Animal Recognition in the Mojave Desert: Vision Tools for Field Biologists

Michael J. Wilber [1], Walter J. Scheirer [1], Phil Leitner [2], Brian Heflin [1], James Zott [3], Daniel Reinke [3], David K. Delaney [4], and Terrance E. Boult [1],

[1] Securics, Inc
[2] California State University-Stanislaus
[3] U.S. Air Force
[4] U.S. Army Engineering Research and Development Center (ERDC)

Abstract

The outreach of computer vision to non-traditional areas has enormous potential to enable new ways of solving real world problems. One such problem is how to incorporate technology in the effort to protect endangered and threatened species in the wild. This paper presents a snapshot of our interdisciplinary team's ongoing work in the Mojave Desert to build vision tools for field biologists to study the currently threatened Desert Tortoise and Mohave Ground Squirrel. Animal population studies in natural habitats present new recognition challenges for computer vision, where open set testing and access to just limited computing resources lead us to algorithms that diverge from common practices. We introduce a novel algorithm for animal classification that addresses the open set nature of this problem and is suitable for implementation on a smartphone. Further, we look at a simple model for object recognition applied to the problem of individual species identification. A thorough experimental analysis is provided for real field data collected in the Mojave desert.

Figure 1. A photograph exemplifying the challenges of identifying targets in the Mojave desert in a real-world setting. Note that vegetation occludes much of this frame and changing shadows complicate the background. The actual target, in this case a White-tailed Antelope Squirrel, is only a few pixels wide and is not facing the frame; see cutaway in the lower-left corner. We must classify whether this target is a squirrel, a tortoise, or neither; this is inherently an open set problem.

1 Introduction

Computer vision is mature enough to create large impacts in other fields. Standing on the shoulders of decades of vision research, we can now develop tools that amplify other scientists' abilities, even across disciplines. Researchers in diverse fields such as biology and ecology can now depend on computer vision techniques to complement their existing scientific approaches.

Despite our wishes, computer vision alone may never "save the environment," but it can play an indirect role in its conservation by helping field biologists with their own efforts. In this paper, we present a snapshot of our ongoing work to design a system that enables field biologists to conduct animal population surveys using conventional hardware more cheaply and efficiently than manually conducting field observations. We are working with a diverse, interdisciplinary team; this work reflects the contribution of biologists, scientists, and land managers.

The purpose of conducting animal population surveys is to generate statistics about resident animal populations, which are important for many reasons [2, 22]. In particular, fauna surveys help public officials gauge the potential impact of planned additions, expansions, and developments to rural or sensitive land, and they help biologists and ecologists to better understand how various animal populations impact each other and the environment. Continuous animal census statistics also give public officials tools to monitor population changes over long periods of time. It is costly to monitor predation rates and habitat saturation due to the use or lack of numbers of adequate personnel and often the decisions are based on assumptions of populations densities and signs of predation. These statistics are vital for measuring the impact of human-caused climate change, pollution, deforestation, and the effects of continued urbanization. If left unchecked, these effects reduce biodiversity in affected environments [7]. Such changes are especially devastating to animals listed on the US Fish and Wildlife Service endangered species list.

Of particular interest to field biologists is the active population of Mohave Ground Squirrels, Desert Tortoises, and other threatened and endangered animals within the boundaries of the Edwards Air Force Base (EAFB) reservation. The Mohave Ground Squirrel is currently listed as "Threatened" and is native only to the western Mojave Desert [12], the smallest range of any ground squirrel species in the US [21]. The Desert Tortoise is also in decline—some areas that used to contain over sixty tortoises per km^2 now contain 1 to 12 per km^2 by recent estimates [9]. To address these issues, EAFB harbored an ongoing monitoring and conservation effort between 2003 and 2007 to conserve the Mohave Ground Squirrel [12] and contributed to similar programs designed to conserve the desert tortoise [17]. Practically, it is difficult to gain an accurate assessment of these animal populations because fauna surveys traditionally depend on complicated quantitative models [24], manual observation [3] or actual trapping [22], which can be expensive, invasive, and time-consuming. As a result, these surveys can provide insufficient coverage of the survey region. Automated surveys do not have these issues. Unlike human volunteers, cameras never grow tired, never look away, and can be deployed in great numbers to survey continuously over long periods of time.

To perform automated population studies, biologists must traditionally invest scarce funding into complicated, specialized hardware. They must also invest their time and energy – both of which are better spent doing research – setting up such capture systems, testing parts, and fixing problems. Incorrectly configured capture hardware may lead to poorly captured data, skewing the results of the survey. However, with the proliferation of cheap cameras such as those found on smartphones, researchers can begin to use familiar, inexpensive capture hardware laying around the lab. When combined with the commoditization of vision algorithms, this can help solve some of these problems.

In our work, we use lightweight algorithms that run on inexpensive mobile hardware. Our eventual goal is to develop the system to the point where field biologists can use inexpensive smartphones or cameras connected to laptops to easily gather census data. Images and video are periodically captured, animals within the frame are detected, and their types are determined using feature extraction and classification algorithms. A second stage classifier identifies the specific species, if needed. Finally, the results are tabulated and presented back to the biologists.

Real-world vision tasks are not trivial. The environment, the experiment, and the animal subjects themselves provide significant vision obstacles. In the Mojave desert, our system faces complex backgrounds and varying illumination. Our images exhibit ambient movement from shadows and plants swaying in the breeze. Occlusion from plants, shrubs, and other vegetation limits the viewing distance. The behavior and geometry of the animals themselves also make classification difficult—due to foreshortening, animals will have differing scales, requiring algorithms that are scale-invariant. Animals may be self-occluded and may assume any pose or posture. Fig. 1 highlights some of these challenges. In this picture, even a human may have a hard time finding the small squirrel among the rocks in the scene.

When testing our system, we are only interested in a few species such as the Mohave Ground Squirrel and the Desert Tortoise. However, there are many other desert objects that may come into contact with our sensors: insects, lizards, snakes, birds, tumbleweeds, thirsty humans, and so on. When classifying target samples, as we determine *what kind* of animal they are, we must also determine *whether* they are interesting at all. As such, an animal population study is an inherently *open set* problem because it is impractical to create a training data set for all possible negative examples (*i.e.* we cannot create a set containing "everything that is not a Mohave Ground Squirrel") [6]. This problem complicates classification algorithms that assume that examples from all possible negative classes are known at training time (for example, a multiclass SVM).

Our main novel contribution is designing and implementing a lightweight system to recognize animals in the desert. We introduce a novel algorithm based on LBP and SIFT for feature extraction, and we utilize 1-class SVMs that re-frame the recognition scenario as an open set problem. We also examine the feasibility of individual species identification using another simple object recognition approach. Finally, we present an analysis of our system's recognition performance using real field data collected in the Mojave desert.

2 Related work

Other systems exist to gather animal population statistics, but none of them are applicable to our situation. Some systems require massive amounts of computing resources to perform well. Others make inappropriate assumptions, for example, only tracking one target at a time in the frame or assuming a known number of targets. Note that very few existing survey systems actually classify the target objects, instead focusing on just detection and segmentation. This may be fine for some constrained problems, but because we intend our system to observe many different kinds of animals, we must take extra pain to classify each.

In [11], Kembhavi *et al*. uses vision techniques to track the Satin Bowerbird's courtship across 200,000 frames of video. Primarily concerned with localizing and counting these birds, this work's segmentation method requires considering every pixel for every frame in the video to build the background model. This is computationally expensive in terms of time, CPU, and memory; to make this feasible, they used a workstation cluster to process multiple videos in parallel. According to [11], "If the entire [model] structure were to be in memory at one time, it would require 100s of GB of memory, rendering this task impossible for even a modern PC." We wish to use lightweight techniques suitable for near-real-time processing using laptops and smartphones without having to rely on computing clusters.

Other work by Dickson *et al*. focuses on segmenting seabirds [4, 5] on a small stretch of seaside cliffs. To segment birds from video, they use region-based segmentation with Markov random fields; for still images, they use haar-like features to detect birds. Later work by Qing *et al*. [16] uses boosted HOG+LBP features and SVM classifiers. They do not attempt to classify the animals.

In [3], Cohen *et al*. present a preliminary system with the similar goal of detecting and identifying threatened animals in the Mojave desert from an engineering and design perspective. This work also briefly outlines methods for identifying individual tortoises (i.e. determining whether a specific individual was seen before), which can provide field biologists with more insightful data [2]. To segment foreground from background, they describe both a codebook segmentation method and a method based on running average background subtraction and bounding box merging. They do not describe the method used to identify species in the captured video, but they do mention that targets are classified based on pixel statistics ("size and color information"). Such statistics may not be appropriate in our case where foreshortening causes targets to assume varying size and where varying lighting, occlusion, and shadows changes the subjects' color.

Biologists can alternatively use off-the-shelf commercial surveillance and tracking software, but because it is usually designed for law enforcement and security scenarios, such software is often ill-suited to performing fauna surveys. For example, "Knight" by Shah *et al*. [20] can detect and track multiple targets through multiple cameras. However, in being designed for "real world scenarios ranging from railway security to law enforcement," it only distinguishes between individual people, groups of people, and vehicles using "color, shape, and motion models." As before, even if classifiers for individual species may be trained, simple statistics alone may not capture enough information to be useful in our scenario.

Some researchers improve accuracy by keeping the humans "in the loop," having humans perform tasks that are difficult for a purely automated system. Branson *et al*. [1] gives human operators high-level questions about the animals being classified ("Is the tail bushy? Is it uniformly brown?") while computer vision techniques narrow the decision tree. This provides stronger performance at the cost of requiring operator intervention. To classify, they use per-class 1-vs-all SVM classifiers. This makes this method applicable only to closed-set recognition problems.

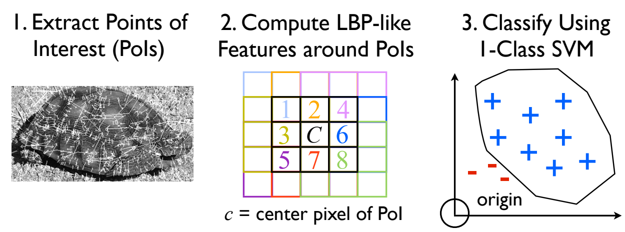


Figure 2. An overview of the animal classification approach. The algorithm is designed to be suitable for near-real time processing on a smartphone for field work, yet accurate enough to yield useful recognition results.

3 A Lightweight but Effective Approach to Recognition

As noted above in Sec. 1, animal population studies in natural habitats present new recognition challenges for computer vision. Specifically, we must address the open set nature of recognition in this scenario, as well as consider classes of algorithms that are "lightweight," meaning they are suitable for quick processing on hardware as limited as a smartphone. At this stage in our project, recognition is a two step process: we first determine if an object in a scene is an animal of interest (Squirrel or Tortoise) and then attempt to distinguish between species (Mohave Ground Squirrel vs. White-tailed Antelope Squirrel vs. Round-tailed Ground Squirrel) if appropriate. Below, we introduce a novel approach for test animal classification and describe our strategy for individual species identification.

3.1 Animal Classification

For animal classification, we make use of a feature-based learning approach incorporating an LBP-like operator and 1-class Support Vector Machines. An overview of this approach is shown in Fig. 2. The underlying features used for classification are generated by extracting points of interest (PoIs) from the images using Difference of Gaussians as proposed in [13] for the well-known SIFT method, and then computing an LBP-like [18] feature descriptor in a window around each detected PoI, somewhat similar to [8]. Feature vectors for learning are composed of histogram bins that summarize the feature descriptor information for each sample image.

Theoretically we could just use SIFT features, but they are "invariant" only for planar objects and our animal objects are not particularly planar. Instead, we opt for an LBP-like feature computation. By starting with the same PoIs that

SIFT uses, we gain the stability of the localized feature regions across objects while leveraging a stronger descriptor. This approach is designed to produce good results with limited computation given the difficulty of this problem: only pixel neighborhoods around each PoI are considered for feature computation. In this work, we rely on the PoIs to focus on a particular object since at this point in our project, we are just concentrating on recognition. However, more sophisticated object detection applied over broad scenes can also accomplish this.

More formally (drawing from [18]), consider a pixel with neighbors $j = 1 \ldots n$. Here, c stands for the center pixel of a PoI determined by the SIFT algorithm of [13] and j for a neighboring pixel. We need to extract a feature representation for each PoI. For each pixel c, the generalized binary representation is defined as:

$$GR(c) = \sum_{j=1}^{n} g_j(c, j) \cdot 2^j. \tag{1}$$

In the standard LBP pattern, a neighboring pixel is defined as a label, 0 or 1, indicating a comparison of its value with the center pixel value c. If the neighboring pixel is greater than c, it is assigned a label of 1; otherwise, it is assigned a label of 0. Using this simple comparison makes LBP undesirably sensitive to noise and scale. The LBP-like operator we use here improves on LBP in three ways. First, we do not compare two individual pixel intensities; rather, we compare the average intensities of neighborhoods of pixels that surround the two. Second, we compare intensities with a threshold, e_N, which is determined by a statistical analysis of the expected level of noise for the sensor data when summed or blurred to level N. Finally, we change the order that neighborhoods are compared when generating the descriptor to ensure that any two neighboring directions in the image are never more than a factor of 4 away in the resulting binary encoding. Each label of this new LBP-like operator is:

$$g_j(c, j) = \begin{cases} 1 & \text{if } \|S(j) - S(c)\| > e_N \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

where $S(x)$ is the average intensity of the neighborhood of pixels surrounding x, as in [18]. If the difference between the center neighborhood intensity and the other neighborhood intensity is more than the threshold, the neighboring pixel is assigned a label of 1. This gives us a binary representation of the center pixel corresponding to the pattern of comparisons of neighborhoods around the center pixel, enabling us to use the surrounding pixels in a more meaningful way than standard LBP. The major advantages of this feature descriptor over standard LBP for our problem are:

1. **Scale Invariance** – we observe a very large variation in object scale for real-world images collected in the field, with animals moving freely towards and away from the acquisition device. This is addressed by calculating the average intensity of the neighborhood of pixels surrounding a center pixel at different operator sizes (3x3, 5x5, 7x7, etc.), which captures multiple scales.

2. **Noise Tolerance** – our imagery is captured in the field, so we expect some measure of minor variation and noise. The threshold e_N defined by Eq. 2 addresses this issue by making sure only statistically significant differences are considered. Comparing neighborhoods of pixel values also reduces the effect of small-scale sensor noise.

3. **Variation Tolerance and Rotation Invariance** – with no control over our animals of interest, different positions and poses will be observed. The alternative numbering of neighborhood pixels (Fig. 2) ensures that any two neighboring directions in the image are never more than a factor of 4 away in the resulting encoding. This increases stability if there are minor variations in the edge features. Adjustment of the operator scale provides a measure of rotation invariance by anchoring pixel orientation.

Each object may have varying numbers of SIFT PoIs, and thus we may have varying numbers of feature vectors generated by our LBP-like operator. The single-vector representation of the object is the normalized sum of the vectors generated by the LBP-like operator on each PoI.

Once we have a feature vector for a given sample, we can apply machine learning. The 1-class SVM introduced by Schölkopf *et al.* [19] adapts the familiar SVM methodology to the open set recognition problem. With the absence of a second class in the training data, the origin defined by the kernel function serves as the only member of a "second class." The goal then becomes finding the best margin with respect to the origin. The resulting function f after training takes the value $+1$ in a region capturing most of the training data points, and -1 elsewhere.

Let $p(x)$ be the probability density function estimated from the training data $\{x_1, x_2, \ldots, x_m \mid x_i \in X\}$, where X is a single class. A kernel function $\Psi : X \to H$ transforms the training data into a different space. To separate the training data from the origin, the algorithm solves a quadratic programming problem for w and ρ to learn f:

$$\min \frac{1}{2} \| w \|^2 + \frac{1}{\nu m} \sum_{i=1}^{l} \xi_i - \rho \tag{3}$$

subject to

$$(w \cdot \Psi(x_i)) \geq \rho - \xi_i \quad i = 1, 2, \ldots, m \quad \xi_i \geq 0 \tag{4}$$

In the 1-class SVM, $p(x)$ is cut by the margin plane minimizing Eq. 3 and satisfying Eq. 4. Regions of $p(x)$ above

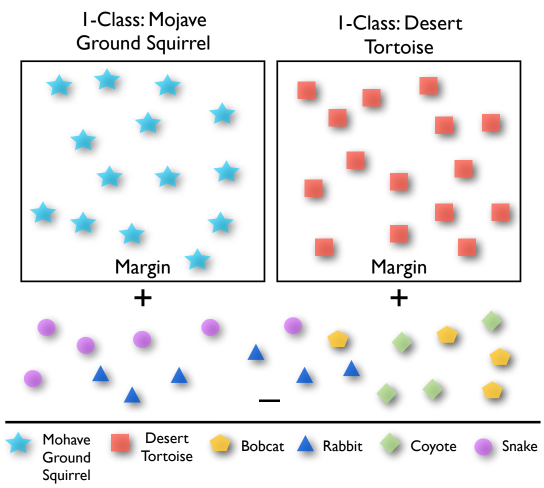


Figure 3. The 1-Class SVM approach for open set problems. In traditional classification problems, we often consider a closed set, where all possible classes are known to the classification system. For the problem we consider here, we must assume that our candidates for recognition can be any creature found in the Mojave desert (squirrels, tortoises, bobcats, humans, etc.). Acquiring the appropriate training data for all possible animals is not feasible, thus, we build classifiers only for the animals of interest.

the margin plane define positive classification and capture most of the training data. The kernel function Ψ impacts density estimation and smoothness. The regularization parameter $\nu \in (0, 1]$ controls the trade-off between training classification accuracy and the smoothness term $\| w \|$, and also impacts the choice and number of support vectors.

We use the 1-class SVM formulation to train our animal classifiers. The 1-class SVM gives us the flexibility to handle any "unknowns" that might be submitted to a classifier (Fig. 3). For our problem, these could be any new object that enters the desert scene: bobcats, rabbits, coyotes, snakes, etc. Acquiring the appropriate training data for all possible animals is not feasible, thus, we build classifiers only for the animals of interest, and make a positive or negative determination with respect to them.

3.2 Individual Species Identification

A particular challenge of this project is the need to distinguish between different species of ground squirrel. Experts use visual criteria to distinguish between squirrels in the Mojave desert (shown in Fig. 4). The common White-tailed Antelope Squirrel has a white lateral stripe on each side of its body, as well as tufts of white hair on the backside of its tail, and ears that stick up a bit higher compared to the two other ground squirrels found in the Mojave. The Mohave Ground Squirrel is uniformly brown on the sides, with light fur on the back of the tail, but not as white as the Antelope Squirrel. The Round-tailed Ground Squirrel is very similar in appearance to the Mohave Ground Squirrel. The main differentiating feature is its tail, which is longer, thinner,

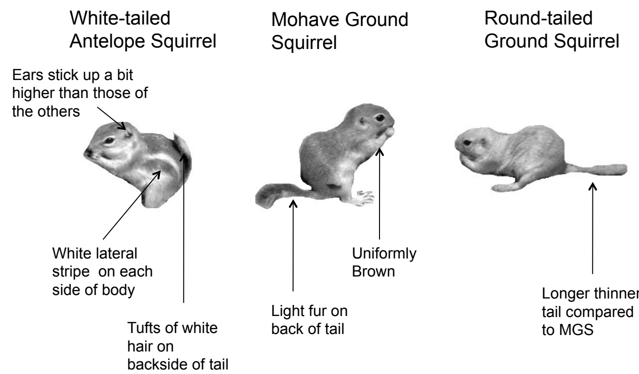


Figure 4. Segmented images and distinguishing characteristics for three different species of ground squirrel. From the left: White-tailed Antelope Squirrel, the Mohave Ground Squirrel, and the Round-tailed Ground Squirrel. Our experiments show that we can distinguish between these three species with approximately 78% accuracy.

and darker. We note that the Round-tailed Squirrel lives to the east of EAFB, but is considered here due to its striking similarity to the Mohave Ground Squirrel. Taking these features into consideration, we selected a different recognition algorithm for this task.

The "V1-like" recognition algorithm of Pinto *et al*. [14, 15] is a simple biologically inspired model of the known properties of the primate first stage visual processing area. For this algorithm, each training image is first filtered by an array of 96 Gabor wavelets, generating a large set of feature vectors. PCA is used to reduce the dimensionality of these feature vectors prior to using them to train a multiclass SVM. Due to the nature of this method of classification, several training images are used for each class so as to increase the accuracy of the SVM's convergence. In the model of Pinto *et al*., the input images during testing are treated the exact same way, with each resulting feature vector classified by the trained SVM. We chose this algorithm as a second-stage classifier for its relative simplicity and excellent baseline performance on popular data sets. Note that assuming we know that a sample is a squirrel, determining its species is a closed set problem.

4 Evaluation

An exciting aspect of this work is the intersection between field work in biology and computer vision. For our experimental evaluation, we relied upon our field biologists to collect animal data in the Mojave desert, but also used a recent imaging technique to create semi-synthetic data that captures even more environmental variation. Our data collection methodology and experimental results are described below.

4.1 Data Collection

Our set of squirrel data was collected during fieldwork in April 2010 at the height of the Mohave Ground Squirrel season around the Fort Irwin area. The data was collected as MPEG video using a trail camera system from RECONYX[1]. Animals were lured into the field of view using feed positioned at the center of the captured frames. In total, 5,362 frames were extracted from the provided videos, including images of the White-tailed Antelope Squirrel, Mohave Ground Squirrel and Round-tailed Ground Squirrel. Our set of tortoise data consists of several long video sequences of different desert tortoises from various sites around the American Southwest, also collected during fieldwork in 2010. In total, 450 frames of tortoises were extracted from those videos for our evaluation. Additional public data collected from the web was used for the following other animals (50 images each): bobcat, coyote, Gila monster, rabbit, and snake. For animal classifier training, 100 images were selected from the squirrel or tortoise pool at random. For testing, 50 images were submitted to the classifiers for each animal (for the squirrel and tortoise classes, these images did not overlap with the training data).

To increase the scope of environmental conditions in the scene for laboratory testing, we adapted "semi-synthetic modeling," a technique used in human biometrics [10, 23]. Semi-synthetic models use measured data, such as 2D images or 3D renderings of an object as the model, but rather than modeling the imaging system, they are incorporated into a real system for evaluations. This is called semi-synthetic because the underlying data is no longer really a synthetic model, but a re-rendering of measured data instead. Semi-synthetic models are derived from individual source object data and hence can capture properties that are never explicitly modeled, *e.g.* distributions of textures, geometry, and, in the case of our fauna problem, hair.

When collecting long-range data for animal population studies, several problems exist including weather and atmospheric effects (distortion caused by thermal aberrations in the atmosphere). Using a version of the above methodology, we collected the data necessary to show the feasibility of our algorithms when considering animals at a distance. We re-imaged our data set outdoors at 100M on a sunny day using the setup described in [10, 23]. This consisted of a 4000 ANSI lumens projector with a resolution of 1024×768 projecting into a custom built screen, imaged by a Canon EOS 7D Camera with a Canon $2\times$ adapter and a Sigma 800mm F5.6 lens. The resulting images (an example is shown in Fig. 5) exhibited realistic levels of atmospheric blur, as well as some motion blur as the wind made contact with the tripod (an important consideration during image acquisition). This produced 450 images across all animals in the same configuration as described above for the original data.

[1] http://www.reconyx.com/

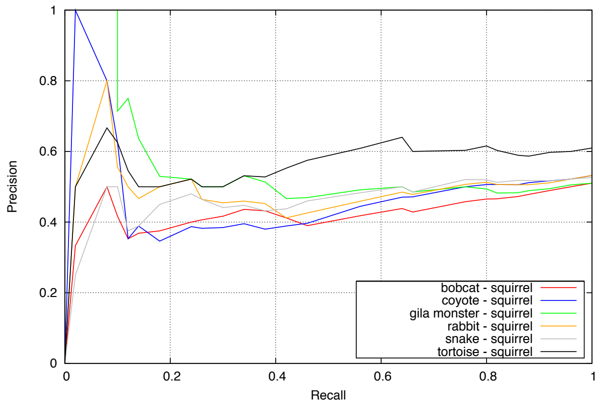
Figure 5. To increase the scope of environmental conditions in our laboratory testing, we adapted an evaluation technique from human biometrics: semi-synthetic models [10, 23]. Semi-synthetic models are derived from individual source object data, but also capture important conditions such as weather and atmospheric effects that impact long-range data for animal population studies. We re-imaged our data set outdoors at 100M on a sunny day, with a setup that consisted of a 4000 ANSI lumens projector with a resolution of 1024×768 projecting into a custom built screen, imaged by a Canon EOS 7D Camera with a Canon $2\times$ adapter and a Sigma 800mm F5.6 lens. This is an example image of our setup; for evaluation, we crop the image and pass it to the classifiers.

4.2 Experimental Results

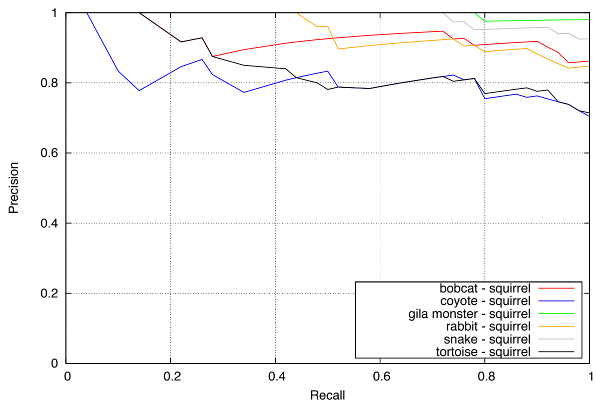
One advantage we have in this work is that we do not need to achieve perfect accuracy on a per frame basis, which would be ideal in other applications like security. In an animal population study, as long as we can track an animal across a contiguous sequence of captured frames, it can be counted correctly if it is recognized the majority of the time. The various precision-recall curves we show below for our proposed animal classification approach reflect acceptable accuracies for the recognition task.

To gain a sense of the impact of potential confuser animals on accuracy, we performed a series of classification experiments on a per animal basis. To begin, we trained 1-class SVMs with RBF kernels for both squirrels and tortoises and then tested each using features derived from all of the animals in our pool. Each curve in Figs. 6 – 8 reflects accuracies at various precision-recall points for 50 negative samples from the confuser animal and 50 positive samples from the animal of interest. Curves were plotted by varying a threshold over the actual decision scores from each SVM. For comparison, we ran a baseline classification experiment using a reference implementation of SIFT [13]. Typically, SIFT features from different images are compared using a distance metric. To create a decision-based classifier, we select the best distance score out of all of the comparisons to the same training set images used for the SVMs. Curves are again plotted by varying a threshold, this time over the distance scores. In all cases, our novel LBP-like features + 1-Class SVM algorithm significantly outperforms SIFT. In the long distance semi-synthetic data test shown in Fig. 7, SIFT simply fails.

We also examined the feasibility of individual species



(a) Baseline SIFT

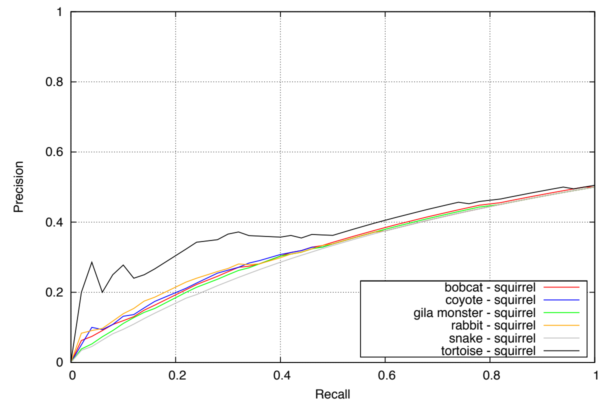


(b) Proposed Approach: LBP-like Features + 1-Class SVM

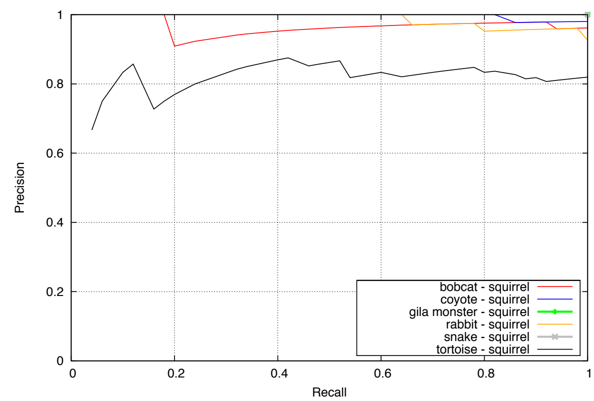
Figure 6. Precision-Recall curves depicting the ability of our squirrel classifiers to discriminate between potential confuser species and the animal of interest. Curves towards the upper right of each plot indicate higher levels of accuracy. (a) shows baseline results for SIFT classification using a distance metric and sliding threshold. (b) shows the results for our proposed approach with an LBP-like operator with a 5×5 neighborhood. Our approach produces significantly higher levels of accuracy in all cases.

recognition for ground squirrels. Using a test set of 100 images for each of the three squirrel species shown in Fig. 4, we ran two series of experiments. In the first experiment, we used the basic V1-like algorithm[2] and an experimental protocol that chooses a random set of testing and training data (45 training, 30 testing) from 100 images of each squirrel. For testing, each candidate image is submitted to a 3-class SVM; the classifier that produces the best score indicates the determination of species. We performed 10-fold cross validation to achieve an average **accuracy of 76.44%**. In our second experiment, we used a modified version of the V1-like algorithm (provided by the same software) that incorporates additional histogram features along with the final outputs to increase accuracy. This experiment followed the same training/testing procedure as the original, achieving an average **accuracy of 77.89%**. Thus, we can conclude that

[2] http://pinto.scripts.mit.edu/Code



(a) Baseline SIFT



(b) Proposed Approach: LBP-like Features + 1-Class SVM

Figure 7. Precision-Recall curves depicting the ability of our squirrel classifiers to discriminate between potential confuser species and the animal of interest for our semi-synthetic data set. While the baseline SIFT approach essentially fails, our approach, with an LBP-like operator with a 5×5 neighborhood, produces high accuracy in all cases (note that for classification involving Gila monsters and snakes, a precision and recall of 1 was always achieved).

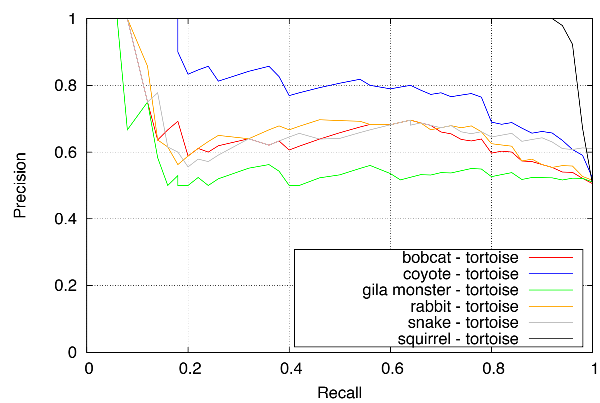


Figure 8. Precision-Recall curves depicting the ability of our tortoise classifiers to discriminate between potential confuser species and the animal of interest. Since tortoises are larger than squirrels, we make use of a larger 9×9 LBP-like operator window here.

while individual species identification can be accomplished with a simple approach, more work must be done to reach higher levels of accuracy.

5 Conclusion

This paper presented a snapshot of our interdisciplinary team's ongoing work in the Mojave Desert to build new vision tools for field biologists for the study of the currently threatened Desert Tortoise and Mohave Ground Squirrel. Besides the obvious application to a real-world problem, another goal of this work is to highlight the outreach potential of computer vision to other areas of science. Even though our algorithmic approach is intentionally simple so it runs on limited hardware, it still produces results that are meaningful enough for field biologists studying animals in the Mojave. Many existing vision solutions can be used immediately for other sciences even though accuracies are lower than what is expected for other application areas.

We emphasize that this project is still underway. The focus of the next phase will be creating a good tracking and detection system and incorporating these algorithms into a mobile smartphone device that will be used in the field for a variety of recognition tasks. While this paper focused on recognition, object detection is still very much of interest here. We are working towards a lightweight object detection approach beyond just relying upon the PoIs from the SIFT method. And finally, as scientists, we are excited about the potential of this approach to generate useful biological studies that will assist in the ongoing conservation efforts for the Desert Tortoise and Mohave Ground Squirrel.

6 Acknowledgments

The research reported in this document was supported by Air Force SBIR FA9302-12-C-0002.

References

[1] S. Branson, C. Wah, F. Schroff, B. Babenko, P. Welinder, P. Perona, and S. Belongie. Visual Recognition with Humans in the Loop. In *ECCV*, Sept. 2010.

[2] T. Clutton-Brock and B. C. Sheldon. Individuals and Populations: the Role of Long-term, Individual-based Studies of Animals in Ecology and Evolutionary Biology. *Trends in Ecology and Evolution*, 25(10):562 – 573, Sept. 2010.

[3] C. Cohen, D. Haanpaa, S. Rowe, and J. Zott. Vision Algorithms for Automated Census of Animals. In *AIPR*, Oct. 2011.

[4] P. Dickinson, R. Freeman, S. Patrick, and S. Lawson. Autonomous Monitoring of Cliff Nesting Seabirds Using Computer Vision. In *Int. Wkshp. on Distributed Sensing and Collective Intelligence in Biodiversity Monitoring*, Dec. 2008.

[5] P. Dickson, C. Qing, S. Lawson, and R. Freeman. Automated Visual Monitoring of Nesting Seabirds. In *Workshop on Visual Observation and Analysis of Animal and Insect Behavior*, Aug. 2010.

[6] R. P. Duin and E. Pekalska. Open Issues in Pattern Recognition. In M. Kurzynski, E. Puchala, M. Wozniak, and A. Zolnierek, editors, *Computer Recognition Systemsn*, pp 27–42. Springer, 2005.

[7] Environmental Protection Authority. Terrestrial Biological Surveys as an Element of Biodiversity Protection. Technical report, Environmental Protection Authority, Perth, 2002.

[8] M. Heikkilä, M. Pietikäinen, and C. Schmid. Description of interest regions with local binary patterns. *Pattern Recognition*, 42(3):425 – 436, 2009.

[9] R. Inman, K. Nussear, and C. Tracy. Detecting Trends in Desert Tortoise Population Growth: Elusive Behavior Inflates Variance in Estimates of Population Density. *Endangered Species Research*, 2009.

[10] V. Iyer, S. Kirkbride, B. Parks, W. Scheirer, and T. Boult. A Taxonomy of Face Models for System Evaluation. In *IEEE AMFG*, June 2010.

[11] A. Kembhavi, R. Farrell, Y. Luo, D. Jacobs, R. Duraiswami, and L. Davis. Tracking Down Under: Following the Satin Bowerbird. In *IEEE WACV*, Jan. 2008.

[12] P. Leitner. Current Status of the Mohave Ground Squirrel. *Trans. Western Section of the Wildlife Society*, 44, 2008.

[13] D. Lowe. Distinctive Image Features From Scale-Invariant Keypoints. *IJCV*, 60(2):91–110, 2004.

[14] N. Pinto, D. Cox, and J. DiCarlo. Why is Real-World Visual Object Recognition Hard? *PLoS Computational Bio.*, 4(1), 2008.

[15] N. Pinto, J. DiCarlo, and D. Cox. How Far Can You Get With a Modern Face Recognition Test Using Only Simple Features? In *IEEE CVPR*, June 2009.

[16] C. Qing, P. Dickinson, S. Lawson, and R. Freeman. Automatic Nesting Seabird Detection Based on Boosted HOG-LBP Descriptors. In *IEEE ICIP*, 2011.

[17] S. Sanchez. Desert Tortoises Get 'Head Start' to Survival. Technical report, Edwards Air Force Base, 2008.

[18] A. Sapkota, B. Parks, W. Scheirer, and T. Boult. FACE-GRAB: Face Recognition with General Region Assigned to Binary Operator. In *IEEE Computer Society Workshop on Biometrics*, June 2010.

[19] B. Schölkopf, J. Platt, J. Shawe-Taylor, A. Smola, and R. Williamson. Estimating the Support of a High-dimensional Distribution. Technical report, Microsoft Research, 1999.

[20] M. Shah, O. Javed, and K. Shafique. Automated Visual Surveillance in Realistic Scenarios. *IEEE Multimedia*, 14(1), 2007.

[21] G. R. Stewart. Petition to list the Mohave Ground Squirrel (Spermophilus mohavensis) as a Federally Endangered Species. Technical report, Defenders of Wildlife, 2005.

[22] G. Thompson and S. Thompson. General Terrestrial Fauna Surveys Protocol. Technical report, Edith Cowan University, 2008.

[23] V. Iyer and W.J. Scheirer and T.E. Boult. Face System Evaluation Toolkit: Recognition is Harder than it Seems. In *IEEE BTAS*, Sept. 2010.

[24] J. Westervelt, B. Hannon, S. Levi, and S. Harper. A Dynamic Simulation Model of the Desert Tortoise (*Gopherus agassizii*) Habitat in the Central Mojave Desert. Technical Report 97/102, US Army Corps of Engineers, 1997.