Vision-Based Biometrics for Conservation*

Sai Ravela, James Duyck, and Chelsea Finn

Earth Signals and Systems Group Earth,
Atmospheric and Planetary Sciences
Massachusetts Institute of Technology

Abstract. Identifying individuals in photographs of animals collected over time is a non-invasive approach that enables ecological studies and conservation planning. Here we propose SLOOP, the first image retrieval system incorporating interactive image processing and matching tools with relevance feedback from crowdsourcing to solve large-scale individual identification for multiple species. One outcome is an advance in matching and image retrieval methodology; another is the creation of a community-based individual identification system that enables conservation planning.

1 Introduction

The development of effective conservation strategies for rare, threatened, and endangered species requires unbiased and precise information on their life history requirements and population ecology. Capture-Mark-Recapture (CMR) studies enable researchers to identify individual animals over time, which is particularly useful for questions related to individual growth and survival, dispersal and movement ecology, and reproductive strategies. CMR studies typically use techniques in which animals are physically marked or tagged. These methods are intrusive to varying degrees and some animals may be difficult to tag efficiently in large numbers. Alternative identification techniques that overcome some of these limitations are needed.

Numerous efforts have been made to identify individual animals using photo-identification methods but often manually and using ad hoc strategies. Manual matching is only feasible for small collections; at $10s$ per comparison, a $1,000$ photograph catalog will take 60 days, and a $10,000$-sized catalog will take about 15 years of nonstop work to analyze. Ad hoc strategies lead to imprecise quantitative analyses and they do not easily port across multiple species. A reusable automated or semi-automated identification system would better advance the application of individual identification in conservation biology.

We contend that image retrieval with crowdsourced relevance feedback offers the basis for large-scale, high performance biometrics for conservation. In this approach, an image retrieval system ranks the images in a database by *visual*

* This material is based upon work supported by AFOSR(FA9550-12-1-0313) and NSF DBI-1146747. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors(s).

J.A. Carrasco-Ochoa et al. (Eds.): MCPR 2013, LNCS 7914, pp. 10–19, 2013.

similarity [1]. Instead of browsing all pairs, the user views a bounded number of retrievals to identify matches. Although not as automated as object recognition, high recall at substantially lower effort than manual matching can be achieved. The throughput can be improved by dynamically incorporating relevance feedback from user-identified matches. A $10,000$-collection will require about three months to analyze. Crowdsourcing can accelerate this process further; the same task is completed in a few hours and at marginal financial cost. To be sure, a *big data* challenge in the form of a large number of visual features and relevance judgments must now be addressed.

The MIT SLOOP[1] system [2–4] is, to the best of our knowledge, the first such image retrieval system for animal biometrics that addresses these challenges. SLOOP has expanded to multiple species. It has an operational implementation at the Department of Conservation, Dunedin, New Zealand and finds mention in popular press [5]. SLOOP is interactive and scalable to distributed implementations and it also easily incorporates new methods in a collaborative framework. In this paper, we describe the SLOOP system architecture and methods, with application to several species.

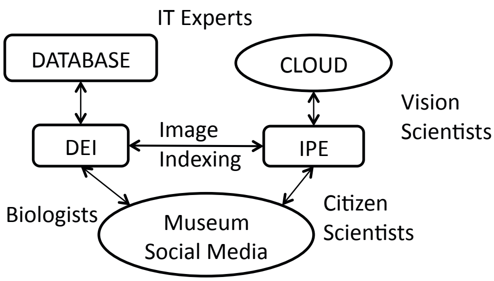

Fig. 1. SLOOP system architecture consists of interactions between a Data Exchange and Interaction (DEI) server and an Image Processing Engine, blending multiple areas of expertise into a single vision application

2 Related Work

There have been many efforts in the ecological community to use pattern recognition methods to aid individual identification. For example, species identification for plants is available [6], but SLOOP predates this system and is geared to individual identification of animals. Several efforts were made to identify individual animals using photo-identification methods [7–11] prior to SLOOP. They are advanced by incorporating new reusable matching techniques with rank aggregation and relevance feedback. The SLOOP system's underlying methods utilize several published techniques [1–4, 12–15].

[1] The name stands for a sloop observed Chris Yang and Sai Ravela saw over the Charles river.

3 The SLOOP System Architecture

The structure of the SLOOP retrieval system is shown in Figure 1. SLOOP is comprised of a Data Exchange & Interaction Server (DEI) and an Image Processing Engine (IPE). The DEI implements the user interface/database (see Figure 5) as a web application running on a GlassFish Server with Postgres binding [16]. The IPE contains segmentation, illumination correction, rectification, matching, relevance feedback, and crowdsourcing tools. It is run as a native MATLAB/Octave server, and can therefore incorporate research codes with relative ease. The separation of IT and Vision components allows independent contributions to be easily absorbed. IPE and DEI interact through a database and, together, they implement a workflow that each image undergoes in its journey from being a photograph to becoming an identity. The end-result is an annotated table is made available to the biologist for subsequent analysis.

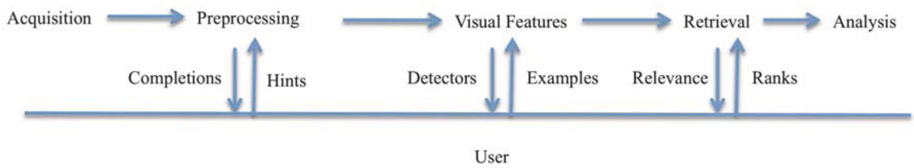

Fig. 2. A SLOOP workflow is an interaction between the users and the system and includes preprocessing, feature extraction, relevance judgement and relevance feedback

The steps involved in a workflow (see Figure 2) typically involve uploading images and metadata, preprocessing images to correct for illumination and geometry, extracting features, producing ranked retrievals, incorporating user judgements, and iterating using relevance feedback. The output of the system is a table with identities associated with each image.

SLOOP enables multiple users to simultaneously work on different aspects of the workflow. IPE and DEI typically operate asynchronously in batch or interactive modes. Some users might be aiding preprocessing, others verification of matches while the IPE might be running rectification and matching jobs on a cloud computing system.

Naturally, synchronization of the system state (tables) becomes necessary for certain tasks. For example, as multiple users verify images, the closure operation linking all matching pairs into a clique through multiple, possibly contradictory, relevance judgements 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 from developing. At the end of each Nightly Build, the SLOOP system is completely consistent, all cohorts share the same identity, new identities have been formed where needed and identities are merged where required. 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, we describe SLOOP's methods along a typical workflow.

3.1 Preprocessing Methods

Images are typically preprocessed to prepare them for feature extraction and matching. Although the preprocessing steps vary between species, segmentation, illumination correction, image rectification and feature extraction are often reused tasks. To facilitate preprocessing, SLOOP contains a Mean-Shift based segmentation algorithm. An additional algorithm is being deployed that uses a combination of SVM and graph-cuts on color-texture features. Rectification includes a spline-based algorithm. For illumination correction, SLOOP includes a variety of global correction techniques, as well as an example-based specularity removal method [12]. Once images are preprocessed, features are extracted. Multiscale patch features, local features including invariant feature histograms [1, 2] SIFT features [13], and point feature extraction methods are most commonly used.

A distinguishing aspect of preprocessing in SLOOP is that it is interactive. In contrast to the traditional view that image processing algorithms need to be fully automated, SLOOP takes an interactive approach where users seed or correct first guess solutions of the algorithms. In this way, SLOOP is designed to maximally reduce human effort but maintain performance. As an example, we show in Figure 3 the specularity removal algorithm within SLOOP [12]. In this approach, the user marks specular and normal regions. Based on this input, SLOOP replaces hypothesized specular spots with information from normal regions to seamlessly in-fill —a task that is otherwise exceedingly difficult [12].

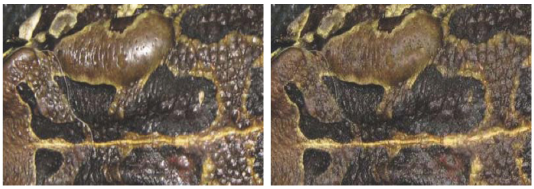

Fig. 3. Interactive example-based specularity removal

3.2 Matching Techniques

SLOOP has appearance-based and geometry-based matching methods, including additions that combine the two. There are four appearance-based methods within SLOOP: multiscale-PCA (MSPCA) [3] on patches and their regularized spatial derivatives, local-feature methods including local-feature histogram [1, 2], SIFT [13], and a new promising method called nonparametric scale-cascaded alignment [4] for deformation invariant matching. In geometry based methods, SLOOP implements an iterative correspondence and affine alignment on point features. These techniques are discussed in more detail in the context of their applications. A common element of all of these methods is that they benefit from interaction in the form of relevance feedback.

3.3 Relevance Feedback

Once matching is complete and ranked hypotheses are produced for photographs in a catalog, the user must verify them. Although the number of images a user must review changes from species to species, typically about $10 - 20$ images at a time are shown. The user selects the matching images (non-matching ones are marked implicitly) and this information is incorporated.

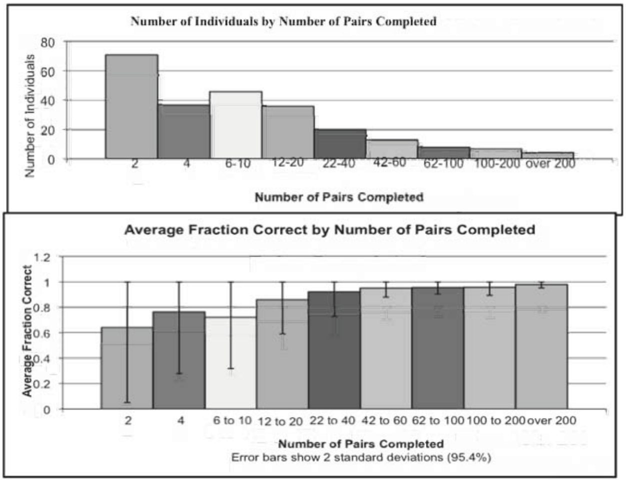

Fig. 4. Crowdsourcing provides sufficient quality for large-scale relevance feedback

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 judgements in two ways: a) the best score from the cohort group is used to rerank images and, b) the population-based prior is replaced by a cohort-based covariance in MSPCA and local feature histogram methods [1]. Relevance feedback turns out to be very useful. It isdemonstrated in the next section.

Crowdsourcing and Social Media: The efficient indexing of large image sets demands a number of rapidly available relevance judgements. Because verification entails ordinary matching skills, crowdsourcing is one way to gather this information. In one experiment, three pairs of images are presented 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. We accept the user's judgement on the *unknown* pair when, akin to ReCAPTCHA, the control performance is perfect on the *known* pair, which is also the condition for payment. The responses are used to update the ranking in the manner described previously.

In Figure 4, we show the results of a crowdsourcing experiment with Mechanical Turk where a HIT is the task as defined above for the marbled salamander.

In a matter of days, we were able to gather hundreds of workers. The same effort took on the order of a year in grant research, and then there were only a handful. We paid the users five to six cents for each verification task. What we found was a natural selection mechanism; the combination of financial incentive and testing produced a situation where users who were not good at matching tried it only a few times before moving on to a different task, and those who were good kept working. One worker in particular, produced 99.96% recognition rate and answered over 1,000 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 Figure 4). The number of people passing this barrier is about a third of the total population, which is very promising.

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 *Oligosoma otagense* (Otago Skink, 8,900 photos/900 individuals), *Oligosoma grande* (Grand Skink, 21,700/2,500), *Naultinus gemmeus* (Jewelled Gecko, 10,000/1,600), *Ambystoma opacum* (Marbled Salamander, 10,000/2,000), and *Rhincodon typus* (Whale Shark, 35,000/3,000). Additional work has been done on Fowler's Toad and is planned for Archey's Frog, Humpback Whales, Giant Manta Ray, Hector's Dolphins, Southern Right Whales and Scree Skinks. Here, we discuss the application of SLOOP to three species. Each uses a different matching technique with the potential for advances to influence methods for all species.

4.1 Grand and Otago Skinks

The SLOOP Skink systems (SLOOP-SK) incorporate photographs of the left and of 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 currently approximately 4,000 captures with both left and right views and approximately 900 captures with only one view. Of all captures with images, approximately 900 individuals have been identified. In the Grand Skink dataset, there are currently approximately 10,100 captures with both left and right views, approximately 1,500 captures with only one view, and approximately 2,500 individual animals.

For each photograph, a worker verifies the image quality. A worker then marks four key points on the image (see Figure 5). These key 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 and scale. A capture may have up to six patches, three for each side. SIFT features [13] are extracted from the patches and corresponding patches on the same side of the animal are compared against each other. The maximum score for any patch is used for ranking, removing the

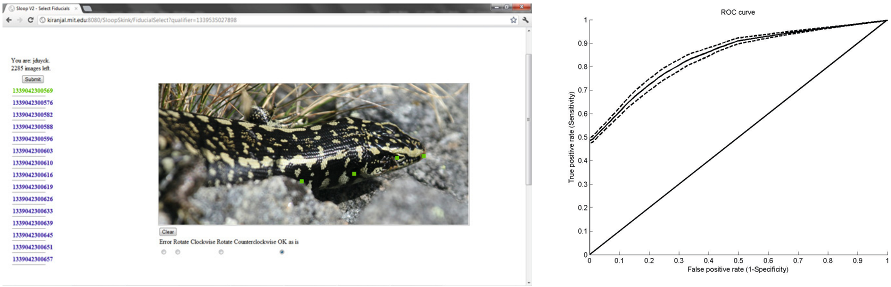

Fig. 5. Otago Skink Dataset with DEI interface showing key points (left) and ROC curve (right)

effect of low-scoring patches. For example, an animal may be partially behind another object, giving a patch an irregularly low score even for photographs of the same individual. After each capture has a ranking for every other capture, workers verify highly-ranked pairs of captures.

The above ROC curve (Figure 5) was found using 1,000 captures (pairs of left and right photographs of an animal) from the Otago Skink dataset. The following groups of animals were used: 282 animals with one capture, 108 animals with two captures, 58 with three captures, 34 with four captures, 17 with five captures, six with six captures, three with seven captures, four with eight captures and two with nine captures. The results indicate that this approach is promising. Additional improvements are possible in terms of rank aggregation, relevance feedback and deformable matching.

4.2 Whale Shark

One application of the SLOOP system is to the whale shark species. The dataset used is a subset of the full data (14,500 images/3,000 individuals). It consists of underwater images of a spot patterning behind the gills of the whale shark on one or both sides. Coordinates of the spots have been specified by users (see Figure 6), and the identification algorithm matches pairs of coordinate sets rather than the images themselves.

The coordinate sets are first normalized for scaling and mean-value. The algorithm then iteratively calculates a correspondence between the sets of points and aligns the points based on this correspondence. In the first iteration, the correspondence is calculated using doubly-stochastic normalization [14] and the points are aligned by a translation. In the second, correspondence calculation is followed by an affine transform calculated with the RANSAC algorithm [15]. The final aligned point sets are scored by determining the area between two CDF curves: the cumulative distribution of distances between corresponding points and that of an ideal perfect match. Thus, lower scores indicate closer matches.

The results of this basic matching algorithm are shown the ROC curves above (see Figure 6). The algorithm was run on seven randomly chosen individuals

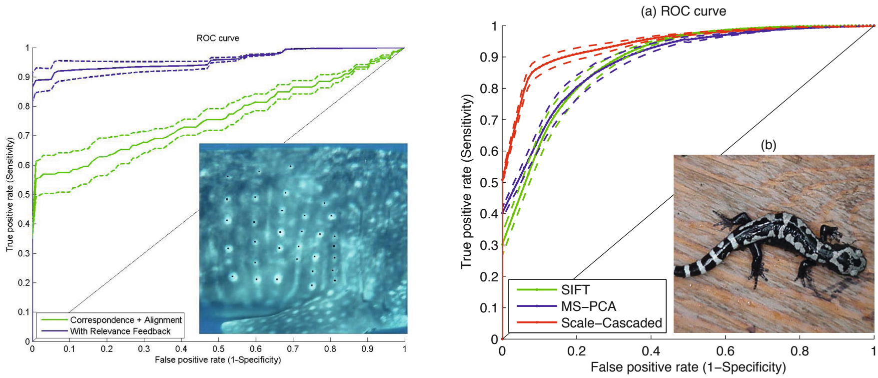

Fig. 6. ROC curves for Whale Shark (left) and Marbled Salamander datasets (right)

each with 5-10 matching images in the database, totaling 56 coordinate sets. The ranks were calculated across all 5,200 or 9,300 photos from the database of left or right side photos respectively. Twenty of these photos and their rankings were randomly chosen to calculate the ROC. To incorporate relevance feedback into the algorithm, the user identifies matches from the 10 best ranking photos, and the ranks are recalculated by taking the minimum score across the new cohort of matching individuals. The user is then shown the 10 new top-ranked photos. This process is repeated until no new matches are identified in the top 10 photos or until the user has viewed a maximum of 50 photos per query. The second ROC curve in Figure 6 reports the improvement achieved through this relevance feedback method. The results show that even for a simple matching procedure, relevance feedback is a powerful addition. Although the improvements to ROC are remarkable, we anticipate adding additional iterations to deal with local deformations in a cascaded framework.

4.3 Marbled Salamander

The marbled salamander (10,000 images/2,000 individuals) was the first species for which SLOOP was developed and applied to [3]. The earliest technique used a multiscale local feature histogram method [1, 2], and a later technique used multiscale PCA (MSPCA) with multiscale Gaussian derivative filter responses on rectified images of animals. MSPCA has also been applied to Fowler's Toad and experimentally to Skinks. It is comparable in performance to SIFT (see Figure 6).

Although the use of rectified images in MSPCA provides some robustness to pose variations, for highly deformable bodies (e.g. salamanders) there is still a substantial amount of local, nonlinear deformation that cannot easily be rectified in advance. Some kind of deformable matching is needed. We developed a scale-cascaded alignment (SCA) method [4] for non-parametric deformable matching which shows considerable promise. To see how it works, consider $J(q) = ||A - T \circ q||_R^2 + L(q)$ an objective that is minimized for deformation q between a template image T and target image A, subject to constraints L, where $T \circ q = T(p - q(p))$.

The vector-field q is dense (and defined at every pixel) and can admit a variety of deformations. For completeness, when the norm R is isotropic and $T \circ q$ is linearized and L represents smoothness, we have classic variational optic flow. If T is not linearized, we have Laplace-Beltrami flow. If R is non-isotropic, we have Ravela's Field Alignment [4] method.

It is useful to note that the objective J can produce a highly selective solution (for example, if L only admits translations) or highly invariant ones (e.g., as a Laplace-Beltrami flow). It follows that the more invariant the matcher is, the less it is able to distinguish dissimilar objects. If the matcher is made more selective, only a few classes of deformations and, thus, only a few objects, can be matched. To optimally adapt invariance and selectivity, we developed Scale-Cascaded Alignment (SCA) [4]. SCA parameterizes the deformation as a weighted sum of Gabor bases with the weights representing a power law turbulence spectrum. This representation can be interpreted as the Green's function of a suitable deformation constraint L. In SCA, the solution is obtained in a coarse to fine manner; the lowest wave number motion solution (translation) is found first followed by higher modes. The total deformation solution or deformation budget is thus dispersed into different modes. Deformations in the low and very high wave numbers are ones to which invariance is desired because they usually correspond to large scale motions or noise. Deformations in middle wave numbers are ones to which selectivity or sensitivity is desired because they help distinguish the natural modes from abnormal ones.

We apply this technique on the Marbled Salamander [3, 4] and the result is shown in Figure 6. SCA outperforms MSPCA and SIFT on the salamander problem and we believe that similar performance improvements could be obtained for skinks and other species. A version for point features is also being developed. If the relevance feedback experiments from Whale-Sharks are also included, we believe all species would substantially benefit.

5 Conclusion and Future Work

The MIT SLOOP system is an extensible community image retrieval system with application 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. SLOOP is being deployed on multiple species, with operational use in two. In these developments we realize the potential for hybrid systems that optimally utilize human interaction and machine skill to deliver high performance recognition systems.

We are developing virtual machines that would simplify future SLOOP deployments. We are investigating rank aggregation from multiple matching methods and a scale-cascaded procedure for iterative correspondence and alignment in the flavor of SCA. We are incorporating graph-based and example-based methods in preprocessing. 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.

References

1. Ravela, S.: On Multi-scale Differential Features and their Representations for Recognition and Retrieval. PhD thesis, University of Massachusetts at Amherst (2002)
2. Ravela, S., Gamble, L.: On recognizing individual salamanders. In: Proc. Asian Conference on Computer Vision (ACCV 2004), vol. 2, pp. 741–747 (2004)
3. 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)
4. Yang, C., Ravela, S.: Spectral control of viscous alignment for deformation invariant image matching. In: Proceedings of International Conference on Computer Vision, vol. 1, pp. 1303–1310 (2009)
5. Staff: Program will identify skinks. Otago Daily Times (June 2012), `http://www.odt.co.nz/news/dunedin/215070/program-will-identify-skinks`
6. 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)
7. Mizroch, S., Beard, J., Lynde, M.: Individual recognition of cetaceans: Use of photo-identification and other techniques to estimate population parameters. In: Hammond, P.S., Mizroch, S.A. (eds.) Computer assisted photo-identification of humpback whales. 12. International Whaling Commission, Cambridge, UK, pp. 63–70 (1990)
8. 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)
9. 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)
10. 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)
11. Kelly, M.: Computer-aided photograph matching in studies using individual identification: an example from serengeti cheetahs. Journal of Mammalogy 82, 440–449 (2001)
12. Runge, J.: Reducing spectral reflections through image inpainting. Master's thesis, Massachusetts Institute of Technology (2009)
13. Lowe, D.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)
14. Rangarajan, A., Chui, H., Mjolsness, E., Pappu, S., Davachi, L., Goldman-Rakic, P.S., Duncan, J.S.: A robust point matching algorithm for autoradiograph alignment. Medical Image Analysis 1 (1997)
15. Hartley, R.I., Zisserman, A.: Multiple View Geometry in Computer Vision, 2nd edn. Cambridge University Press (2004) ISBN: 0521540518
16. Ravela, S., Duyck, J., Yang, C., Gamble, L.: Sloop system web (June 2009), `http://tinyurl.com/mitsloop`