



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
- 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

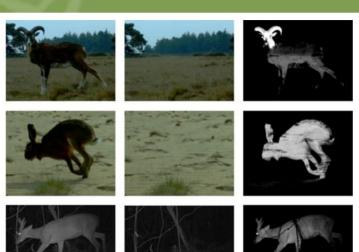


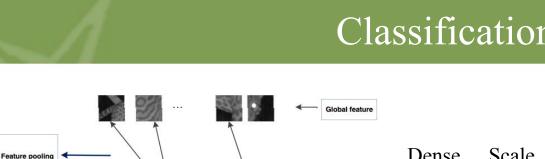
Figure 4: Previous works Source: [Zhang et al., 2016]

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

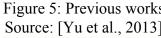


Feature map

Local feature

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

Invariant Transform Dense Scale Feature 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].





Feature coding

Local feature



Methods Multi-Layer RPCA

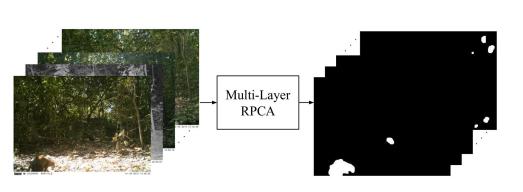


Figure 6: Segmentation algorithm, Multi-Layer Robust Principal Component Analysis (Multi-Layer RPCA)

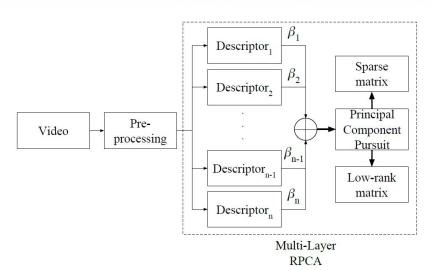


Figure 7: Multi-Layer RPCA method





Methods (2) CNN

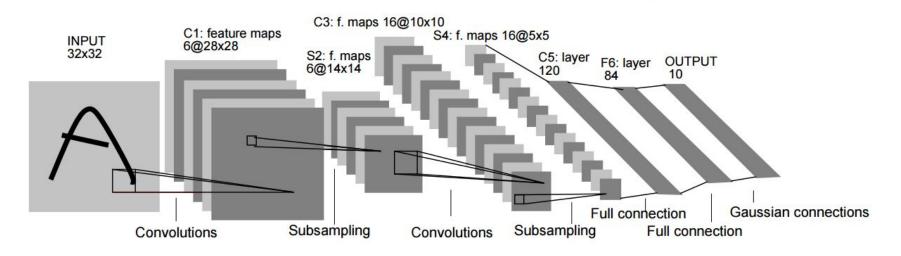
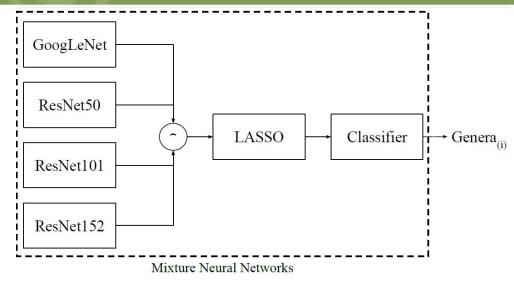


Figure 8: Convolutional Neural Network (LeNet) [LeCun et al., 1998]





Methods (3) 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 (4)

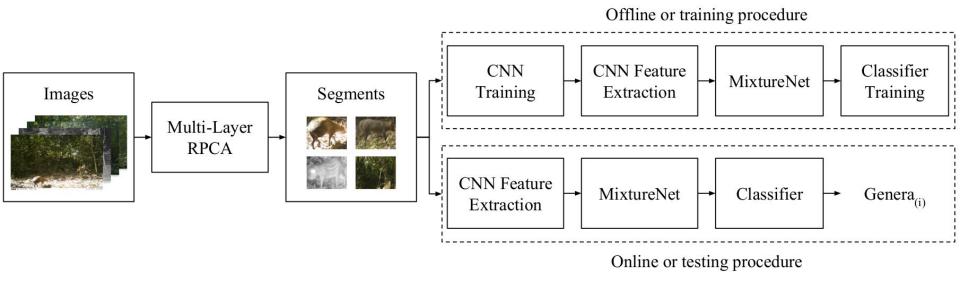


Figure 10: Pipeline of the mammal classification method presented in this work





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.
 - o 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).
 - o Direct evaluation of CNNs on test (non-optimistic results), without LASSO.





Results Segmentation

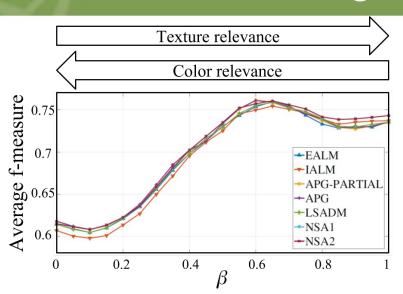


Figure 14: Results with color images

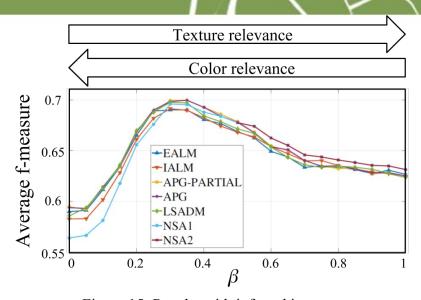


Figure 15: Results with infrared images





Results (2) Segmentation

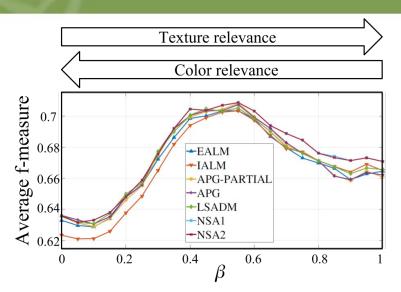


Figure 16: Results with color and infrared images

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)

	Accuracy [%]		Accuracy LASSO [%]			T.	
CNN	ANN	LSVM	RSVM	ANN	LSVM	RSVM	Kappa
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

	Accuracy [%]			Accuracy LASSO [%]			¥7.
CNN	ANN	LSVM	RSVM	ANN	LSVM	RSVM	Kappa
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 IoU > 0.5 and IoU = 0 for animals and False Positives, respectively. Certainly, regions with $0 < IoU \le 0.5$ are limitations of our method.





Publications



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

Hony-Heriberto Giraldo-Zalunga, Augusto Salazar und Alexander Gomez Gropo de Investigación SISTEMIC, Inversidad de Antioquia, Medellis, Colombia, heriberto, giraldo/forbaccola co Od. gustosularar@udea.edu.co, alvianles@gmail.com

Instituto de Investigación de Recursos Biológicos Alexander von Humboldt, Begoti D.C., Colombia. adiax@humboldt.org.co

Above—The expensation and durabative of amount from the contract which to leavage or their. The end present is required and the contract of the contract to the contract of the contract of

Local Bases, Protons a future creation from conversion was to be a final point of the control of

we confused the same specie-recognition problem from the with a CNN arithment with 3 consolutional and 3 mas congress existing leavage should rely experiment and problem of the problem of the contract with a CNN and the contract was in Alberday the current maps, should rely experiment and the problem of the problem of the contract with a contract w

Conference: IEEE International Conference on Tools with Artificial Intelligence 2017 Hindex: 28 (Scimago)



Camera-Trap Images Segmentation using Multi-Layer Robust Principal Component Analysis

Jhony-Heriberto Giraldo-Zulunga † - Augusto Salazar † - Alexander Gomez † - Angelica Diaz-Pullido 2

Abstract The segmentation of animals from conservation images in a difficult task. To Bastrate, there are various challenges due to entimentated conditions. Studying and menitoring salimal species can be p 100 WHIGH GEREIGES (2010 to sourcement seasons and landrous limitation in these issues, we proposed as Multi-Layer Robert Principal Component Analysis their layer Robert Principal Component Analysis the strength of the Component Analysis and the season of the Component Analysis and the Component Principal Robert Principal Component (Component Principal Component Principal Component Principal Component (Component Principal Component Principal Component Principal Component (Component Principal Component Principal Component Principal Component (Component Principal Component Principal Component (Component Principal Component Principal Component Principal Component (Component Principal Component Principal Component (Component Principal Component Principal Component Principal Component (Component Principal Component Principal Component (Component Principal Component Principal Component Principal Component Principal Component (Component Principal Component Principal Component Principal Component Principal Component (Component Principal Component Principal Component Principal Component Principal Component (Component Principal Component Principal Component Principal Component Principal Component Principal Component Principal Component (Component Principal Component Principal Component Principal Component Princi On the trade complex special configuration of the control of the c

Keywords Camer-trap images - Multi-Layer Ro-bust Pitaciyal Component Asalysis - bankgawan subtraction image agraentation This work was supported by the Colombian National Head for Science, Technology and Innevention, Prancisco Ford de Colom-Science, Technology and Innevention, Prancisco Ford de Colom-COLCHENCIAS (Colombia), Prajest No. 111171451061.

Science Verific Associated and Prajest Praject Praje

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





To do

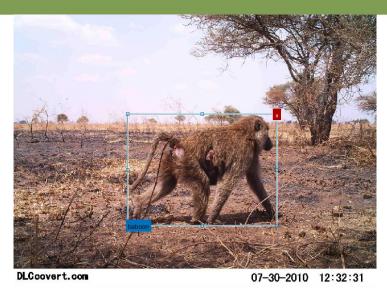


Figure 17: Snapshot Serengeti Source: [Swanson et al., 2015]





Acknowledgements

This work was supported by the Colombian National Fund for Science, Technology and Innovation, Francisco José de Caldas - COLCIENCIAS (Colombia). Project No. 111571451061.







Contacts

Jhony Heriberto Giraldo Zuluaga

email: heriberto.giraldo@udea.edu.co

email: jhonygiraldoz@gmail.com

webpage: https://sites.google.com/view/jhonygiraldo

Prof. Dr. Augusto Enrique Salazar Jiménez

email: augusto.salazar@udea.edu.co

Prof. Dr. Claudia Victoria Isaza Narvaez

email: victoria.isaza@udea.edu.co

















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)

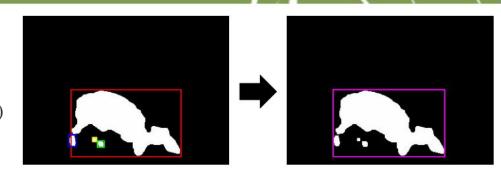
Precision: TP/(TP + FP)

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

minimize
$$||L||_* + \lambda ||S||_1$$

subject to $L + S = M$.

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



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



