



**UNIVERSIDAD
DE ANTIOQUIA**

1 8 0 3

Recognition of Mammal Genera on Camera-Trap Images using Multi-Layer Robust Principal Component Analysis and Mixture Neural Networks

Student: Jhony Heriberto Giraldo Zuluaga

Advisor: Augusto Enrique Salazar Jiménez

Departamento de ingeniería Electrónica

Facultad de Ingeniería

Universidad de Antioquia

Content

- Motivation
- Problem
- State of the Art
- Objectives
- Methods
- Experimental Framework
- Results
- Conclusions



Figure 1: Artiodactyla Cervidae Mazama
Source: Instituto Alexander von Humboldt

Motivation



Figure 2: Animals in the wild
Creative Commons Zero (CC0) license

Non-invasive animal genera recognition,
using computer vision and pattern
recognition algorithms.

Problem



Figure 3: Animal detection

Source: Instituto Alexander von Humboldt

Detection problem with intra class variation due to different poses. Context problems: illumination changes, dynamic background, shadows.

Detection of mammal genera from camera trap images, using computer vision and pattern recognition.

State of the Art Segmentation



Figure 4: Previous works
Source: [Zhang et al., 2016]

Bag of Words, Histogram of Oriented Gradients, and graph cut energy minimization [Zhang et al., 2015]. Iterative embedded graph cut, histogram of oriented gradient, and convolutional neural networks [Zhang et al., 2016].

State of the Art (2)

Classification

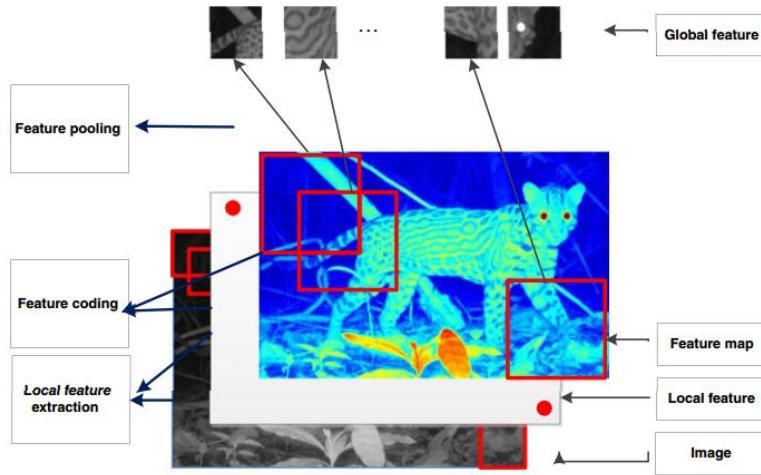


Figure 5: Previous works
Source: [Yu et al., 2013]

Dense Scale Invariant Feature Transform and cell-structured Local Binary Pattern with multi-class SVM to classify the features [Yu et al., 2013]. Convolutional Neural Networks (CNN) [Chen et al., 2014]. Deep CNN [Gomez et al., 2017].

Objectives

General Objective

- Classification of mammal genera on camera-trap images using Convolutional Neural Networks to help in conservation tasks.

Specific Objectives

- Design of experiments with segmentation algorithms to segment the animals in the database.
- Validation of the segmentation algorithm applied on camera-trap images.
- Design of experiments with Deep Learning architectures to separate mammal genera.
- Validation of the Deep Learning architectures applied on camera-trap images.

Methods

Multi-Layer RPCA

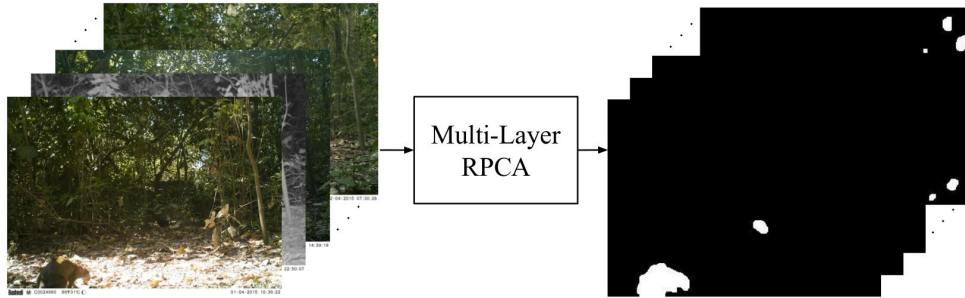


Figure 6: Segmentation algorithm, Multi-Layer Robust Principal Component Analysis (Multi-Layer RPCA)

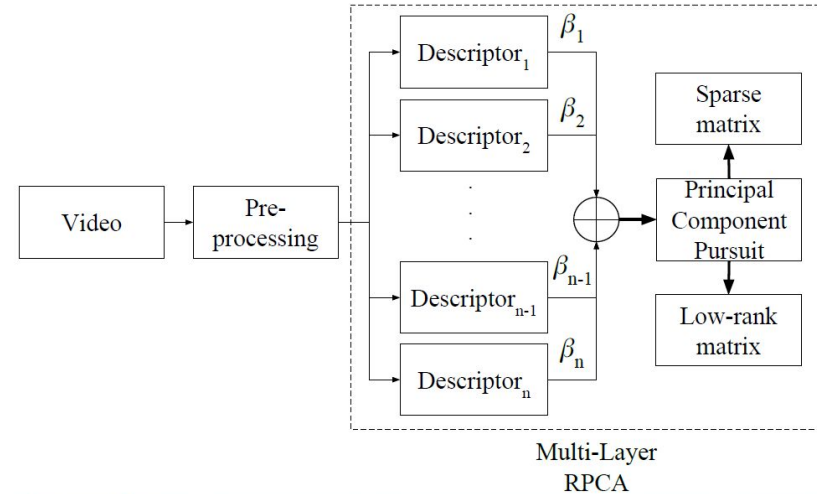


Figure 7: Multi-Layer RPCA method

Methods (3)

CNN

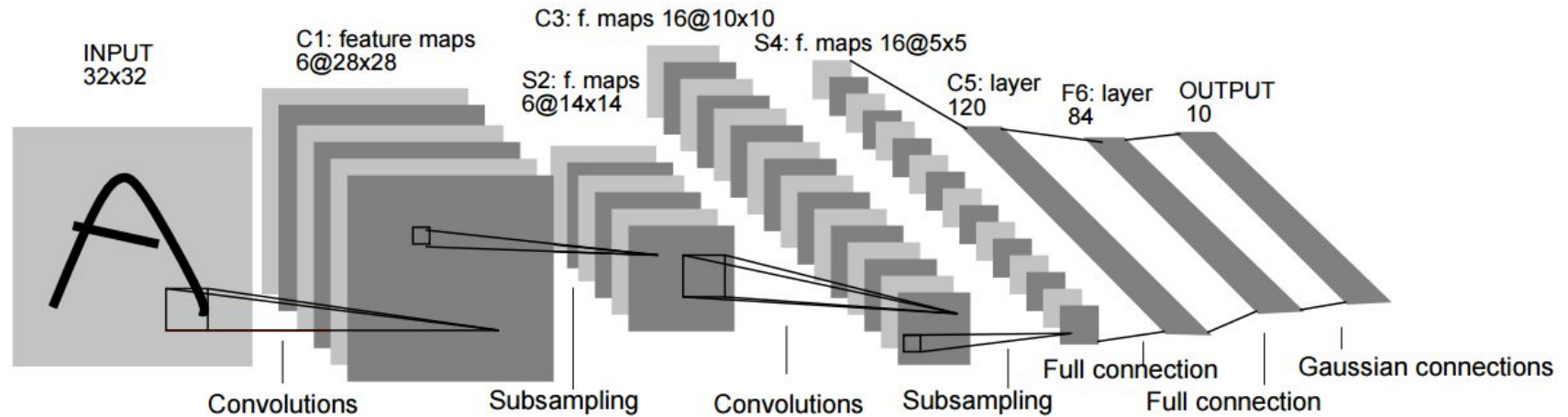
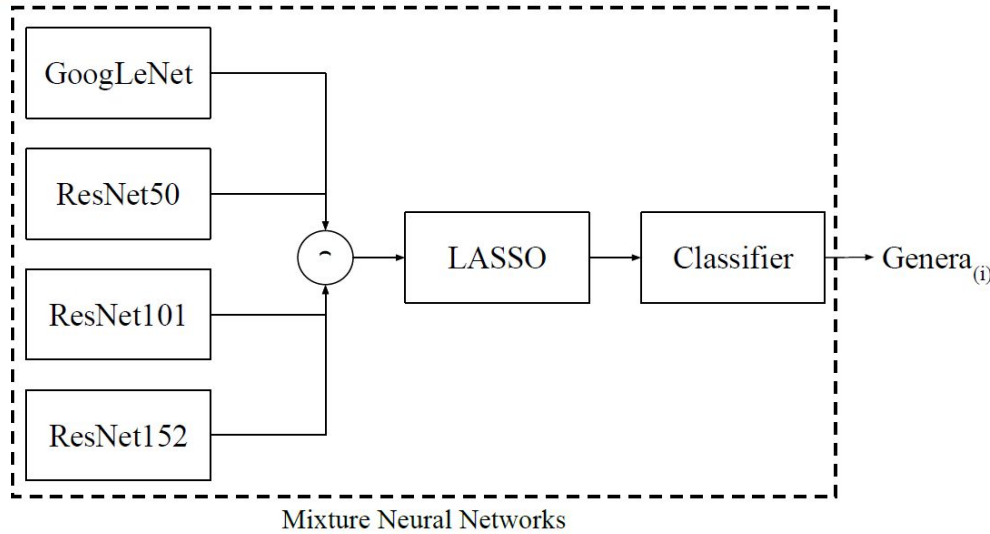


Figure 8: Convolutional Neural Network (LeNet) [LeCun et al., 1998]

Methods (4)

MixtureNet



CNN: GoogLeNet [Szegedy et al., 2015], ResNet50, ResNet101, y ResNet152 [He et al., 2016]. Least Absolute Shrinkage and Selection Operator (LASSO) [Tibshirani, 1996].

Figure 9: MixtureNet

Methods (5)

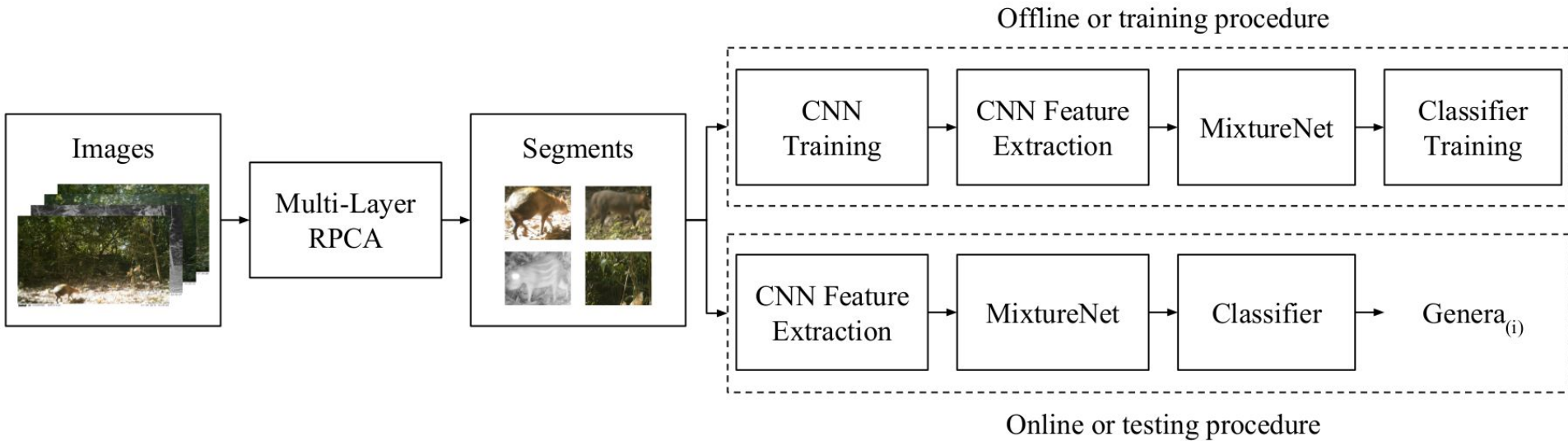


Figure 10: Pipeline of the mammal classification method presented in this work

Experimental Framework

Database Segmentation



Figure 11: Ground Truth images

The database consists of 1065 images from 30 color and infrared sequences. The length of each sequence data set varies from 9 to 72 images, depending on the animal activity in that sequence. We randomly chose the 60 sequences, ensuring animal activity in each sequence. Each image of the database has a Ground Truth.

Experimental Framework (2)

Database



Figure 12: *Dasyprocta* (4228, 3396), *Mazama* (441, 292), *Pecari* (712, 343), *Cerdocyon* (288, 167), *Leopardus* (284, 207), *Dasypus* (741, 389), *Didelphis* (688, 207), *Proechimys* (472, 229), *Cuniculus* (1150, 883), *Tamandua* (204, 125)

Experimental Framework (3)

Metrics

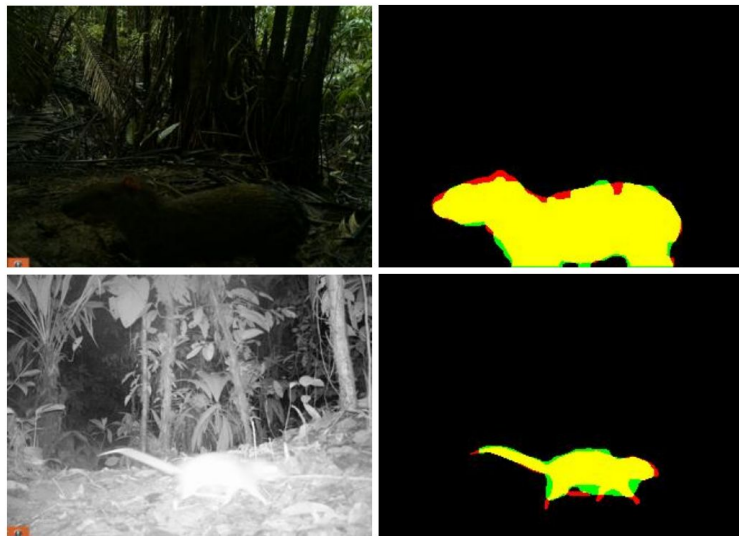


Figure 13: Extracting f-measure

Segmentation:

- F-measure per sequence.
- Average f-measure.

Classification:

- Accuracy.
- Kappa index (our classification method against a random classifier).

Experimental Framework (4)

Experiments

Segmentation:

- Multi-Layer RPCA with 2 descriptors and exhaustive search of the β parameter with 7 Principal Component Pursuit algorithms of the state-of-the-art.
 - Color, infrared, and color & infrared sequences.

Classification:

- Classification with expert segmentation (10 genera categories).
- Classification using automatic segmentation with intersection over union greater than 50% (8 genera categories and a false positive category).
 - Artificial Neural Network (ANN), Linear and Radial Basis Function Support Vector Machines (LSVM and RSVM): hyperparameters-optimization with test (optimistic results).
 - Direct evaluation of CNNs on test (non-optimistic results), without LASSO.

Results

Segmentation

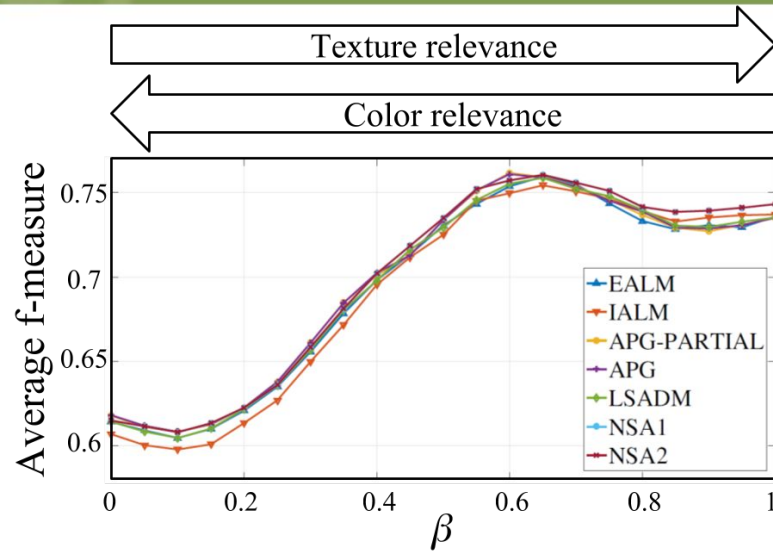


Figure 14: Results with color images

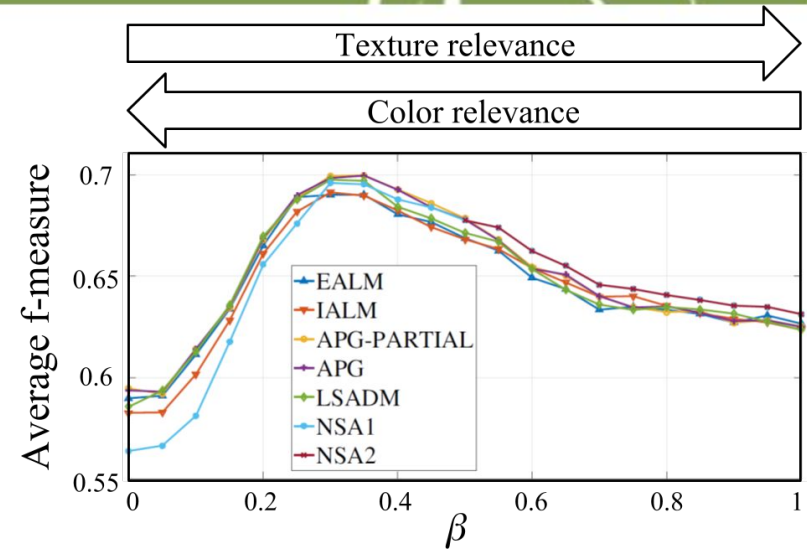


Figure 15: Results with infrared images

Results (2)

Segmentation

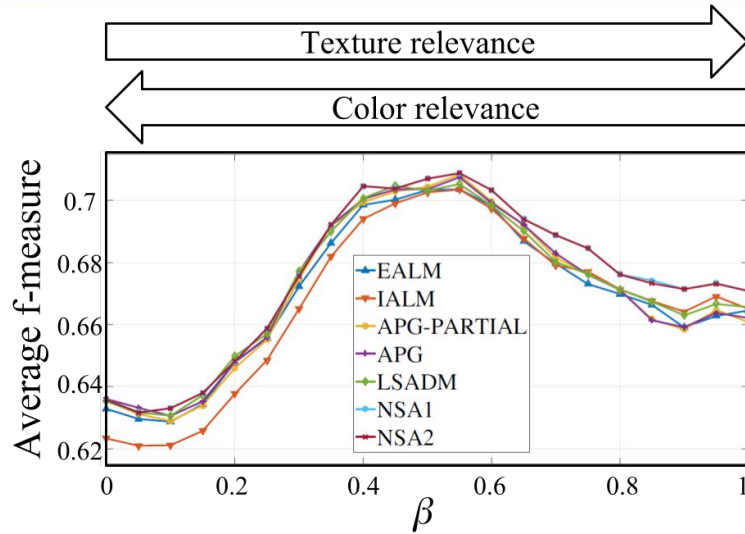


Figure 16: Results with color and infrared images

Experiment	Algorithm	β	Avg F-measure
Color images	APG-PARTIAL	0,6	0,7617
Infrared images	APG-PARTIAL	0,35	0,6997
Color & Infrared images	NSA2	0,55	0,7088

Table 1: Summary of the results

Results (3)

Classification (optimistic results)

CNN	Accuracy [%]			Accuracy LASSO [%]			Kappa
	ANN	LSVM	RSVM	ANN	LSVM	RSVM	
GoogLeNet	10	10	10	86,72	10	85,74	0,8525
ResNet50	88,85	1,15	88,36	86,39	10	85,41	0,8761
ResNet101	90,49	89,34	89,34	87,87	86,39	86,89	0,8944
ResNet152	90,66	89,51	90,16	89,67	88,52	88,85	0,8962
MixtureNet	10	10	10	89,67	10	87,87	0,8852
MixtureResNet	90,82	84,59	87,7	89,34	10	87,7	0,898

Table 2: Results using expert segmentation

CNN	Accuracy [%]			Accuracy LASSO [%]			Kappa
	ANN	LSVM	RSVM	ANN	LSVM	RSVM	
GoogLeNet	11,29	11,11	11,11	89,78	11,11	88,53	0,8851
ResNet50	91,94	19,89	91,22	89,43	11,11	89,07	0,9093
ResNet101	85,66	86,38	85,48	83,87	83,51	84,05	0,8468
ResNet152	92,83	92,83	92,65	91,58	91,4	90,32	0,9194
MixtureNet	11,29	11,11	11,11	92,83	11,11	91,04	0,9194
MixtureResNet	94,09	88,53	86,02	92,83	11,11	91,58	0,9335

Table 3: Results with automatic segmentation and false positive class

Results (4)

Classification (non-optimistic results)

Redes	Accuracy [%]	Kappa index	CNN time training (600,000 iterations)
GoogLeNet	84,59	0,8288	2 days, 26 minutes
ResNet50	88,2	0,8689	3 days, 15 hours, 52 minutes
ResNet101	88,85	0,8761	3 days, 13 hours, 33 minutes
ResNet152	88,85	0,8761	3 days, 11 hours, 54 minutes

Table 4: Results using expert segmentation, direct evaluation of the CNNs

Redes	Accuracy [%]	Kappa index	CNN time training (600,000 iterations)
GoogLeNet	88,17	0,8669	1 day, 23 hours, 28 minutes
ResNet50	90,32	0,8911	3 days, 14 hours, 55 minutes
ResNet101	85,48	0,8367	3 days, 12 hours, 52 minutes
ResNet152	91,94	90,93	3 days, 11 hours, 47 minutes

Table 5: Results with automatic segmentation and false positive class, direct evaluation of the CNNs

Conclusions

- Multi-Layer RPCA can handle the dynamic background of the camera-trap images.
- LASSO selection make the method more robust against bad features.
- LASSO selection is useful for combining CNNs (trained separately), when any CNN has harmful features. In another way, the MixtureNet inherits the harmful features of each CNN.
- Although the Multi-Layer RPCA generates thousands of false positives, the CNNs can handle this problem.
- The MixtureResNet exhibits the best kappa indexes in both classification experiments.
- The classification method fails in images with two or more challenges (e.g. overexposed and partially-occluded images).
- We tested images with $\text{IoU} > 0.5$ and $\text{IoU} = 0$ for animals and False Positives, respectively. Certainly, regions with $0 < \text{IoU} \leq 0.5$ are limitations of our method.

Angélica Díaz-Palido
Instituto de Investigación de Recursos Biológicos
Alexander von Humboldt, Bogotá D.C., Colombia.
adiaz@humboldt.org.co

database [11], where only 28.8% of the available images contain animals. As a result, wildlife scientists must analyze thousands of photographs, of which a high percentage does not show wildlife. This problem, albeit very well known in the camera-trap community, is far from being solved. Furthermore, biologists must classify tens of animal species or genera from thousands of images. An automatic segmentation and classification system might accelerate the professional work, allowing the biologists to concentrate in data analysis.

The pattern recognition community has approached the camera-trap recognition problem in two ways: segmentation (to detect animals in images and to segment them into species or genera identification (classification)). Yu et al. [3] proposed a segmentation method that is based on species from a camera-trap database taken in Patagonia. They used a wavelet-based Invariant Feature Transform and cell-convolutional Local Binary Pattern as feature extractions, and multi-

SVM to classify the features, they achieved 92% of accuracy. Note that Yu et al. assumed a perfect segmentation algorithm, doing a manual selection of which objects resulted in a perfect segmentation. This is not realistic. In this paper, we used a fully automatic segmentation algorithm, which also confronted species identification. They classified 30 animal species with a human-aided segmentation method and extracting 8 Local Feature Transforms for feeding a Random Forest Classifier. The proposed Convolutional Neural Network, they reached an 87.27% of classification accuracy.

Chen et al. [37] chose the segmentation and identification problems using 20 animal species from a camera-trap database. They used a deep convolutional neural network to extract the Ensemble Video Object Cut and then classified the segmented images comparing the performance of Bag of Visual Words with a CNN architecture with 3 convolutional and 3 max pooling layers. The proposed method achieved a 90.2% of accuracy, reducing false positives, the reached performance was lower (38.31%) of accuracy compared with manual segmentations.

Genere et al. confronted the species identification problem using 1000 images of 100 different species from the Snapshot Serengeti dataset [38], using very deep

Jhony-Heriberto Giraldo-Zubaga¹ · Augusto Salazar¹ · Alexander Gomez² · Annelisa Díaz-Palido²

Abstract The segmentation of animals from camera trap images is a difficult task. To illustrate, the new various challenges due to environmental conditions and hardware limitation in these images. We propose a Multi-Layer Robust Principal Component Analysis (Multi-Layer RPCA) approach for background subtraction. The proposed method is able to deal with the background clutter and low-level camera motion from a weighted sum of descriptors, using color and texture descriptors as one of study for camera-trap images segmentation. The segmentation algorithm uses histogram equalization or Gaussian filtering as pre-processing, and morphological filters with active contour as post-processing. The experiments obtained 100% accuracy for the Multi-Layer RPCA approach in the first search, using different amounts of camera-trap images. The database consists of camera-trap images from the Colombian forest taken by the Instituto de Investigación de Recursos Biológicos Alexander von Humboldt. The proposed method reached 76.17% and 60.87% accuracy for the first and second search, respectively. In our best knowledge, this paper is the first work proposing Multi-Layer RPCA and using it for camera-trap images segmentation.

Keywords: Camera-trap images - Multi-Layer Robust Principal Component Analysis - background subtraction - image segmentation

This work was supported by the Colombian National Fund for Science, Technology and Innovation, Francisco José de Caldas - COLCIENCIAS (Colombia), Project No. 11117431001.

Corresponding author: J. B. González-Zuluaga
E-mail: jgonzalez@uniag.org

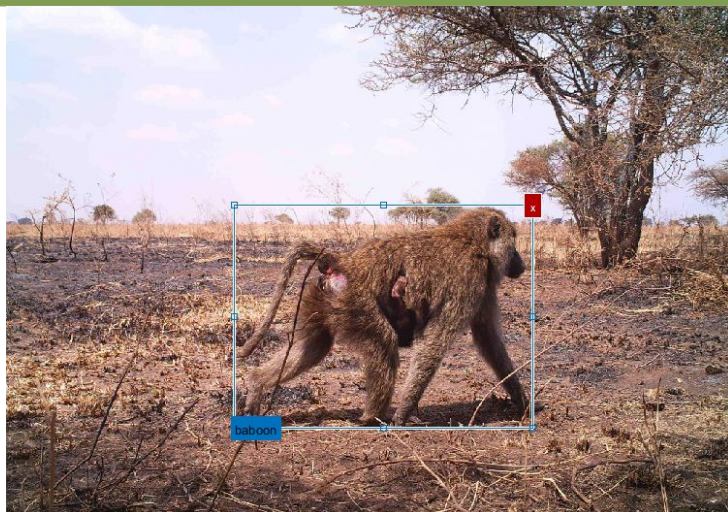
¹ Grupo de Investigación SISTEMAG, Facultad de Ingeniería, Universidad de Antioquia, Medellín, Colombia
² Instituto de Investigación de Recursos Biológicos Alexander von Humboldt, Bogotá D.C., Colombia

[illegible]

Journal: Visual Computer
Quartile: Q2, A2 (Colciencias)
Hindex: 53 (Scimago)



To do



DLCcovert.com

07-30-2010 12:32:31

Figure 17: Snapshot Serengeti
Source: [Swanson et al., 2015]

Thanks.
Questions?



References

- [Chen et al., 2014] Chen, G., Han, T. X., He, Z., Kays, R., & Forrester, T. (2014, October). Deep convolutional neural network based species recognition for wild animal monitoring. In Image Processing (ICIP), 2014 IEEE International Conference on (pp. 858-862). IEEE.
- [Gomez et al., 2017] Gomez, A., Salazar, A., & Vargas, F. (2017). Towards automatic wild animal monitoring: identification of animal species in camera-trap images using very deep convolutional neural networks. Ecological Informatics.
- [He et al., 2016] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 770-778).
- [LeCun et al., 1998] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.
- [Swanson et al., 2015] Swanson, A., Kosmala, M., Lintott, C., Simpson, R., Smith, A., & Packer, C. (2015). Snapshot Serengeti, high-frequency annotated camera trap images of 40 mammalian species in an African savanna. Scientific data, 2.
- [Szegedy et al., 2015] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1-9).

References

- [Tibshirani 1996] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological), 267-288.
- [Yu et al., 2013] Yu, X., Wang, J., Kays, R., Jansen, P. A., Wang, T., & Huang, T. (2013). Automated identification of animal species in camera trap images. EURASIP Journal on Image and Video Processing, 2013(1), 52.
- [Zhang et al., 2015] Zhang, Z., Han, T. X., & He, Z. (2015, September). Coupled ensemble graph cuts and object verification for animal segmentation from highly cluttered videos. In Image Processing (ICIP), 2015 IEEE International Conference on (pp. 2830-2834). IEEE.
- [Zhang et al., 2016] Zhang, Z., He, Z., Cao, G., & Cao, W. (2016). Animal detection from highly cluttered natural scenes using spatiotemporal object region proposals and patch verification. IEEE Transactions on Multimedia, 18(10), 2079-2092.

Attachments

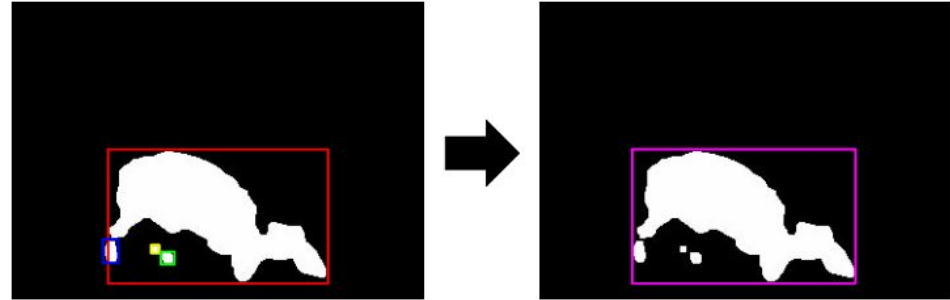
Recall : $TP / (TP + FN)$

Precision : $TP / (TP + FP)$

F-Measure : $(2 * Precision * Recall) / (Precision + Recall)$

minimize $\|L\|_* + \lambda \|S\|_1$
subject to $L + S = M.$

$$IoU = \frac{A_{pred} \cap A_{gt}}{A_{pred} \cup A_{gt}}$$



$$\min_{\alpha_0, \alpha} \left(\frac{1}{2N} \sum_{i=1}^N (y_i - \alpha_0 - \mathbf{x}_i^T \alpha)^2 + \gamma \sum_{j=1}^p |\alpha_j| \right)$$