Remote Sensing of Environment 200 (2017) 341–351

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Detecting animals in African Savanna with UAVs and the crowds

Nicolas Rey[a], Michele Volpi[b], Stéphane Joost[a,c], Devis Tuia[b,d,*]

[a] Laboratory of Geographic Information Systems (LASIG), School of Architecture, Civil and Environmental Engineering (ENAC), Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland
[b] MultiModal Remote Sensing, Department of Geography, University of Zurich, Switzerland
[c] The Savmap Consortium
[d] Laboratory of Geo-Information Science and Remote Sensing, Wageningen University & Research, The Netherlands

ARTICLE INFO

Keywords:
Animal conservation
Wildlife monitoring
Object detection
Active learning
Crowd-sourcing data
Unmanned aerial vehicles
Very high resolution

ABSTRACT

Unmanned aerial vehicles (UAVs) offer new opportunities for wildlife monitoring, with several advantages over traditional field-based methods. They have readily been used to count birds, marine mammals and large herbivores in different environments, tasks which are routinely performed through manual counting in large collections of images. In this paper, we propose a semi-automatic system able to detect large mammals in semi-arid Savanna. It relies on an animal-detection system based on machine learning, trained with crowd-sourced annotations provided by volunteers who manually interpreted sub-decimeter resolution color images. The system achieves a high recall rate and a human operator can then eliminate false detections with limited effort. Our system provides good perspectives for the development of data-driven management practices in wildlife conservation. It shows that the detection of large mammals in semi-arid Savanna can be approached by processing data provided by standard RGB cameras mounted on affordable fixed wings UAVs.

1. Introduction

In the fragile ecosystems of semi-arid Savanna, any change in the equilibrium between precipitation, grazing pressure and bush fires can lead to long-term land degradation, such as the reduction in grass cover and bush encroachment (Trodd and Dougill, 1998). To avoid overgrazing, the populations of grazers must be kept in adequacy with the grass availability, which is subject to meteorological conditions. For this purpose, land managers need to regularly estimate the amount of wildlife present on their territory. Thus, monitoring wildlife populations is crucial towards conservation in wildlife farms and parks.

To carry out wildlife censuses, traditional methods include transect counts on land or from a helicopter, and camera traps. While a total count is usually not possible over large areas, these methods estimate the population density based on observations localized along a predefined path (see (Aebischer et al., 2017; Alienor et al., 2017) and references therein). These methods are expensive (e.g. in the case of the Kuzikus reserve considered in this paper, helicopter costs for a single survey are between 1000$ and 2500$), require trained human experts to screen large amounts of data and are consequently not suitable for regular censuses over large areas.

In recent years, unmanned aerial vehicles (UAVs) have been used to detect and count wildlife such as birds, marine mammals, and large herbivores (Linchant et al., 2015). Compared to traditional methods, UAVs offer several advantages: they cover large areas in a short amount of time and can be used in inaccessible and remote areas, yet they are cheaper and easier to deploy than helicopters. Moreover, they are safer for the pilot, who can stay on the ground and avoiding retaliations from poachers.

However, UAVs collect large amounts of color images with sub-meter to sub-decimeter spatial resolution, of which only few contain animals. Furthermore, the animals cover only an infinitesimal area of the images and their color might blend in smoothly with background vegetation and soil. Therefore, identifying and counting single animals across large collections of images is extremely complex and time-consuming, preventing land managers from using UAVs on a regular basis.

Despite these challenges, recent developments in object detection pipelines in both computer vision (Girshick et al., 2014; Malisiewicz et al., 2011) and remote sensing (Akçay and Aksoy, 2016; Moranduzzo and Melgani, 2014; Tuermer et al., 2013), provide promising techniques to semi-automatically localize and count animals. We refer to these methods as semi-automated and not as fully automated since they rely on supervised learning paradigms, thus requiring annotated ground truth to be trained. Still, as the human effort required to make sense of the aerial images is reduced, the overall benefits of using UAVs are significantly increased.

* Corresponding author.
E-mail address: devis.tuia@wur.nl (D. Tuia).

http://dx.doi.org/10.1016/j.rse.2017.08.026
Received 27 March 2017; Received in revised form 18 July 2017; Accepted 16 August 2017

The use of UAVs in wildlife monitoring and conservation is well documented (e.g. Linchant et al. (2015)), but only few studies have implemented semi-automatic detection pipelines. Grenzdörffer (2013) proposes to detect seagulls by combining supervised classification of RGB images with geometric rules. Kudo et al. (2012) present a pipeline to count salmons in aerial images using simple color thresholding after contrast adjustment. Such approaches are only possible if the animals are visually very similar and exhibit distinctive colors that contrast with the background. Chabot and Bird (2012) detect geese by manual counting single animals in UAV images. Maire et al. (2014) adopt more advanced machine learning tools for the detection of dugongs. They obtain promising results by training a deep neural network and address the problem of scarcity of training samples by replicating them through random rotations and scalings applied to confident missclassifications (a technique related to hard negative mining (Malisiewicz et al., 2011)).

In this paper, we present a data-driven machine learning system for the semi-automatic detection of large mammals in the Savanna ecosystem characterized by complex land-cover. We perform animal detection on a set of sub-decimeter resolution images acquired in the Namibian Kalahari desert and train our system using animals annotated by digital volunteers using the MicroMappers crowdsourcing platform (Ofli et al., 2016). We focused on large mammals for two main reasons: first, they stood out compared to the background, while smaller animals such as meerkats are not clearly visible and could be too easily confused with rocks or bushes by the volunteers. Secondly, larger animals also mean more pixels available to learn the appearance of the animals' furs, which leads to less signal mixing, to more discriminative features and to a more accurate system overall. We show that the system achieves high recall rate, and high overall accuracy can be obtained if a human operator can verify the detections, reduce the false positives and verify true negatives, and retrain the detector. This last technique, known as active learning (Tuia et al., 2011), aims at focusing the operator's effort on instances with low detection confidence and its benefits are shown by our experimental results, where only 1 h was required to correct the crowd-sourced dataset of several errors (mainly animals missed by the volunteers). The main contributions of the paper are:

- A pipeline for semi-automatic animal detection in semi-arid Savanna that uses affordable UAV platforms with off-the-shelf RGB cameras;
- A complete study of the model's parameters to provide intuitions about the trade-offs between acquisition settings, image resolution and the complexity of the appearance descriptors involved;
- A discussion of the promising performances of the system in a real deployment scenario in the Kuzikus reserve in Namibia, including the quasi real time improvement of the model.

2. Study area and data

2.1. The Kuzikus wildlife reserve

Kuzikus is a private wildlife reserve that covers 103 km^2 (10,300 ha), located on the edge of the Kalahari in Namibia. The Kalahari is a semi-arid sandy Savanna that extends across Botswana, South Africa and Namibia. It is home of an important variety of animals, including many large mammal species (Reinhard, 2016). About 3000 individuals from more than 20 species populate the reserve, including Common Elands (*Taurotragus oryx*), Greater Kudus (*Tragelaphus strepsiceros*), Gemsboks (*Oryx gazella*), Hartebeests (*Alcelaphus buselaphus*), Gnus (*Connochaetes gnou* and *Connochaetes taurinus*), Blesboks (*Damaliscus albifrons*), Springboks (*Antidorcas marsupialis*), Steenboks (*Raphicerus campestris*), Common Duickers (*Sylvicapra grimmia*), Impalas (*Aepyceros melampus*), Burchell's Zebras (*Equus quagga burchellii*), Ostriches (*Struthio camelus australis*) and Giraffes (*Giraffa camelopardalis giraffa*).

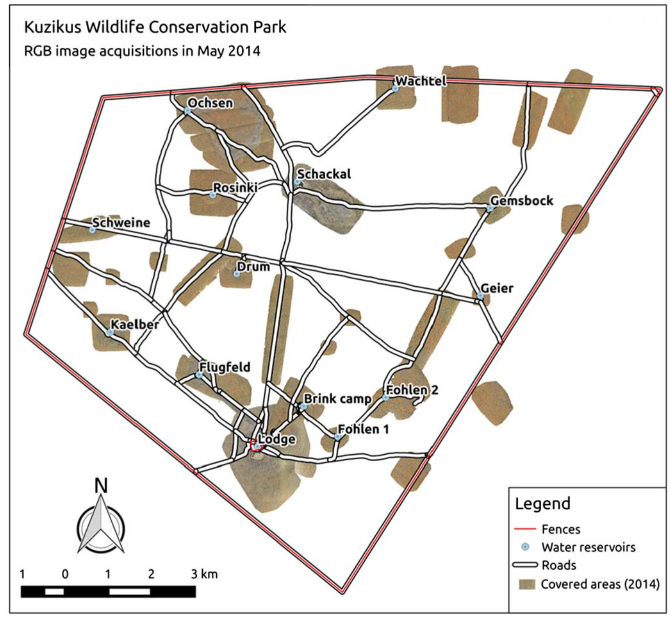

Fig. 1. Map of the Kuzikus Wildlife Conservation Park and areas covered by the 2014 RGB dataset.

2.2. The SAVMAP 2014 UAV campaign

The SAVMAP Consortium (http://lasig.epfl.ch/savmap) acquired a large aerial image dataset during a two-week campaign in May 2014. It is composed of five flights, between May 12 and May 15, 2014. The images were acquired with a Canon PowerShot S110 compact camera mounted on an eBee, a light UAV commercialized by SenseFly (https://www.sensefly.com). Each image is 3000×4000 pixels in size and comprises three bands in the Red Green and Blue (RGB) domains, with a radiometric resolution of 24 bits. The ground sampling distance is approximately 4 cm per pixel. The extent of the reserve mapped by the 2014 SAVMAP campaign is illustrated in Fig. 1.

2.3. Animals annotation via crowd sourcing

In order to obtain a ground truth of the position of all large animals, a crowd-sourcing campaign was set by MicroMappers (https://micromappers.wordpress.com/). A total of 232 digital volunteers participated in the operation. The volunteers were asked to draw a polygon around each animal they detected in the images, without distinction between species. They did not have to report signs of animal presence, such as Aardwolves' holes or termite mounds. Each image was inspected by at least three volunteers, with a maximum of ten. On average, the images were seen by five volunteers (Ofli et al., 2016).

The volunteers visually analyzed 6500 images and drew 7474 polygons in 654 images containing animals. After merging the overlapping polygons and removing objects tagged only by a single volunteer (as the bottom right annotation in Fig. 2), 976 annotations were kept. Since the number of volunteers per image varied between three and ten, we used as ground truth the surface that was tagged as “animal” by at least half of the annotators who considered it (areas in green-to-yellow colors in the right panel Fig. 2). To avoid false annotations, we visually inspected them to confirm or infirm animals presence. It took 30 min to verify the 976 annotations, leading to the removal of 21 spurious ones. More details on the acquisition of annotations can be found in Ofli et al. (2016). Note that the same animals could be observed in several consecutive, overlapping images. This effect is beneficial when training appearance models, since the different angles and poses characterizing animals better cover the appearance variability of the class of interest. However, note that the