A BIRDSAI View for Conservation

Elizabeth Bondi[*], **Milind Tambe**
Harvard University

Raghav Jain, Palash Aggrawal, Saket Anand
IIIT-Delhi

Robert Hannaford
Air Shepherd

Ashish Kapoor, Jim Piavis, Shital Shah, Lucas Joppa
Microsoft

Bistra Dilkina
University of Southern California

Abstract

Monitoring of protected areas to curb poaching is a monumental task for law enforcement authorities. To augment existing manual patrolling efforts to locate humans in protected areas, unmanned aerial surveillance using visible and Thermal Infrared (TIR) spectra cameras is increasingly being adopted. However, it is still a challenge to accurately and quickly process large amounts of the resulting TIR data. In this paper, we present a challenge based on the first large dataset collected using a TIR camera mounted on a fixed-wing UAV, which was flown over multiple protected areas in African national parks. This dataset includes TIR videos of humans and animals with several challenging scenarios like scale variations, background clutter due to thermal reflections, large camera rotations and occasional motion blur. We also provide synthetically-generated videos with the publicly available Microsoft AirSim simulation platform using a 3D model of an African savanna and a TIR camera model.

1 Introduction

In this paper, we introduce a challenge based upon Benchmarking IR Dataset for Surveillance with Aerial Intelligence (BIRDSAI, pronounced "birdseye"), a large, challenging aerial TIR video dataset that will help with benchmarking of algorithms for automatic detection and tracking of humans and animals (Bondi et al. (2020)). To our knowledge, this is the first large-scale aerial TIR dataset, with multiple unique features: It has 48 real aerial TIR videos of varying lengths, carefully annotated with objects like animals and humans and their trajectories. These were collected by conservation organizations during their regular surveillance efforts by flying a fixed-wing UAV over national parks in Southern Africa. Finally, we augment it with 80 synthetic aerial TIR videos generated from AirSim-W (Bondi et al. (2018)), an Unreal Engine-based simulation platform. Two example images from real videos are shown in Fig. 1 depicting a herd of elephants and a human. Realistic and challenging benchmarking datasets have had tremendous impact on the progress of a research area. Synthetic datasets like Richter et al. (2016); Ros et al. (2016) along with real ones like Cordts et al. (2016) have accelerated the progress in unsupervised domain adaptation techniques (Lee et al. (2019); Tsai et al. (2018)). Similarly, the Caltech-UCSD Bird (CUB-200) dataset (Wah et al. (2011)) has helped advance an exciting and important area of fine-grained visual recognition (Zhao et al. (2017)). With more wildlife monitoring datasets (Swanson et al. (2015); Beery et al. (2018); Witham (2017)) becoming publicly available, we may expect rapid progress in areas like visual animal biometrics (Crall et al. (2013); Cheema and Anand (2017); Kumar and Singh (2017)). Inspired by these instances,

[*]ebondi@g.harvard.edu

33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.

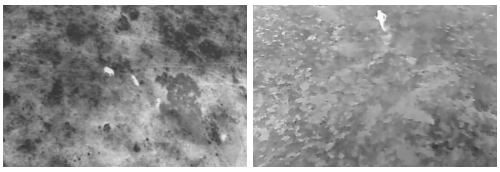

Figure 1: Example images from BIRDSAI: elephants and a human, respectively, from an aerial perspective.

we anticipate the challenge will promote advances in both (i) algorithm development for the general problems of object detection, single and multi-object tracking in aerial videos and their domain adaptive counterparts, and (ii) the important application area of aerial surveillance for conservation.

2 Dataset Description

2.1 Real Data

Data were collected throughout protected areas in the countries of South Africa, Malawi, and Zimbabwe using a battery powered fixed-wing UAV. Specific locations are withheld for security. All flights took place at night, with an individual flights lasting for about 1.5 - 2 hours, but no longer than 2.5 hours. Various environmental factors such as wind resistance determined this variation in flying time. Throughout the night, there were typically 3 to 4 flights, and the altitude ranged from approximately 200 to 400 ft (60 to 120m), and flight speed ranged from 12 to 16 m/s depending on conditions such as wind. Temperature ranged from less than 0°C to 4°C in winter at night, though typically closer to 4°C. There was often with a shift of approximately 5°C throughout the course of the night in the winter. For reference, daytime temperatures were typically approximately 15°C to 16°C. During summer, the temperature ranged from 18°C to 20°C at night, and 38°C to 40°C during the day. When flying just after sunset, the ground temperature is warm and can make it more difficult to spot objects of interest due to the lack of contrast. However, by about 10:30-11PM, there is typically sufficient contrast for easier visibility. Fog was present in some rare cases, which could cause "whiteouts" in images. The FLIR Vue Pro 640 was the primary sensor utilized.

We used VIOLA (Bondi et al. (2017)) to assist in labeling detection bounding boxes in the thermal infrared imagery. After labels were made by one person, two other people reviewed the labels. Other methods of labeling were tested, but this was found to be the most effective. All labels were finally confirmed and checked for quality for use in the dataset by the authors.

2.2 Synthetic Data

To generate synthetic data with AirSim, we utilized the African savanna environment introduced in Bondi et al. (2018). In brief, the environment is not based on a particular area of interest, but rather represents the variety of environments found in Southern Africa, such as wide-open plains to dense forest, flatland to mountainous terrain, roads, and water. The AirSim platform has a TIR model that was introduced in Bondi et al. (2018). We used this TIR model to generate images of the objects in the scene as the TIR camera mounted UAV flew through the environment.

3 Challenge

A challenge will be hosted based on this dataset. It will be posted on a standard challenge-hosting website (e.g., Kaggle), and will also use a hackathon event (e.g., Zoohackathon) to kick off the challenge. Training data will be freely available to all to facilitate wide participation, as this dataset and methodology will lead to real-world benefit.

References

Sara Beery, Grant Van Horn, and Pietro Perona. 2018. Recognition in Terra Incognita. In *The European Conference on Computer Vision (ECCV)*.

Elizabeth Bondi, Debadeepta Dey, Ashish Kapoor, Jim Piavis, Shital Shah, Fei Fang, Bistra Dilkina, Robert Hannaford, Arvind Iyer, Lucas Joppa, and Milind Tambe. 2018. AirSim-W: A Simulation Environment for Wildlife Conservation with UAVs. In *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies (COMPASS '18)*. Article 40, 12 pages.

Elizabeth Bondi, Fei Fang, Debarun Kar, Venil Noronha, Donnabell Dmello, Milind Tambe, Arvind Iyer, and Robert Hannaford. 2017. VIOLA: Video Labeling Application for Security Domains. In *Proceedings of the 8th Annual Conference on Decision Theory and Game Theory for Security (GameSec)*.

Elizabeth Bondi, Raghav Jain, Palash Aggrawal, Saket Anand, Robert Hannaford, Ashish Kapoor, Jim Piavis, Shital Shah, Lucas Joppa, Bistra Dilkina, and Milind Tambe. 2020. BIRDSAI: A Dataset for Detection and Tracking in Aerial Thermal Infrared Videos. In *(To Appear) Proceedings of the IEEE Winter Conference on Applications of Computer Vision*.

Gullal Singh Cheema and Saket Anand. 2017. Automatic Detection and Recognition of Individuals in Patterned Species. In *ECML PKDD*. `https://doi.org/10.1007/978-3-319-71273-4{_}3`

Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. 2016. The Cityscapes Dataset for Semantic Urban Scene Understanding. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

J.P. Crall, C.V. Stewart, T.Y. Berger-Wolf, D.I. Rubenstein, and S.R. Sundaresan. 2013. HotSpotter – Patterned species instance recognition. In *Applications of Computer Vision (WACV), 2013 IEEE Workshop on*. 230–237.

S. Kumar and S. K. Singh. 2017. Visual animal biometrics: survey. *IET Biometrics* 6, 3 (2017), 139–156.

Kuan-Hui Lee, German Ros, Jie Li, and Adrien Gaidon. 2019. SPIGAN: Privileged Adversarial Learning from Simulation. In *International Conference on Learning Representations*. `https://openreview.net/forum?id=rkxoNnC5FQ`

Stephan R. Richter, Vibhav Vineet, Stefan Roth, and Vladlen Koltun. 2016. Playing for Data: Ground Truth from Computer Games. In *European Conference on Computer Vision (ECCV) (LNCS)*, Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling (Eds.), Vol. 9906. Springer International Publishing, 102–118.

German Ros, Laura Sellart, Joanna Materzynska, David Vazquez, and Antonio M. Lopez. 2016. The SYNTHIA Dataset: A Large Collection of Synthetic Images for Semantic Segmentation of Urban Scenes. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

AB Swanson, M Kosmala, CJ Lintott, RJ Simpson, A Smith, and C Packer. 2015. Data from: Snapshot Serengeti, high-frequency annotated camera trap images of 40 mammalian species in an African savanna.

Y.-H. Tsai, W.-C. Hung, S. Schulter, K. Sohn, M.-H. Yang, and M. Chandraker. 2018. Learning to Adapt Structured Output Space for Semantic Segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie. 2011. *The Caltech-UCSD Birds-200-2011 Dataset*. Technical Report.

Claire L Witham. 2017. Automated face recognition of rhesus macaques. *Journal of neuroscience methods* (2017).

Bo Zhao, Jiashi Feng, Xiao Wu, and Shuicheng Yan. 2017. A survey on deep learning-based fine-grained object classification and semantic segmentation. *International Journal of Automation and Computing* 14, 2 (2017), 119–135.