Classifying Wildlife Camera Trap Images for Conservation International’s Tropical Ecology Assessment and Monitoring Network

Anthony F. McGovern

Harvard University Extension School  
https://canvas.harvard.edu/courses/18127  
(202) 594-6194

tonmcg@gmail.com

**ABSTRACT**

Monitoring wildlife species in a non-intrusive and cost-effective manner has been made possible through camera trapping methods, which employ cameras set off by sensors to take bursts of images of animals in their natural habitat. Photo-identification of wildlife species yields valuable information about certain species, including population size, behavior, migration patterns, and human encroachment, particularly of endangered species. Camera trapping technologies produce a high volume of data in the order of millions of images that must be analyzed by a human expert. Moreover, given the need to establish an image proprietary mechanism that prevents opportunistic poachers and illegal traders of wildlife from further encroaching on endangered species, correctly classifying wildlife image data remains a challenging problem for wildlife researchers and enthusiasts.

In this paper, I describe a method to classify the *Loxodonta* genuswithin the *Elephantidae* family of African elephants from camera trap image data using a simple classification pipeline: manual segmentation of images; feature extraction using SIFT; and classification using the Bag of Visual Words and linear SVM models. These methods are fast, reliable, and have broad application potential to classify other wildlife species from camera trap image data.

**Keywords**

Animal species identification, image segmentation, feature extraction, SIFT, camera trap, Tropical Ecology Assessment and Monitoring Network, Conservation International

# INTRODUCTION

The Tropical Ecology Assessment and Monitoring (TEAM) Network managed by Conservation International (CI) monitors long-term trends in biodiversity, land cover change, climate, and ecosystem services in tropical forests. The network has built a system of camera trap stations within sixteen tropical forest sites across Africa, Asia and Latin America that feed image data to experts at the regional hub field sites, which then make their way into an image repository at CI for further analysis and research. The scientists use standard measures on this image data to quantify how animals respond to pressures like climate change impacts and habitat degradation.

Camera trap technologies have provided a non-intrusive and cost-effective method to view and collect huge volumes of image data, opening the way for more avenues to monitor specific animal species and answer questions about them including biometric features, behaviors, their resource selection, population size, migration patterns, as well as information about their environment [10]. These stationary cameras reside in small boxes (traps) affixed to a natural structure, such as rocks or trees and are set off by a motion-triggered sensor. When a sensor detects movement by an animal coming in to view, the cameras record a series of image sequences by deploying a burst of up to 10 individual shots within a matter of seconds. These images are later downloaded by experts in the field for manual processing and classification.

Humans can easily track and classify wildlife species from a series of individual shots of the animal walking in and out of the frame. Even with distorted images that are rotated or scaled, humans see the object recognition task as a trivial exercise [6]. In other words, humans understand that context matters: bursts of individual shots are understood as a sequence of related images containing the identified wildlife species either in part (just an elephant leg or trunk, for example) or whole. Yet this is one of the hardest challenges for computer vision systems today.

The TEAM Network relies on human experts to identify and classify species in each individual image. While this task is trivial for a human, it is still extraordinarily time-consuming, expensive, and error-prone. Furthermore, given that each image provides a rich array of metadata about the animal in the shot – including the geolocation of certain endangered species – the TEAM Network is precluded from engaging “citizen scientists” to help collaborate in the identification and classification tasks through collaborative, open-source methods [11]. Put another way, each image contains valuable information to would-be poachers and illegal traders that would further endanger the wildlife species in the image.

More to the point: an opportunity currently exists for computer vision tools to solve this complex problem though a mechanism that both removes humans from the task of manually classifying wildlife species from images while still ensuring sensitive image data containing endangered species remain hidden from poachers and illegal animal traders.

The current wildlife image classification task consists of three distinct stages: segmentation, feature extraction, and classification [7]. Segmentation attempts to remove the animal foreground from the dense background as an input into a computer vision tool for an object recognition exercise. The goal of segmentation is to feed a classifier an image with the animal of interest as the focus, discarding as much as possible all background features. The TEAM Network images of the *Loxodonta* genus are surrounded by plants, trees, and thickets of the Nouabalé-Ndoki National Park in the Republic of Congo, where the camera traps are stationed. Feature extraction attempts to extract properties of the data that describe the images themselves for classification. The last stage in the pipeline attempts to find a suitable classifier to apply to the segmented and feature extracted images that yields accurate results.

In this paper, I describe a method to classify the *Loxodonta* genus of African elephants from camera trap image data using a simple classification pipeline: foreground segmentation using manual cropping; feature extraction using SIFT; and classification using the Bag of Visual Words and linear SVM models. The rest of this paper is organized as follows: Section 2 mentions related work in this field; Section 3 provides an analysis and describes the limitations of the image data; Section 4 describes the segmentation, feature extraction, and classification methods in the pipeline; Section 5 presents the results and a discussion of the results; and finally, Section 6 presents conclusions and future work.

# RELATED WORK

Automatic classification of animal species is an ongoing problem that remains unsolved. Bolger et.al [1] worked to identify wildlife species with distinctive coat patterns. They used Scale Invariant Feature Transformation (SIFT) key points extraction and matching for the feature extraction portion of the pipeline. Yu et.al [8] manually cropped selected camera trap images in the segmentation portion of their work and combined SIFT and cLBP (cell-based Local Binary Pattern) descriptors as feature extractors. They used a linear SVM classifier to classify the images. Parikh et al. [8] used template matching as their model for wildlife species classification. The pipeline described in this paper closely follows the methods of Yu et.al.

More recently, advances in the deep learning literature for computer vision have been applied as the classification step in the pipeline. Chen et.al [2] used a deep convolutional neural network (DCNN) to classify wildlife species based on images segmented using an Ensemble Video Object Cut method. They claimed to achieve a 38% accuracy using this pipeline. Gomez, et.al. [5] used a very deep CNN for image classification using architectures pre-trained on the ImageNet dataset. They also manually cropped images in the segmentation portion of the pipeline like Yu et.al. They claimed to have achieved an 88.9% accuracy using this pipeline.

# IMAGE DATA ANALYSIS & LIMITATIONS

Much of the TEAM Network’s repository of wildlife image data is available to the public through a data query and download UI on their website. As well as the *Loxodonta* genus of images, the TEAM Network provides millions of images for download across dozens of terrestrial vertebrate in the *Mammalia* class, including lemurs, tigers, and Asian elephants. A report on all TEAM Network terrestrial vertebrate image data shows that classes are highly unbalanced within order, family, genus, and species of animals. Given both the random nature of wildlife species passing in front of a camera trap and the non-invasive method for capturing these images, it is impossible for field experts to provide a balanced class of images for each wildlife species. Some wildlife species only have a very small number of images, while others – like the *Loxodonta* genus – have many.

For example, the table below shows an unbalanced count of classes for some of the more popular wildlife species in the TEAM Network’s repository:

Table . Unbalanced Count of Classes in Wildlife Images

|  |  |  |  |
| --- | --- | --- | --- |
| **Species** | **Common name** | **No. photos** | **No. sites** |
| nebulosa | Clouded Leopard | 6 | 1 |
| planiceps | Flat-headed Cat | 21 | 1 |
| marmorata | Marbled Cat | 55 | 3 |
| serval | Serval | 178 | 3 |
| tigris | Tiger | 205 | 1 |
| bengalensis | Leopard Cat | 642 | 3 |
| onca | Jaguar | 3252 | 7 |
| pardalis | Ocelot | 16473 | 8 |

Moreover, many wildlife species images show the same species from several different field sites, which presents another set of problems that includes variability in camera trapping technologies, dissimilar habitats showing different background patches of trees, rocks, and thickets, as well as dissimilar lighting based on the location of the image. The table below provides metadata samples for the *Elephantidae* images showing a high number of camera technology attributes that can vary across field sites:

Table . Camera Metadata for *Elephantidae* Images

|  |  |
| --- | --- |
| **Column name** | **Sample values** |
| Genus | Loxodonta |
| Species | Africana |
| Camera Serial Number | RM13AB01002923 |
| Memory Card Serial Number | BE090412797B |
| Camera Start Date and Time | 2/20/10 13:21 |
| Camera End Date and Time | 3/27/10 12:53 |
| Camera Make | 'Reconyx' |
| Camera Model | 'RM45 RAPIDFIRE' |
| Flash | 25 |
| Exposure Time | 448/13440 (0.033) |
| Sequence Info | RM13AB01002923 |
| Triplet Number | M 1/3 |

Human classification of species poses the biggest problem as field experts are manually isolating, identifying, and classifying thousands of images. In each image, experts must determine if there is an animal in the image and to which order, family, genus, and species the mammal belongs. This image classification process can vary in output quality due to environmental conditions in the image, limits of knowledge, and fatigue. This means that sometimes the problem of species classification is difficult and ambiguous even for experts. Frequently, misclassification occurs.

The table below shows environmental variability in the image metadata, which can lead to misclassification of the image:

Table . Environmental Metadata for *Elephantidae* Images

|  |  |
| --- | --- |
| **Column name** | **Sample values** |
| Latitude | 2.6976397 |
| Longitude | 16.6181041 |
| Genus | Loxodonta |
| Species | Africana |
| Number of Animals | 1 |
| Person Identifying the Photo | Patrick Boundja |
| Moon Phase | 2 |
| Temperature | 23 |

Finally, the TEAM Network repository contains a class of unlabeled or unidentified images that cannot be classified for a variety of reasons, including image distortion, bad lighting, inoperable technology, or the existence of more than one species of wildlife animal within the image. These images present a challenging problem in the computer vision literature and for the purposes of this project, these images are discarded from the repository of images used to train a classification model.

Given the data quality and limitations across all images described above, I chose the *Loxodonta* genus of African elephants as the sole wildlife animal to classify. The TEAM Network repository includes 31,485 images of this genus alone, all of which were captured by a camera trap in only one field site: Nouabalé-Ndoki National Park in the Republic of Congo. This large number of images provides a good training set for testing this classification pipeline. The focus on one genus from one field site limits the variability in environmental conditions. Furthermore, limited variability in camera technology ensures data quality is more consistent across all images. Taken together, the selection of these images should ensure a higher classification accuracy compared to attempting to classify multiple species in images from across multiple field sites.

Most images of the *Loxodonta genus* fell between 500kb and 1Mb in file size and were highly dimensional. The figure below shows a representative image from the TEAM Network repository; note the image has a file size of 749kb and features a 2048 × 1536 dimensionality across 3 channels:



Figure . Sample *Loxodonta* Image from the Repository

Note the animal is surrounded by plants, trees, and thickets of the Nouabalé-Ndoki National Park, which makes the classification task more challenging.

# IMAGE CLASSIFICATION PIPELINE

This project relied almost exclusively on a the OpenCV library for image pipeline tasks. OpenCV offers many methods for image recognition, particularly its implementation of SIFT. While not as robust as the original algorithm, OpenCV’s implementation is extremely fast and reliable.

For the segmentation portion of the pipeline, I manually chose and cropped 287 images of the *Loxodonta* genus for processing. Following the work of Yu, et.al [8], manual cropping provides a reliable method to strip the background from each image, while retaining the object of interest in the foreground for feature extraction. The figure below is a representative example of the cropping done on each image:  
  


Figure 2. Sample Cropped *Loxodonta* Image

Cropping images manually reduced the file size to about 250kb to 300kb per image, allowing the feature extraction portion of the pipeline to execute much faster.

While strides were made using SIFT for feature extraction, implementations of histogram and template matching did not yield positive results and were therefore abandoned. Using OpenCV’s implementation of SIFT, particularly the key points and SIFT descriptors methods, provided the feature extraction pipeline a robust and reliable set of image features by which the classifiers could do their work. The image below provides an example of the key points and features drawn from each image using SIFT:  
  


Figure 3. Key Points and SIFT Descriptors Highlighted

As the figure shows, the *Elephantidae* image presents key features around the eyes, trunk and ears for easier detection by classifiers. The figure below shows how SIFT connects these key features through image descriptors, highlighted by the bright lines:  
 

Figure 4. Descriptors using OpenCV’s Brute-Force matcher

Brute-Force matching takes the descriptor of one feature in the first image and matches it with other features in second image using some distance calculation. In the image above, a Norm L2 distance measure is used, which is the Euclidean distance between two points (square root of sum of squares).

Lastly, for the classification step of the pipeline, I combined the Bag of Visual Words (BoVW) image mining technique with a linear classifier to achieve results. To achieve more robust results, images from the CalTech 101 Objects were introduced into the pipeline, which contained images of other elephants. This paper trained and tested the images on a fixed number of pictures and repeated the classification with different random selections of images.

The BoVW is the analogous technique to the Bag of Words text mining framework. BoVW quantizes the SIFT descriptor space into a number of examples and assigns each descriptor in the image to one of those examples. Examples are determined by analyzing a training set of images that are considered like visual words. A visual vocabulary is built (visual codebook) with the set of all visual words clustered together to create a representative of the larger image. The clustering method employed in this paper is k-means.

Once the BoVW model is implemented, a linear SVM using scikit-learn’s GridSearch and cross-validation provided the following results:

Best parameters: {'gamma': 0.1, 'C': 0.1, 'kernel': 'linear'}

Best cross-validation score: 0.91

Setting the number of clusters *a priori* to 10 for the k-means technique yielded a 98% accuracy on the training set and a 95% accuracy on the test set. For exploratory points, AdaBoost classification was introduced and k-means clustering was further tuned to 20, 40, and 80 clusters. Results using the different variations of clusters hovered around 98% for the training set and 98% for the test set. In other words, AdaBoost did not seem to help the accuracy of the image classification.

Finally, visualizing the results using a confusion matrix confirmed the accuracy of the results:  
  


Figure 5. Confusion Matrix for Classification Results

# RESULTS AND DISCUSSION

In this paper, I described a method to classify the *Loxodonta genus* of African elephants from camera trap image data using a simple classification pipeline: foreground segmentation using manual cropping; feature extraction using SIFT; and classification using the Bag of Visual Words and linear SVM models. This classification pipeline provided a fast and reliable set of results. Moreover, this pipeline could have possible applications to other wildlife species.

Results of the classification pipeline are shown below:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Precision** | **Recall** | **F1-Score** | **Support** |
| **Loxodonta** | 0.98 | 0.98 | 0.98 | 43 |
| **Not Loxodonta** | 0.98 | 0.98 | 0.98 | 43 |
| **Avg / Total** | 0.98 | 0.98 | 0.98 | 86 |

# CONCLUSIONS AND FUTURE WORK

The short-term goals of this project are to help the TEAM Network verify the classification of *Loxodonta* genus of African elephants in their repository. Given the volume of image data in their repository, as well as the proprietary nature of their endangered species images, much of the classification tasks are left to experts in the field sites who must pore over millions of individual images and manually classify the order, family, genus, and species of the wildlife animal in the image. A long-term goal of this project is to automate the task of classifying TEAM Network images from start to finish.

Future work to help with the short-term goal should include increasing training set image data with graph cut images instead of manually cut images. Having more training data with the focus of the object clearly present should help with overall accuracy. Using graph cut will obviate the need for a manual intervention in the classification pipeline. Given the choice of a linear SVM as the classifier, training data should increase memory load and computational time only trivially.

Future work should also attempt to implement more feature extraction methods like Speeded up robust features (SURF) and LBP texture descriptor to help with the second part of the classification pipeline. Once accuracy reaches a pre-defined threshold, more should be done to incrementally include other species from the same family: for example, the classification pipeline described in this paper should be extended to include the *Elephas* genus of Asian elephants, which also belongs to the *Elephantidae* family.

Finally, future work should explore using deep CNNs for the feature extraction and classification parts of the pipeline, particularly those models already trained on ImageNet or other publicly available wildlife image data. Recent literature describes multilayer architectural approaches that achieve very high accuracy. Currently, the computational costs to run a CNN with the architectures described are much too prohibitive for a class project.

# ACKNOWLEDGMENTS

All data in this publication were provided by the Tropical Ecology Assessment and Monitoring (TEAM) Network, a collaboration between Conservation International, the Smithsonian Institution, and the Wildlife Conservation Society, and partially funded by these institutions, the Gordon and Betty Moore Foundation, and other donors.

My thanks to Eileen Larney, TEAM Technical & Communications Manager and Ph.D. in Physical and Biological Anthropology, for providing image data and the necessary context of the animal species in the images themselves. My appreciation also goes to Ashleigh McGovern, Senior Director at the Center for Oceans at CI, and to whom this project is dedicated. Ashleigh introduced me to the brilliant work being done at TEAM, facilitated conversations with Jorge Ahumada, Executive Director of the TEAM Network, and provided the needed proofing and editing of this report.

My thanks also go to Peter Henstock of the Harvard University Extension School and his hard-working team of teaching assistants: Dave Downey, Xintao Qiu, Rashmi Banthia, and Claudio Rosenberg. Peter’s class is rigorous and intellectually rewarding. Combined with Peter’s excellent class lectures and notes, the TA’s sections, notebooks, and Piazza answers provided a great foundation from which to tackle this project.

For more information on Conservation International:  
<http://www.conservation.org/>

For more information on the TEAM network:  
<http://www.teamnetwork.org/>

# DATA AND CODE DIRECTORY

The 31,485 images in the *Loxodonta* dataset were downloaded as a 7.33 GB compressed zip file from the TEAM Network’s “Data Query and Download” website using Terrestrial Vertebrate for the Data Products parameter, Nouabalé-Ndoki for the TEAM Sites parameter, and Loxodonta for the Advanced Options > Taxonomic Search > Terrestrial Vertebrate > Genus parameter:  
<http://www.teamnetwork.org/gridsphere/gridsphere?cid=search>.

The code for the image classification pipeline was written in Python and can be viewed online as a [Jupyter iPython notebook](http://nbviewer.jupyter.org/github/tonmcg/Classify-Terrestrial-Vertebrate/blob/master/Classifying%20Wildlife%20Camera%20Trap%20Images%20for%20Conservation%20International%E2%80%99s%20Tropical%20Ecology%20Assessment%20and%20Monitoring%20Network.ipynb).

Python version 3.5 was used running in a Jupyter notebook environment on a 64-bit MacBook Pro with an Intel i386 processor using four cores. The following libraries were also used for the project:

Table . Python Libraries and Version Numbers

| **Library** | **Version** |
| --- | --- |
| pandas | 0.19.1 |
| numpy | 1.11.2 |
| cv2 | 3.1.0 |
| sklearn | 0.18.1 |
| matplotlib | 1.5.3 |

# REFERENCES

1. Bolger, D. T., Morrison, T. A., Vance, B., Lee, D., & Farid, H. A computer‐assisted system for photographic mark–recapture analysis. *Methods in Ecology and Evolution*, *3*(5), 813-822.
2. Chen, G., Han, T. X., He, Z., Kays, R., & Forrester, T. Deep convolutional neural network based species recognition for wild animal monitoring. In *2014 IEEE International Conference on Image Processing (ICIP)* (pp. 858-862). IEEE.
3. Figueroa, K., Camarena-Ibarrola, A., García, J., & Villela, H. T. Fast Automatic Detection of Wildlife in Images from Trap Cameras. In *Iberoamerican Congress on Pattern Recognition* (pp. 940-947). Springer International Publishing.
4. Forrester, T., McShea, W. J., Keys, R. W., Costello, R., Baker, M., & Parsons, A. eMammal–citizen science camera trapping as a solution for broad-scale, long-term monitoring of wildlife populations. *Sustainable Pathways: learning from the past and shaping the future*.
5. Gomez, A., & Salazar, A. Towards Automatic Wild Animal Monitoring: Identification of Animal Species in Camera-trap Images using Very Deep Convolutional Neural Networks. *arXiv preprint arXiv:1603.06169*.
6. Kamencay, P., Benco, M., Matuska, S., Hudec, R., & Radilova, M. A Novel System for Non-Invasive Method of Animal Tracking and Classification in Designated Area Using Intelligent Camera System. *Radioengineering*.
7. Kumar, Y. S., Manohar, N., & Chethan, H. K. Animal Classification System: A Block Based Approach. *Procedia Computer Science*, *45*, 336-343.
8. Parikh, M., & Patel, M. Animal detection using template matching algorithm. *MBICT, Gujarat Technological University, India*.
9. Yu, X., Wang, J., Kays, R., Jansen, P. A., Wang, T., & Huang, T. Automated identification of animal species in camera trap images. *EURASIP Journal on Image and Video Processing*, *2013*(1), 1.
10. Zhang, Z., He, Z., Cao, G., & Cao, W. Animal Detection from Highly Cluttered Natural Scenes Using Spatiotemporal Object Region Proposals and Patch Verification. *IEEE Transactions on Multimedia*, *18*(10), 2079-2092.