Ecological Informatics 34 (2016) 108–116

Contents lists available at ScienceDirect

Ecological Informatics

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

Classifying and ranking audio clips to support bird species richness surveys

Liang Zhang *, Michael Towsey, Jinglan Zhang, Paul Roe

Eco-acoustics group, Science and Engineering Faculty, Queensland University of Technology, Australia

ARTICLE INFO

Article history:
Received 24 February 2016
Received in revised form 25 May 2016
Accepted 25 May 2016
Available online 27 May 2016

Keywords:
Acoustic sampling
Classification
Acoustic monitoring
Ranking
Bird species richness

ABSTRACT

Advances in programmable field acoustic sensors provide immense data for bird species study. Manually searching for bird species present in these acoustic data is time-consuming. Although automated techniques have been used for species recognition in many studies, currently these techniques are prone to error due to the complexity of natural acoustics.

In this paper we propose a smart sampling approach to help identify the maximum number of bird species while listening to the minimum amount of acoustic data. This approach samples audio clips in a manner that can direct bird species surveys more efficiently. First, a classifier is built to remove audio clips that are unlikely to contain birds; second, the remaining audio clips are ranked by a proxy for the number of species. This technique enables a more efficient determination of species richness.

The experimental results show that the use of a classifier enables to remove redundant acoustic data and make our approach resilient to various weather conditions. By ranking audio clips classified as “Birds”, our method outperforms the currently best published strategy for finding bird species after 30 one-minute audio clip samples. Particularly after 60 samples, our method achieves 10 percentage points more species. Despite our focus on bird species, the proposed sampling approach is applicable to the search of other vocal species.

1. Introduction

Bird species are good indicators of environmental health and have been used to monitor the dynamic change of the natural environment (Carignan and Villard 2002; Catchpole and Slater 2003). The use of acoustics to monitor birds confers several advantages (Bardeli et al. 2010). First, it allows for covering a large spatial area; second, it provides continuous recordings for a long period of time; third, it functions well even with poor lighting or visual impediment; and fourth, audio signals are cheaper to store and compute than visual signals. With acoustic recordings, the cost of in-the-field observation is translated into an analysis problem. This paper focuses on the study of bird species richness (Spellerberg and Fedor 2003), which aims to determine the number of unique bird species in a specific area within a specific period of time using post-analysis of acoustic recordings.

1.1. Bird species richness survey

Due to spatiotemporal limitations, conducting an in-the-field bird species richness survey requires effective sampling protocols (Schneider 1994). Point count is one of the most popular sampling protocols where skilled bird observers document bird species they encounter in a specific site at fixed period of time (Huff et al. 2000). Normally, these recorded bird species can be subjective and hard to verify.

Acoustic sensor offers an effective approach to collect data at large spatiotemporal scales (Acevedo and Villanueva-Rivera 2006). The recorded acoustic data can be stored permanently and provide a convenient way to verify bird species. However, the increased dataset also necessitates the development of efficient techniques.

The difficulty of conducting the bird species richness survey by acoustics lies in the diversity of bird vocalizations. Competition for the acoustic space and environmental constraints, such as temperature and vegetation structure, may lead to significant variations within and between species vocalizations (Farina 2014). Additionally, simultaneous vocalizations could make the acoustic recognition of bird species even more difficult.

Manually listening to audios and inspecting the corresponding spectrograms for bird species identification is reliable, if experienced persons are involved, but time-consuming. For example, a one-minute audio clip often requires twice the time to investigate because people frequently replay the audio to identify which species are vocalizing (Wimmer et al. 2013). Although automated techniques offer computational power to alleviate this problem, their development is still in infancy. Unlike human speech and music, bird vocalizations are less structured and their repetition is unpredictable. These problems hinder the use of automated techniques for bird species identification.

Several studies have contributed to developing automated techniques for bird species recognition (Fagerlund 2007; Kasten et al.

* Corresponding author at: Room 1002, Level 10, S Block, Gardens Point Campus, 2 George Street, Brisbane, QLD 4000, Australia.
E-mail address: l68.zhang@hdr.qut.edu.au (L. Zhang).

http://dx.doi.org/10.1016/j.ecoinf.2016.05.005
1574-9541/

2010; Somervuo et al. 2006). These methods work well when vocalization structures are simple. For example, only a single species that has multiple vocal types is calling or several different species sharing a common type of vocal structure. Automated call recognition is a promising alternative for acoustic data analysis, but accuracy is still far from perfect, especially on detecting vocal species recorded from the natural environment. Recently, a multi-label classification method has been introduced to detect bird vocalizations (Briggs et al. 2012). Unlike prior work on single-label classification where there is only one label associated with a study object, multi-label classification is possible to associate the study object with multiple labels, providing a potential solution to recognize simultaneous bird vocalizations. However, this method is a supervised machine learning that can only predict predefined labels in the training data; consequently, it cannot handle any unexpected vocalizations that may appear.

Confronted with a large volume of data, Wimmer et al. (2013) first introduced sampling methods to assist bird species richness survey. They compared five temporal sampling strategies on a one-day recording, pointing out that the most efficient strategy to find bird species is dawn sampling. Here dawn sampling is referred to randomly selecting audio clips 3 h after dawn. However, this is an intuitive approach based on the fact that many bird species vocalize during dawn. There are no further instructions on how to effectively investigate these 3-hour acoustic data. They also suggested that using automated techniques to locate periods that are likely to contain unique species might improve the efficiency of bird species surveys. The use of automated techniques to direct the sampling of acoustic data for bird species surveys is called "*smart sampling*". Prior works have shown the ability of using the linear regression (Towsey et al. 2013) or the clustering technique (Eichinski et al. 2015) for smart sampling, but they did not take into consideration of various weather conditions such as heavy rain and strong wind gusts that can affect the efficiency of these methods.

1.2. Acoustic characteristics of audio clips

Direct use of automated techniques may not be effective in the case of non-targeted and multi-species inventories; nevertheless, we can still utilize these techniques to make bird species recognition more efficient than manual analysis. Fig. 1 shows five commonly encountered examples of one-minute audio clips. We categorize them as "Birds", "Insects", "Low activity", "Rain", and "Wind". In temperate woodland ecosystems in spring, a prior study describes the daily acoustic activity to be birds vocalizing during the day and insects chirping from sunset to the sunrise of the next day (Wimmer et al. 2013). Occasional heavy rain and strong wind gusts are important acoustic information because they may interrupt bioacoustics activities. We also define low activity as the time when little amount of acoustic energy is recorded. Apparently, removing data that do not contain bird species can improve the efficiency of species finding. Since these five acoustic patterns have discriminative time-frequency characteristics, it is possible to filter the other four patterns from "Birds" using a classification method.

1.3. Indicators of acoustic diversity

Biodiversity assessment is one of the most challenging problems that ecologists are confronted with. Indices have been used to

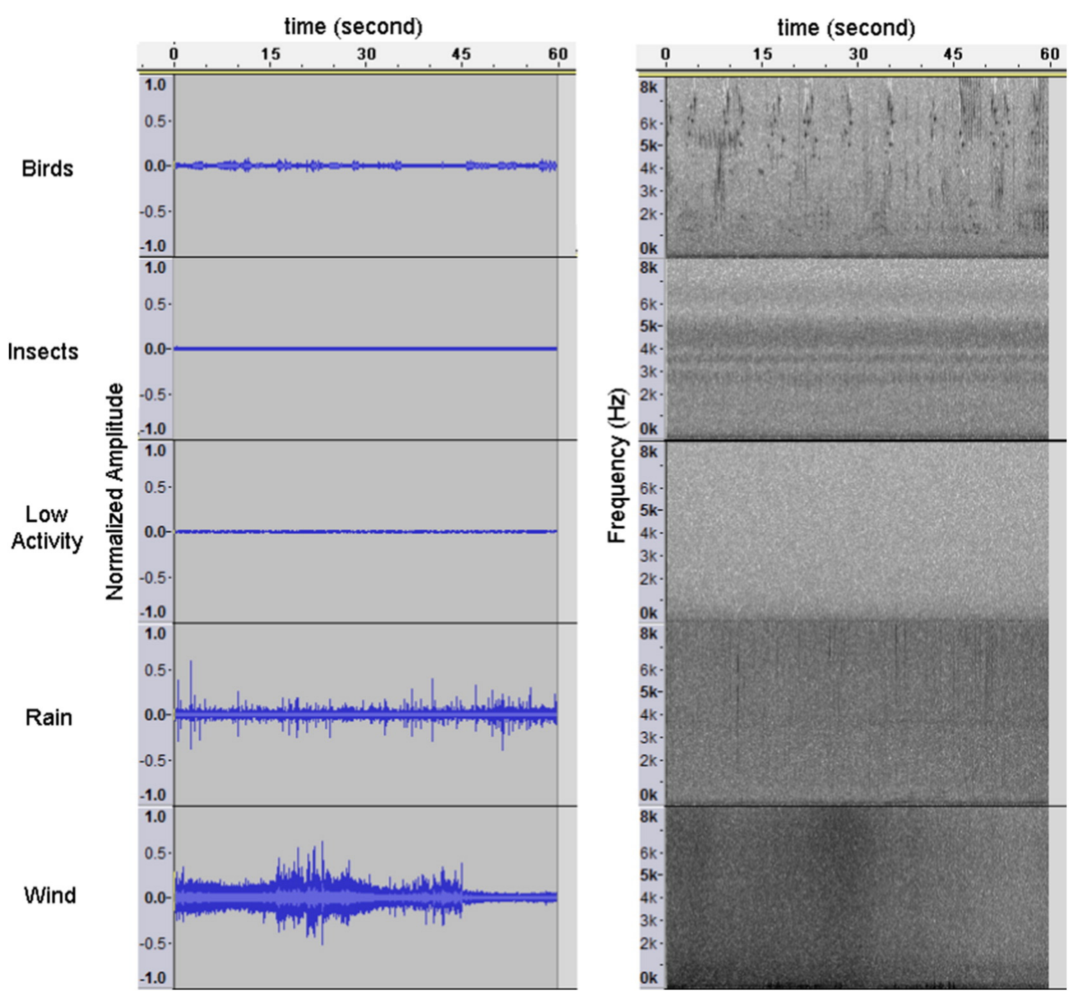

Fig. 1. Characteristics of five acoustic patterns in one-minute audio clips (left) and their corresponding spectrograms (right).