feedback, the likelihood $P(Q=q|C=c_i)=\max_e P(Q=q|I=I_{e,c_i})$. If we assume the user-selected images form "collocation" points in an abstract feature space over which the probability measure is defined, then several measures including kernel density estimate or geodesic distance formulations could be used. The maximal score, tested in the context of trademark retrieval [11], was also found to be robust in other work [35]. The primary reason for the robustness as argued [11] is that each perturbation may reveal a different systematic aspect of the individual in question. This is unlike the situation where an ensemble of normally distributed random perturbations about a statistic (here, the mean) are used. For example, if one ensemble member was partially occluded then, as an outlier, it would not be useful, and the process by which a user construes relevance may decidedly not follow Gauss–Markov statistics. Estimates, such as the mean score over the cohort, therefore, are largely irrelevant.

Aggregating scores by the best match relies on the fact that, given the high-dimensional space that defines a feature, the probability of coming very close to an exemplar is extremely rare and thus, extremely significant. Other ordered statistics were

[3] Nayna Patel, personal communication

attempted but this version performed the best. It is easy to implement and can be used in run-time. For example, in the Sloop system, as the user highlights other matches, the system immediately recalculates. It is frequently seen that the correct match starts to "bubble" up the ranks, as examples in the application will show.

There are, of course, other elements of the representation that can be updated with relevance feedback. In the second approach that is complimentary to the first [11], feature vectors of the cohort are used to calculate new individual centered representations (*e.g.*, covariance or leading modes for PCA) using the perturbed feature vectors gleaned from the selected ensemble. 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, $P(C=c_i|Q=q) \propto P(Q=q|C=c_i)P(C=c_i)$. The right-hand terms are evaluated respectively by comparing the query, using class c_i's principal components, to each member in the relevance ensemble (for likelihood) and by evaluating the probability of the ensemble member using the population statistics (prior). This step has the effect of localizing the principal components to observed variability within a class. Relevance feedback turns out to be very useful. Its role in lifting the performance is illustrated in Section 4.

3.4.1. *Crowdsourcing and social media*

The efficient indexing of large image sets demands a number of rapidly available relevance judgments. Because verification entails ordinary matching skills, crowdsourcing is one way to gather this information. The key question, answered here, is whether there will be sufficiently high quality recognizers in the crowd, and whether they can be efficiently found.

In one experiment, three pairs of images are presented successively to a crowd-user in random order: a positive control pair of a known match, a negative control pair of a known non-match and the experimental pair of unknown status. The user's judgment on the *unknown* pair is accepted when, akin to ReCAPTCHA, the control performance is perfect on the *known* pairs, which is also the condition for payment. The responses are used to update the ranking in the manner described previously.

Fig. 5 shows the results of a crowdsourcing experiment with Mechanical Turk where a HIT is the task as defined above for the marbled salamander. As shown, hundreds of workers can be gathered in a matter of days! The users are paid five to six cents for each verification task, leading to a natural selection mechanism. The same effort would take on the order of a year in grant research with only a handful of workers. The combination of financial incentive and testing produces a situation where candidates who are not good at matching try it only a few times before moving on to a different task while those who are good keep working. One worker in particular is seen to produce a 99.96% recognition rate, answering over 1000 tests in one day! By the time someone performs 40 known-pairs (twenty tests) of comparisons, statistically their recall is likely to be 95% on average (see Fig. 5 in dots). The number of people passing this barrier is about a third of the total population (Fig. 5 in bars).

The self-regulating mechanism for selecting quality workers combined with the accelerating effect of relevance feedback shows the possibility of indexing large databases efficiently. Where human effort can span about 100 images and small teams about 1000 images, this approach can scale well into the 100,000s. Efforts are currently underway with large databases to provide higher levels of recall by the use of human–machine pattern recognition systems. Some applications are discussed next.

4. Application to individual identification

The Earth Signals and Systems Group's Sloop research program is currently engaged in developing individual identification algorithms for several species. These include *Naultinus gemmeus* (jewelled gecko, 10,000 photos/1600 individuals), *Oligosoma otagense* (Otago skink, 8900 photos/900 individuals), *Oligosoma grande* (grand skink, 21,700/2500), *Rhincodon typus* (whale shark, 35,000/3000) and *Ambystoma opacum* (marbled salamander, 10,000/2000). Additional work is in progress on Fowler's toad and is planned for Archey's frog, humpback whales, Hector's dolphins, southern right whales and scree skinks. Here, we discuss the application of Sloop to five species.

4.1. *Geckos*

As a demonstration of baseline performance, a randomized SIFT with bidirectional match is used with and without relevance feedback on a set of 246 extracted gecko patches. The cohort distribution shown in Fig. 6 and patches are exhaustively compared. Relevance feedback is then applied by examining the top five ranked retrievals. ROC curves for SIFT alone and SIFT with relevance feedback (Fig. 7) show that SIFT with relevance feedback is far more effective for the gecko dataset, resulting in an AUC of 99.5%. A full-fledged deployment Sloop-GK is in progress.

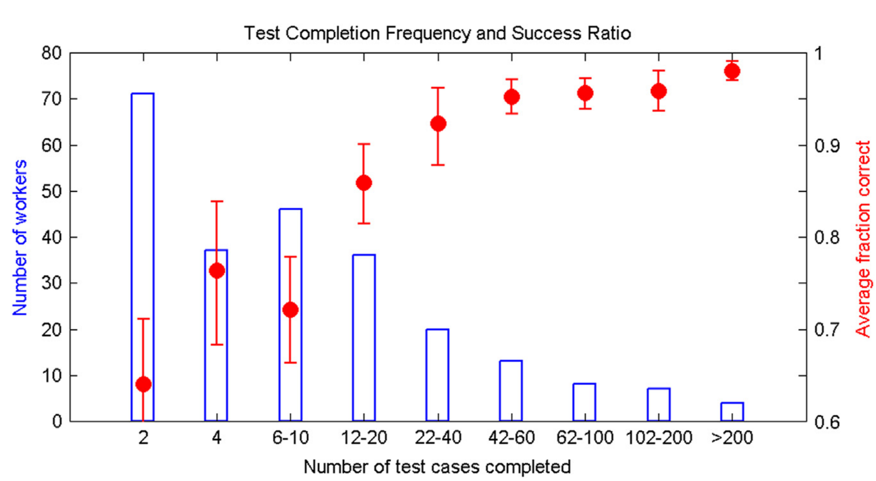

Fig. 5. Persistent test takers are somewhat infrequent (bars). They are also the most skillful (spots). The success rate is strongly correlated to the number of tests that a candidate undertakes, suggesting that persistent candidates are also the most skillful at identification.

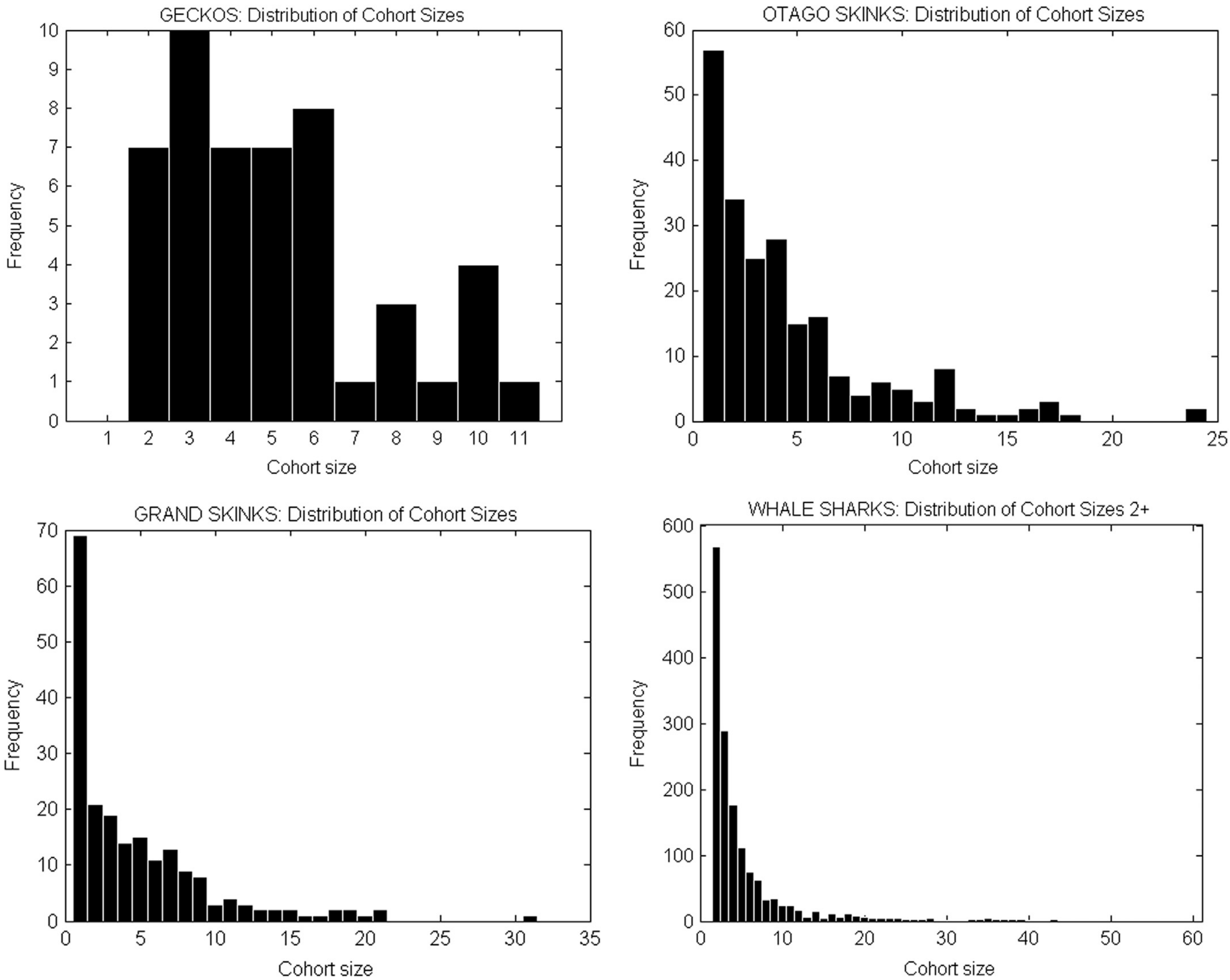

Fig. 6. Cohort size distributions for gecko, Otago skinks, grand skinks and whale sharks. These are the datasets used for experiments described in this paper, and are smaller than the full datasets.

4.2. Grand and otago skinks

There are two Sloop skink systems (Sloop-SK, 2011) currently in use by the New Zealand Department of Conservation for identifying individual Otago and grand skinks. They incorporate photographs of the left and right sides of the animals. These are grouped into capture events which may include photographs of one or both sides of an individual animal. In the Otago skink dataset, there are approximately 4000 captures with both left and right views and approximately 900 captures with only one view. Of all captures with images, approximately 900 unique individuals are identified and associated with more than one capture. In the grand skink dataset, there are currently approximately 10,100 captures with both left and right views, approximately 1500 captures with only one view, and approximately 2500 individual animals with multiple captures. Operational experience with Sloop-SK is discussed in Section 5.

For each photograph, a worker verifies the image quality. A worker then marks four fiducial points on the image (see Fig. 2). These points are used to define patches between the nostril and eye, between the eye and ear, and between the ear and shoulder. The patches are only approximately rigid but are assumed as such and are normalized in orientation

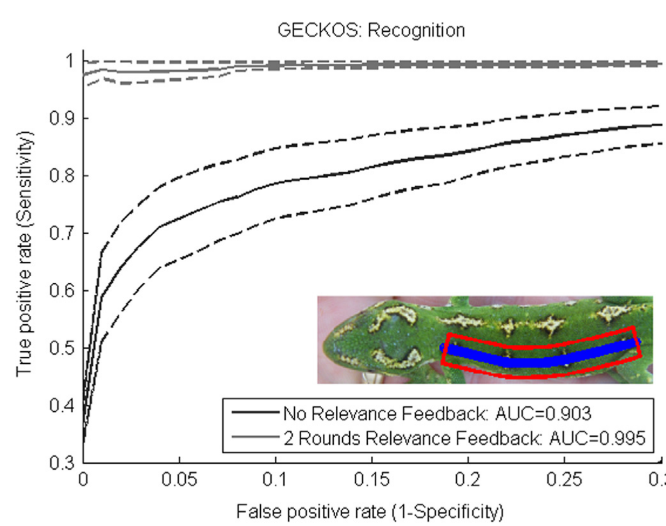

Fig. 7. ROC curve for gecko dataset using SIFT with and without relevance feedback. These graphs indicate that the baseline performance is quite good and improves with coupled human–machine relevance feedback cycles. Two cycles of relevance feedback were applied and the ROC improvement is statistically significant and nonlinear.

and scale. A capture may have up to six patches, three for each side.

In operational use, SIFT features [20] are extracted from the patches and corresponding patches on the same side of the animal are matched. The maximum score over the patch ensemble is used for ranking, removing the effects of low-scoring patches and features. After each capture has a ranking for every other capture, workers verify highly ranked pairs of captures. SURF and ORB features were also extracted for comparisons in this paper, see Section 4.3.

Since the Otago and grand skink systems are active, complete cohort information is not available. Performance is, therefore, tested on subsets of the total Otago and grand skink data where cohort information is completely known. The Otago skink dataset consists of 1002 captures, with the cohort distribution shown in Fig. 6. The grand skink dataset consists of 1008 captures, with the cohort distribution shown in Fig. 6. Following exhaustive comparison and relevance feedback using ten ranked retrievals, ROC curves are generated for captures grouped by number of cohorts (Fig. 8 for Otago skinks and Fig. 9 for grand skinks). The results show that relevance feedback is very effective for the skink datasets across all cohort size categories. Also notable is that SIFT alone produces slightly higher AUC values for grand than for Otago skinks. This agrees with the operational experience described in Section 5, where a slightly higher success rate for grand skinks is found.

As an example in Fig. 10, the value of relevance feedback is shown for a query (left patch image) in the Otago dataset. Relevance feedback causes matches that were not found to bubble up to the top, dramatically accelerating the process of indexing. This is very encouraging particularly when noting that users typically desire performance above 90%. The feedback mechanism can be applied to other methods, too, as discussed in the next section.

4.3. LIFE comparisons for skinks and geckos

In addition to SIFT, SURF and ORB techniques are tested on the skink and gecko datasets. All techniques are applied both with and without relevance feedback as described above. Relevance feedback (on top ten matches for skinks and top five matches for geckos, as above) is shown to improve AUC values by 3–10% (Fig. 11) after one round.

SIFT features are calculated using the VLFeat library [37]. SURF and ORB features are calculated using the mexopencv library [41]. SIFT and SURF features are matched using `vl_ubcmatch`, with the default threshold of 1.5. ORB features are matched within a Hamming distance of 44 (chosen experimentally using a subset of the grand skink dataset).

Although all techniques benefit from relevance feedback, the primary observation of this experiment, we also observe that SIFT performs best overall, a conclusion shared by Town et al. [17] for manta rays. There is a large body of vision work engaged in the

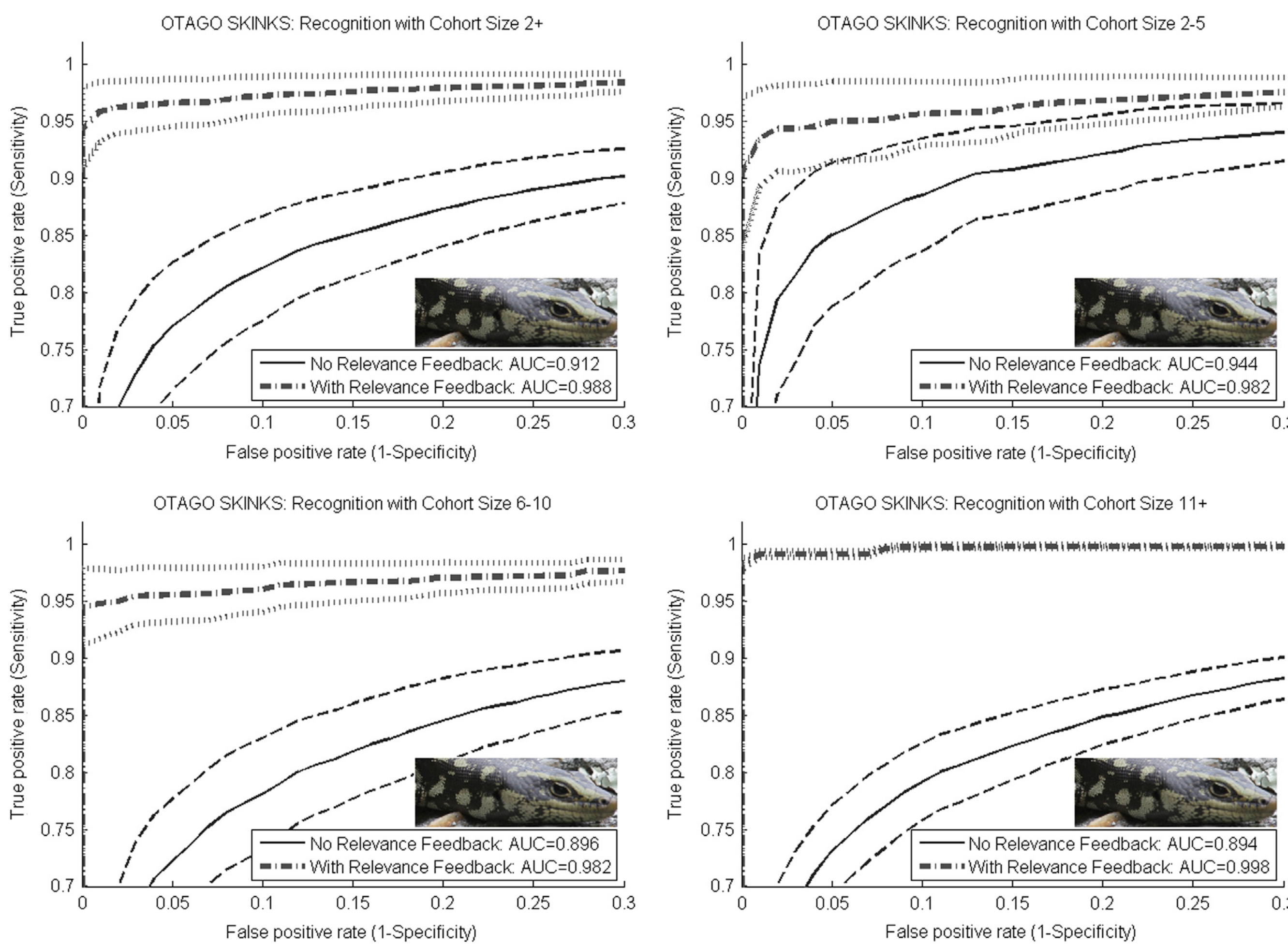

Fig. 8. ROC curve for Otago skink dataset using SIFT with and without relevance feedback at various cohort sizes. These graphs indicate that the baseline performance is quite good and improves with coupled human-machine relevance feedback cycles. In each of these cases a single cycle of relevance feedback was applied and the ROC improvement is statistically significant and nonlinear.

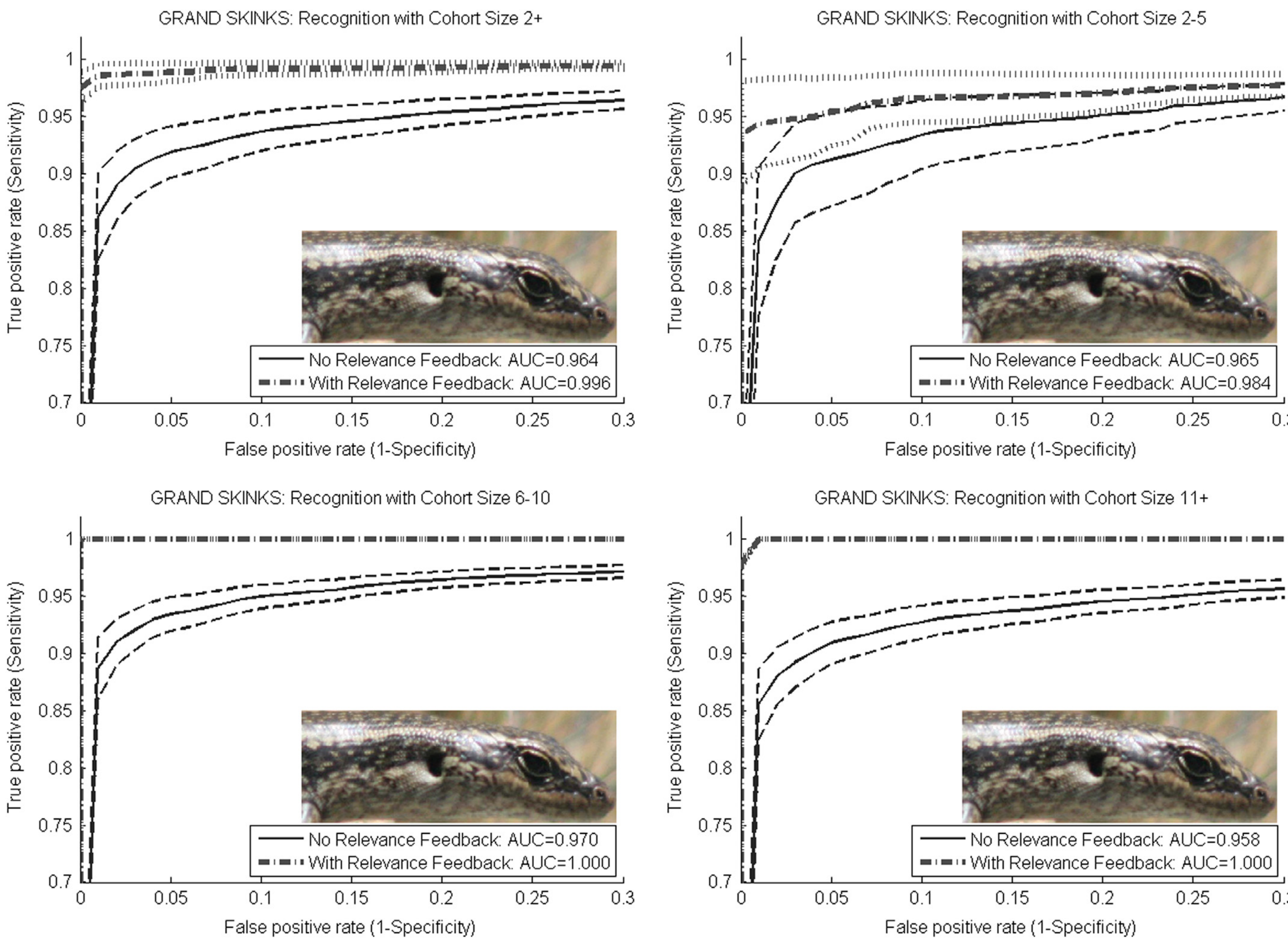

Fig. 9. ROC curve for grand skink dataset using SIFT with and without relevance feedback at various cohort sizes.

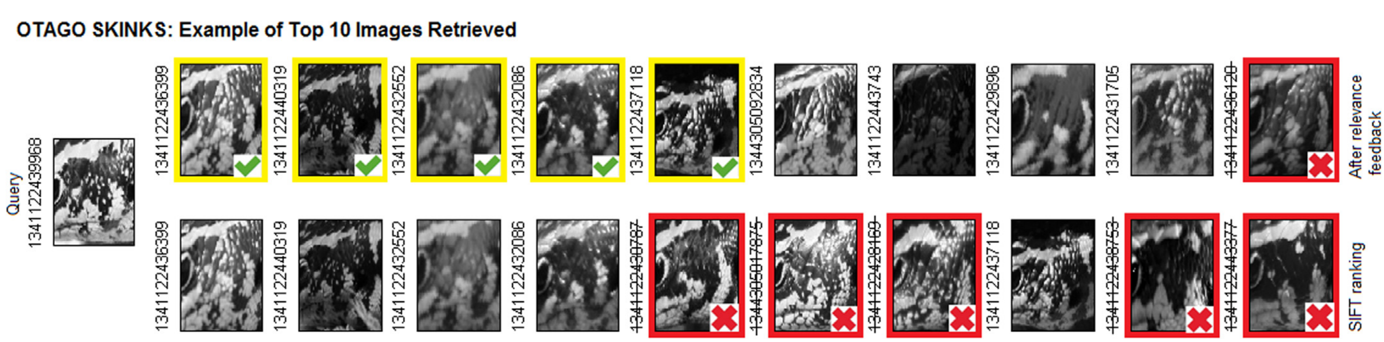

Fig. 10. In response to the query the top ten ranked retrievals show a few mismatches (crosses). Relevance feedback incorporating the few discovered cohorts (check marks) results in retrieval of additional cohorts that the system fails to retrieve at sufficiently high ranks the first time.

development of local features and variants. We emphasize here that, whereas these are claimed to be "generic" methods, often their combination performs better with the coupled human-computational system, lifting performance to the high recall regime that is typically desired.

4.4. Whale shark

The whale shark system (Sloop-WS, 2013) consists of underwater images of a spot patterning behind the gills of the whale shark on one or both sides [6]. Coordinates of the spots are specified by users or extracted (see Fig. 12), and the identification algorithm matches pairs of coordinate sets rather than the images themselves. Spot detectors have been developed, but the provided feature data is used to facilitate direct comparisons of the hybrid context, aggregation and relevance feedback stages with earlier work [6].

Through the process in Fig. 13, a query image's position and scale normalized spot distribution is compared to every other whale shark's similarly normalized spot pattern in the dataset. Pairs of coordinate sets are matched and scored using cascaded correspondence and alignment. In the Eulerian approach, the shape-context is constructed with 16 overlapping bins at two ranges of distances from the spot of interest, and correspondence

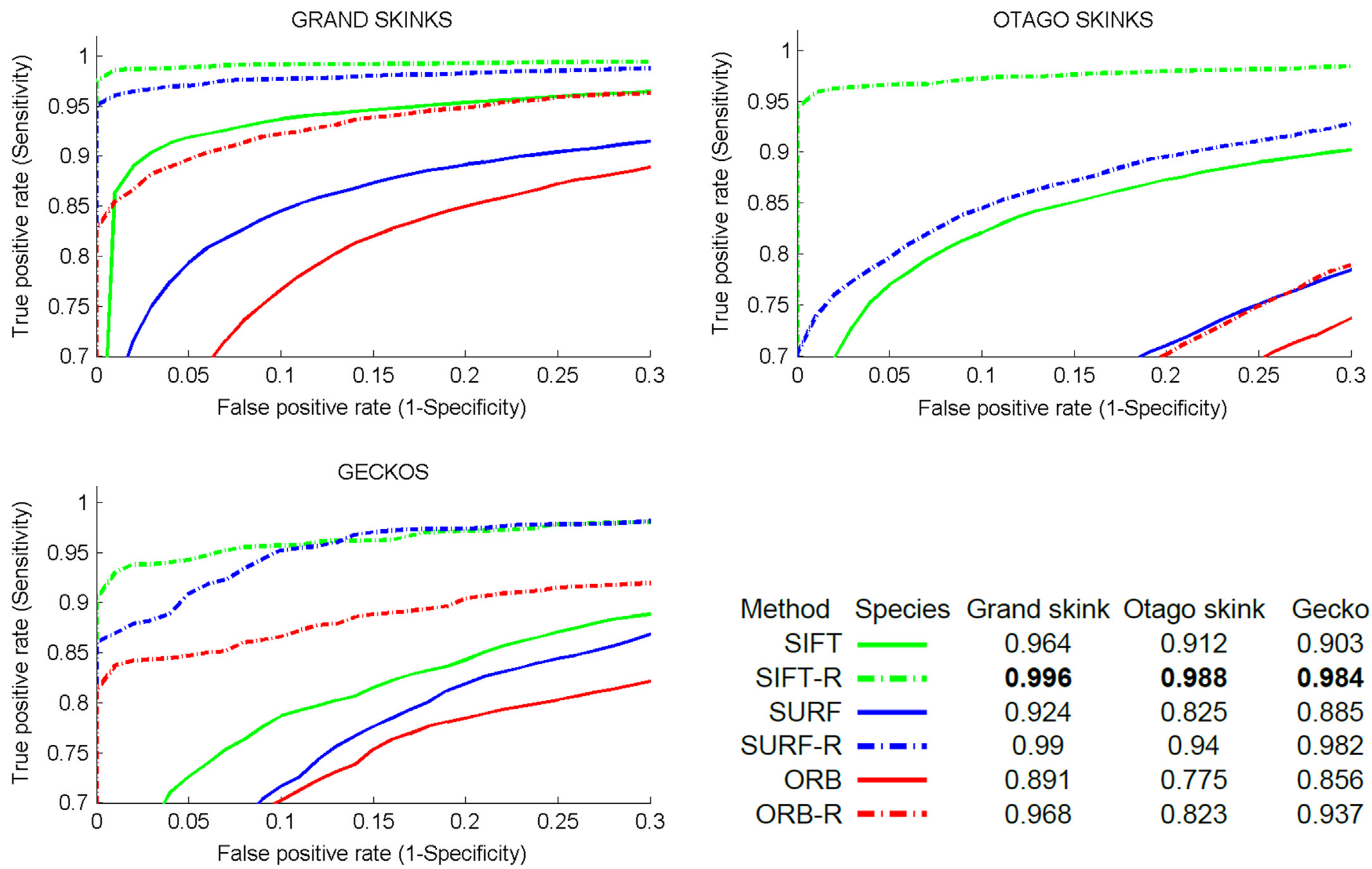

Fig. 11. ROC curves comparing SIFT, SURF, and ORB methods for skink and gecko datasets, with and without relevance feedback. R in the legend indicates that relevance feedback was used. The key shows the AUC values.

between contexts is established using doubly stochastic normalization. The median translation between all corresponding pairs is used to align the points. In the Lagrangian method, a spot's context consists of its five nearest neighboring point coordinates; subsequently, the top three highest matching contexts are used for alignment. After the 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.

To test the algorithm, queries are made on 147 images of 18 individual whale sharks with known cohorts against 5200 or 9300 photos from the database of left or right side photos respectively. The cohort size distribution is shown in Fig. 6. The results of the individual, weak matching algorithms compared to the aggregated version are shown in ROC curves (see Fig. 13). Aggregation of the two algorithms demonstrates a significant improvement over both of the individual weaker algorithms, a benefit that will be observed again in marbled salamanders.

The second set of ROC curves in Fig. 12 reports the improvement achieved through a single cycle of relevance feedback that sets aside previous results [6]. A mechanism of relevance feedback is illustrated in Fig. 14. In this example, five matching encounters are identified in the top 20 results. Through relevance feedback, new matches are retrieved in the top 20. Four of these five new matches are not detectable in the original search because they were solely a left-side view, opposite to the right-sided query. These four matches are identified through the dual-sided records within the cohort of six. One of the five new matches is identified simply by the existence of a closer-matching right-sided image in the new cohort. The value of relevance feedback is greater when there are more encounters of an individual in the dataset (see Fig. 12). This is consistent with findings with the skink species.

4.5. Marbled salamander

The marbled salamander (*Ambystoma opacum*) is an endangered species in Massachusetts. Due to long-term ecological studies in western Massachusetts, there exist a large number of photographs. Earlier work demonstrates the design of an acquisition system in the field, marking medial axes and straightening the salamander as preprocessing, and using patches on the dorsal patterns to compare photographs. The marbled salamander (10,000+ images/2000 individuals) is the first species for which Sloop was developed [1] (Sloop-A, 2009). The earliest technique uses a multiscale local feature histogram method [8,11], and a later technique uses multiscale PCA (MS-PCA) with multiscale Gaussian derivative filter responses on rectified images of animals [1]. MS-PCA applications exist for Fowler's toad and experimentally for skinks. It is comparable in performance to SIFT (see Fig. 15).

To facilitate comparisons, we begin with earlier experiments [9] reproduced here. Approximately 150 queries were run on a dataset of 6000 images and show that, although the classical methods SIFT and MS-PCA are comparable and give reasonably good results, the ROC curves also suggest the need for improvements. Typically for applications, a 90% AUC is competitive but these methods fall

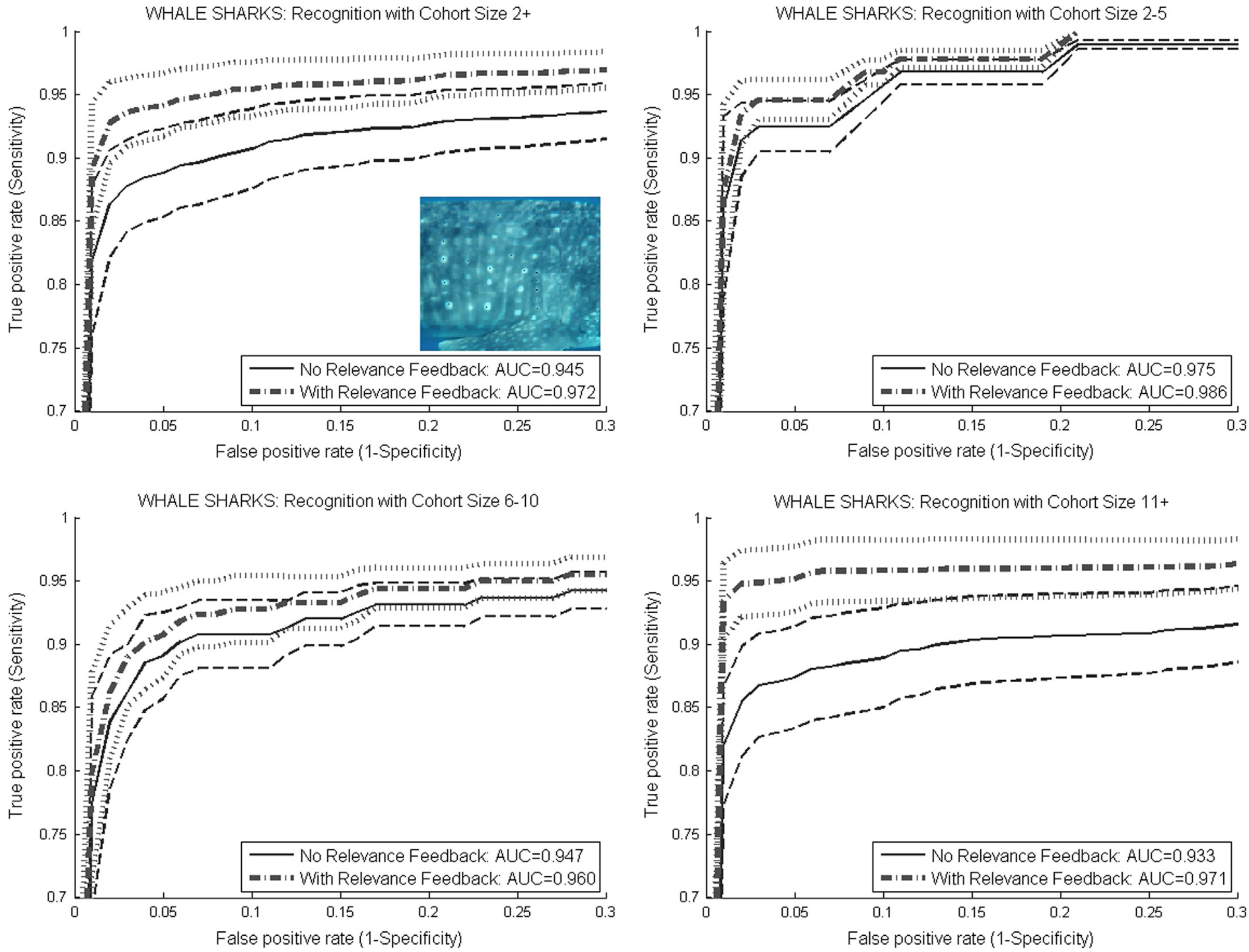

Fig. 12. ROC curve for whale shark with and without relevance feedback at various cohort sizes. These graphs indicate that the baseline performance is quite good and improved with one cycle of relevance feedback.

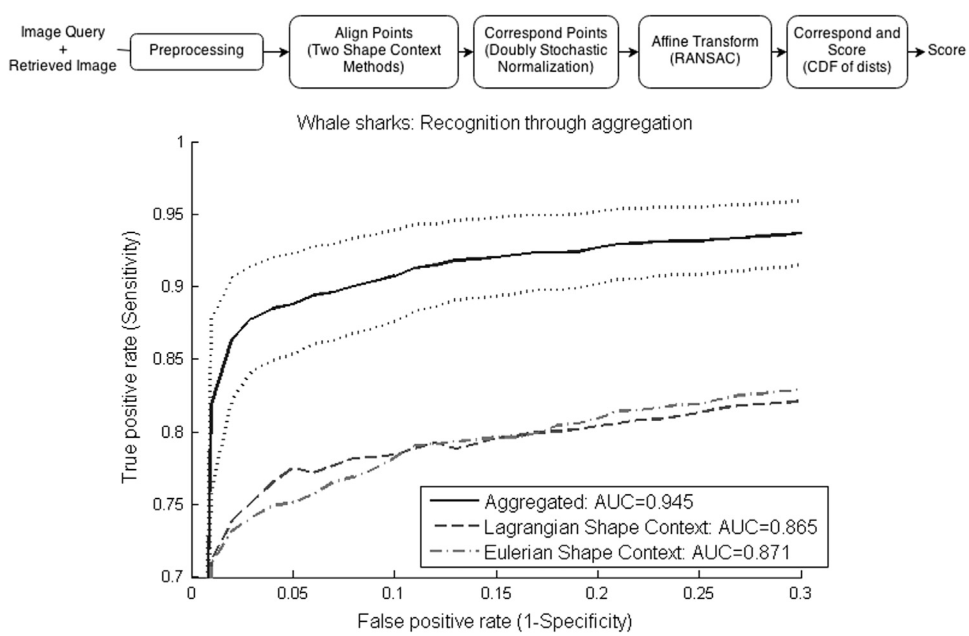

Fig. 13. Flow diagram of weak whale shark score matching algorithm. This process is repeated twice with two different initial shape context alignment methods, and the minimum of the two resulting scores is used to rate the pair of images. This aggregation achieves much higher performance, as seen in the ROC curves.

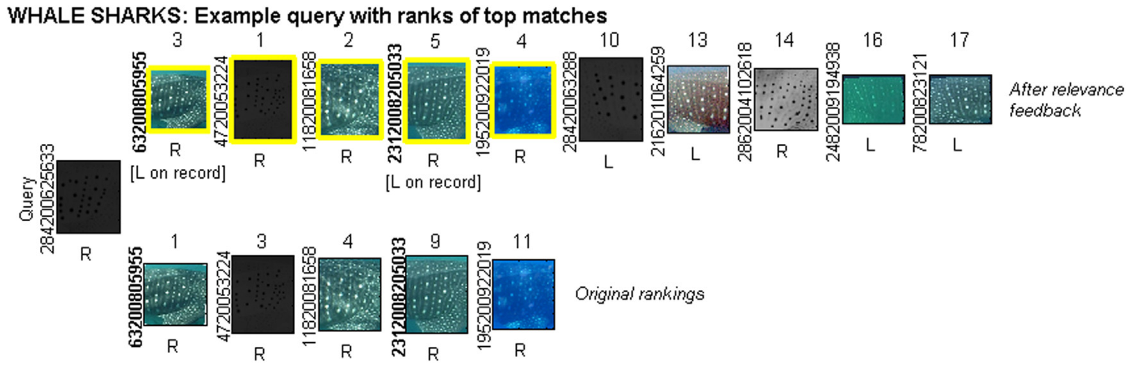

Fig. 14. Retrieval example for whale shark species. Note that there are two modes of relevance feedback in play: the presence of dual-sided records to retrieve left-side views of the particular whale shark, and the original mode of improved matching through increased cohort size.

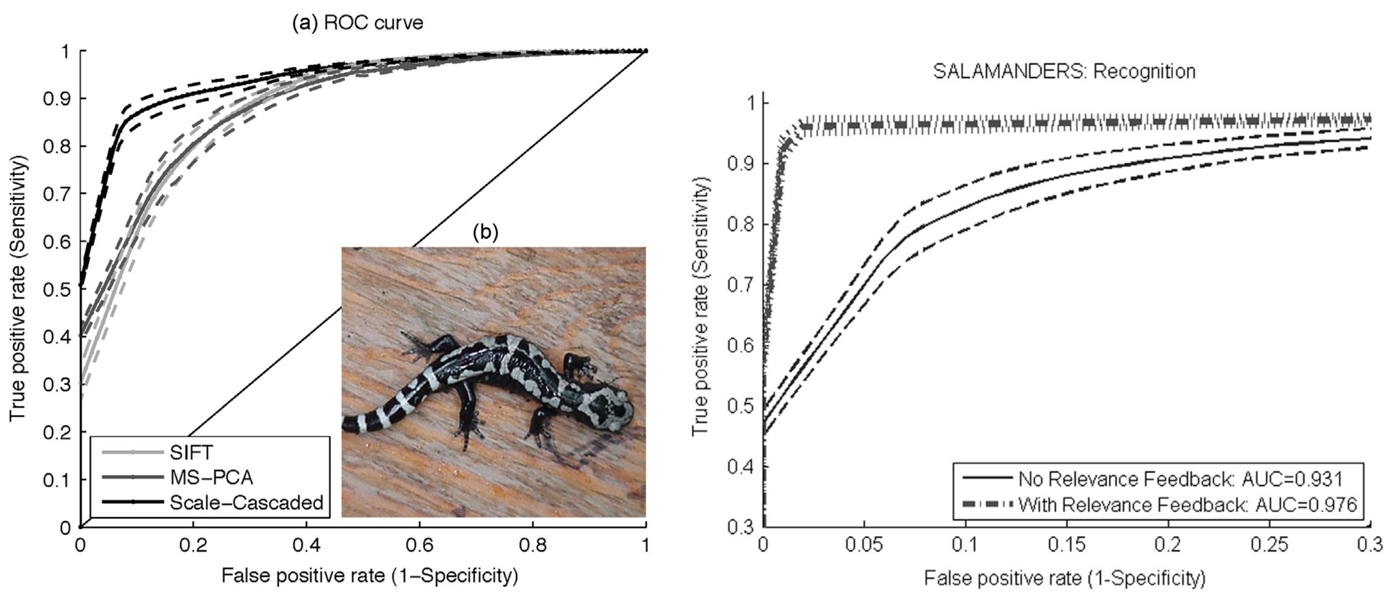

Fig. 15. Reported results on the marbled salamander (left) and additionally with relevance feedback (right).

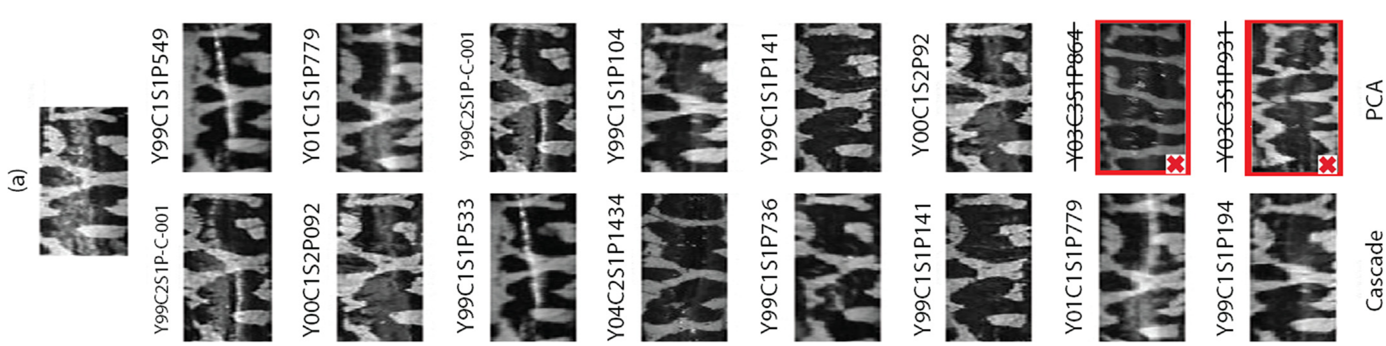

Fig. 16. SCA boosts previous methods, which is further improved by aggregation and relevance feedback in this work.

short. Because the deformations are nonlinear, a boosted approach is used in which SIFT or MS-PCA is followed by SCA refinement. The performance improves, as Fig. 16 shows and as is discussed in [9] to about 90%. Here, the MS-PCA and SCA ranks are aggregated and the result is further improved to AUC of 93.1%. The addition of relevance feedback on the top 5 ranked retrievals results in additional improvement to 97.8%. These results are highly encouraging and they scale to operational experience, discussed next.

5. Sloop in practice: the grand and otago skink recovery programme

The Grand and Otago Skink (GAOS) Recovery Programme was a major effort by the New Zealand Department of Conservation to understand and, if possible, reverse the decline towards imminent extinction of the two critically endangered endemic skink species, *Oligosoma grande* and *O. otagense*. With remnant populations restricted to around 8% of their former range [42], a major protection operation was instigated at Macraes Flat, Otago, NZ. From 2005 to 2008, protected populations increased significantly [43], a result that supported expanding the protection operation to cover approximately 4500 hectares of skink habitat plus buffer zone. That enabled community engagement to institute protection of other isolated remnant populations. The GAOS programme was based on best-practice control of the introduced pests of concern together with robust monitoring to detect skink population and survival changes at various treatment and non-treatment sites [43]. The behavior of the skinks makes full census impossible, but *mark-recapture* techniques can be used to obtain high quality data.

The photographic technique was chosen over approaches based on physical capture and actual marking of animals based on both ethical concerns (disturbance effects), data quality (captured animals become capture-averse, affecting recapture rates), and efficiency.

The photographic technique [43] involved multiple visits to a site on successive days with suitable weather. On each visit, left and right (if possible) lateral photographs were taken of every skink that was found. Because the skinks have markings that are unique to the individuals, these photographs could then be transformed into encounter histories suitable for analysis in a robust design model using program MARK [44]. This analysis produced estimates of both population and annual survival for each site.

The big overhead in the photographic technique became the matching of images to identify individuals. As population increased so did the numbers of photos to match, and the programme became a victim of its own conservation success. By 2011, skilled and dedicated conservationists were undertaking *around half a million manual image comparisons per year and rising*, taking at least twice as much time as the field work required to obtain the photos.

When we switched to Sloop, the Otago skink dataset comprised around 6000 images of 1850 known individuals and the grand skinks around 18,000 of 4750 individuals. The import into Sloop gave an opportunity to check that data, looking for individuals with double identities and individuals that had migrated between sites. Sloop was able to highlight double identities, *finding 15 additional Otago skinks and 67 grand skinks*. This improved data quality and thus provides more accurate results going forwards.

Searching for migrations between sites manually would have required two million extra comparisons and would likely have found few results. Because Sloop presents ranked matches, it was practical to manually check only the top-ranked matches. All correct matches between sites were found in the top three ranked matches. Doing this, several Otago and grand skink movements between study sites were detected, adding to our knowledge of the movement behavior of these animals.

For the 2012/2013 survey year, 550 Otago skinks and 2122 grand skinks required matching. Looking through the whole list of presented possible matches was untenable, especially for the grand skinks. Therefore, if a skink was not matched within the top 100, a manual search took place. If that failed to locate a match then it was likely a new skink. We returned to complete the whole list for the Otago skinks but not the Grands.

Sloop performed better with Grands than Otagos, with Grand matches most often within the top three presented and most Otago matches appearing in the top ten. The matching success rate for Otagos was approximately 96%. For the Grands, Sloop failed to match only seven skinks in the top 100 (approx.), a 99% match rate. The matching effort was reduced to around 4% for Otagos and 1% for grands, however we carried out much more. As this was the first year of using Sloop, we manually searched for every single animal in order to double-check/test the system and (apart from two cases of human error) did not find any errors. At the same time, advantage was taken of the move to the Sloop platform to interrogate the underlying database and feed the encounter histories into script-driven analysis using Rmark [45] as a front end to MARK. This both saved a great deal of time (and error-prone manual process) and allowed a much larger set of candidate MARK models to be explored.

Additionally, in Sloop, matches are based on aggregated scores from all the known cohorts. With manual identification, there is only one library photograph to compare to at a time. Consequently, in cases where a photograph may be poor quality it could take up to three rounds of checking each library photograph before it was picked up.

6. Conclusion

The MIT Sloop system is a community extensible image retrieval system with applications to large-scale conservation. Its main benefits include independent components in the IT, biology and vision realms, and ease of incorporation of new research methods including a number of tools and new algorithms developed by our group, such as the novel cascaded correspondence method discussed above. Sloop is being deployed on multiple species, with operational use in two.

The success of Sloop in the Grand and Otago Skink Recovery Programme is highly encouraging. As the initial discussion showed, perturbed representations, aggregation, and relevance feedback enhance performance, where baseline techniques appear to underperform. As new cohorts form, the complexity of matching will approach the true distribution of identities (as opposed to the collection of photographs) and produce an efficient scalable system.

In these developments, we realize the potential for hybrid systems that optimally utilize human interaction and machine skill to deliver high performance recognition systems. Sloop is available as a web application and a standalone application to users. We invite motivated vision researchers to join Sloop in an Earth Vision endeavor to develop and apply vision and learning tools for effective stewardship of our Earth System.

Conflict of Interest

None declared.

Acknowledgment

This project is performed at the Earth Signals and Systems Group and directed by Sai Ravela, the original designer of Sloop with Chris Yang [9]. This project is supported in part by AFOSR (FA9550-12-1-0313) and NSF DBI-1146747, MIT-Mexico Seed Fund (MISTI), FOMIX CONACYT-GDF (189085) and SIP-IPN (20140325). Any opinion, findings, and conclusions in this material are those of the authors(s). We thank Jason Holmberg for whale shark data, the reviewers for many constructive comments and Julie Marquardt for editorial support.

References

[1] L. Gamble, S. Ravela, K. McGarigal, Multi-scale features for identifying individuals in large biological databases: an application of pattern recognition technology to the marbled salamander *Ambystoma opacum*, J. Appl. Ecol. 45 (1) (2008) 170–180.
[2] M. Kelly, Computer-aided photograph matching in studies using individual identification: an example from serengeti cheetahs, J. Mammal. 82 (2001) 440–449.
[3] L. Hiby, P. Lovell, A note on an automated system for matching the callosity patterns on aerial photographs of southern right whales, J. Cetacean Res. Manag. 2 (2001) 291–295.

[4] S. Mizroch, J. Beard, M. Lynde, Individual recognition of cetaceans: use of photo-identification and other techniques to estimate population parameters, in: P. Hammond, S. Mizroch, G. Donovan (Eds.), Computer Assisted Photo-Identification of Humpback Whales, vol. 12, International Whaling Commission, Cambridge, UK, 1990, pp. 63–70.
[5] B. Araabi, N. Kehtarnavaz, T. McKinney, G. Hillman, B. Wursig, A string matching computer-assisted system for dolphin photoidentification, Ann. Biomed. Eng. 28 (2000) 1269–1279.
[6] Z. Arzoumanian, J. Holmberg, B. Norman, An astronomical pattern-matching algorithm for computer-aided identification of whale sharks *Rhincodon typus*, J. Appl. Ecol. 42 (2005) 999–1011.
[7] N. Kumar, P. Belhumeur, A. Biswas, D. Jacobs, J. Kress, I. Lopez, J. Soares, Leafsnap: a computer vision system for automatic plant species identification, in: European Conference on Computer Vision, 2012, pp. 502–516.
[8] S. Ravela, L. Gamble, On recognizing individual salamanders, in: Asian Conference on Computer Vision, vol. 2, 2004, pp. 741–747.
[9] C. Yang, S. Ravela, Spectral control of viscous alignment for deformation invariant image matching, in: International Conference on Computer Vision, vol. 1, 2009, pp. 1303–1310.
[10] S. Ravela, J. Duyck, C. Finn, Vision-based biometrics for conservation, in: Lecture Notes in Computer Science, Springer, 2013, pp. 10–19.
[11] S. Ravela, On multi-scale differential features and their representations for recognition and retrieval (Ph.D. thesis), University of Massachusetts at Amherst, 2002.
[12] Staff, Program will Identify Skinks, Otago Daily Times ⟨http://www.odt.co.nz/news/dunedin/215070/program-will-identify-skinks⟩.
[13] L. Hiby, P. Lovell, N. Patil, S. Kumar, A. Gopalaswamy, U. Karanth, A tiger cannot change its stripes: using a three-dimensional model to match images of living tigers and tiger skins, Biol. Lett. 5 (3) (2009) 383–386.
[14] L. Hiby, W. Paterson, P. Redman, J. Watkins, S. Twiss, P. Pomeroy, Analysis of photo-ID data allowing for missed matches and individuals identified from opposite sides, Methods Ecol. Evol. 4 (3) (2012) 252–259.
[15] M. Lahiri, C. Tantipathananandh, R. Warungu, D. Rubenstein, T. Berger-Wolf, Biometric animal databases from field photographs: identification of individual zebra in the wild, in: International Conference on Multimedia Retrieval, ACM, 2011, p. 6.
[16] T. Redd, B. Dubansky, M. Osborn, T. Tully, D. Homberger, A registration algorithm for the identification of individual parrots based on the patterns of filing ridges on the internal surface of their upper bill tip, Int. J. Biom. Bioinform. 6 (3) (2012) 68–91.
[17] C. Town, A. Marshall, N. Sethasathien, Manta matcher: automated photographic identification of manta rays using keypoint features, Ecol. Evol. 3 (7) (2013) 1902–1914.
[18] T. Burghardt, N. Campbell, Generic phase curl localisation for an individual identification of turing-patterned animals, in: Visual Observation and Analysis of Animal and Insect Behavior, 2010, pp. 17–21.
[19] H. Kühl, T. Burghardt, Animal biometrics: quantifying and detecting phenotypic appearance, Trends Ecol. Evol. 28 (7) (2013) 432.
[20] D. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vis. 60 (2) (2004) 91–110.
[21] H. Bay, T. Tuytelaars, L. Van Gool, Surf: speeded up robust features, in: European Conference on Computer Vision, Springer, 2006, pp. 404–417.
[22] E. Rublee, V. Rabaud, K. Konolige, G. Bradski, ORB: an efficient alternative to SIFT or SURF, in: International Conference on Computer Vision, IEEE, 2011, pp. 2564–2571.
[23] C. Finn, J. Duyck, A. Hutcheon, P. Vera, J. Salas, S. Ravela, Relevance feedback in biometric retrieval of animal photographs, in: MCPR2014, Springer, in press, http://link.springer.com/chapter/10.1007%2F978-3-319-07491-7_29.
[24] S. Belongie, J. Malik, J. Puzicha, Shape context: a new descriptor for shape matching and object recognition, in: Neural Information Processing Systems, vol. 2, 2000, p. 3.
[25] R. Sherley, T. Burghardt, P. Barham, N. Campbell, I. Cuthill, Spotting the difference: towards fully-automated population monitoring of African penguins *Spheniscus demersus*, Endanger. Species Res. 11 (2) (2010) 101–111.
[26] J. Holmberg, B. Norman, Z. Arzoumanian, Estimating population size, structure, and residency time for whale sharks *Rhincodon typus* through collaborative photo-identification, Endanger. Species Res. 7 (2009) 39–53.
[27] A. Loos, A. Ernst, An automated chimpanzee identification system using face detection and recognition, EURASIP J. Image Video Process. 12 (1) (2013) 1–17.
[28] A. Turing, The chemical basis of morphogenesis, Bull. Math. Biol. 52 (1) (1990) 153–197.
[29] R. Schapire, Y. Singer, Improved boosting algorithms using confidence-rated predictions, Mach. Learn. 37 (3) (1999) 297–336.
[30] S. Xie, S. Shan, X. Chen, J. Chen, Fusing local patterns of gabor magnitude and phase for face recognition, IEEE Trans. Image Process. 19 (5) (2010) 1349–1361.
[31] J. Wright, A. Yang, A. Ganesh, S. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Trans. Pattern Anal. Mach. Intell. 31 (2) (2009) 210–227.
[32] C.-C. Chang, C.-J. Lin, LIBSVM: a library for support vector machines, ACM Trans. Intell. Syst. Technol. 2 (3) (2011) 27.
[33] J. Runge, Reducing spectral reflections through image inpainting (Master's thesis), Massachusetts Institute of Technology, 2009.
[34] S. Ravela, Shaping receptive fields for affine invariance, in: IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, 2004, pp. 725–730.
[35] S. Ravela, C. Yang, J. Runge, Visual recapture for movement ecology at interannual timescales, Workshop on Visual Observation and Analysis of Animal and Insect Behavior, 2008.
[36] S. Ravela, J. Duyck, C. Yang, L. Gamble, SLOOP: photographic animal ID system, ⟨http://tinyurl.com/mitsloop⟩, June 2009.
[37] A. Vedaldi, B. Fulkerson, VLFeat: an open and portable library of computer vision algorithms, ⟨http://www.vlfeat.org/⟩, 2008.
[38] R. Hartley, A. Zisserman, Multiple View Geometry in Computer Vision, second ed., Cambridge University Press, 2004.
[39] S. Belongie, J. Malik, J. Puzicha, Shape Matching and Object Recognition using Shape Contexts, IEEE Trans. Pattern Anal. Mach. Intell. 24 (4) (2002) 509–522.
[40] A. Rangarajan, H. Chui, E. Mjolsness, S. Pappu, L. Davachi, P. Goldman-Rakic, J. Duncan, A robust point matching algorithm for autoradiograph alignment, Med. Image Anal. 1 (4) (1997) 379–398.
[41] K. Yamaguchi, Mexopencv, ⟨http://www.cs.stonybrook.edu/~kyamagu/mexopencv/⟩, 2013.
[42] A. Whitaker, G. Loh, Otago Skink and Grand Skink recovery plan (*Leiolopisma otagense* and *L. grande*), Threatened Species Unit, Department of Conservation, 1995.
[43] J. Reardon, N. Whitmore, K. Holmes, L. Judd, A. Hutcheon, G. Norbury, D. Mackenzie, Predator control allows critically endangered lizards to recover on mainland New Zealand, N. Z. J. Ecol. 36 (2) (2012) 141.
[44] G. White, K. Burnham, Program MARK: survival estimation from populations of marked animals, Bird Study 46 (S1) (1999) S120–S139.
[45] L.J., RMark: an R interface for analysis of capture-recapture data with MARK, AFSC Processed Report 2013-01, Alaska Fish. Sci. Cent., NOAA, Natl. Mar. Fish. Serv., Seattle, WA, 2013.

James Duyck is an undergraduate researcher at MIT, working for the Signals and Systems Group at the Department of Earth Atmospheric and Planetary Sciences, majoring in Mathematics with Computer Science. He has been an intern at Synaptics.

Chelsea Finn is an undergraduate researcher at MIT CSAIL, majoring in Electrical Engineering and Computer Science. She has been intern at Google, and Sandia National Laboratories.

Andy Hutcheon is the Program Manager for the Grand and Otago Skink Recovery Programme at the Department of Conservation, Dunedin, New Zealand.

Joaquin Salas area of research is Computer Vision. Because of his professional work, he received in 2012, from the President of México, the Lazaro Cardenas del Rio award. He has been a consultant for the Palo Alto Research Center (PARC), guest Professor at Oregon State University, Universidad Autónoma de Barcelona, and the University of Duke, and Visiting Scholar at ENST at Brest, and Stanford University.

Sai Ravela directs the Earth Signals and Systems Group (ESSG) in the Earth, Atmospheric and Planetary Sciences at the Massachusetts Institute of Technology. His primary research interests are in stochastic systems science with application to Earth, Atmospheric and Planetary Sciences. He conducts this research through various projects including Autonomous Observation (caos.mit.edu), Animal Biometrics (sloop.mit.edu), Fluid Imaging (flux.mit.edu), Coherent Fluids (stics.mit.edu) and Hurricane Risk (hazmet.mit.edu). Dr. Ravela received a PhD in 2003 in Computer Science from the University of Massachusetts at Amherst, specializing in Computer Vision, Multimedia Retrieval and Robotics.