Entomological remote dark field signal extraction by maximum noise fraction and unsupervised clustering for species identification

Benoit Kouassi Kouakou^{*†}, Eliane Koko Nadine Assoi[†], Adolphe Yatana Gbogbo[†] and Jeremie Zoueu[†]

Laboratoire d'Instrumentation, Image et Spectroscopie; Institut National Polytechnique Houphouët Boigny, BP 1093 Yamoussoukro, Côte d'Ivoire

*Author for correspondence (benoitkouakou@gmail.com).

†Present address : Laboratoire d'Instrumentation Image et Spectroscopie, Institut National Polytechnique Houphouët Boigny, BP 1093, Yamoussoukro, Côte d'Ivoire

1. Abstract

Characterization of flying insects *in-situ* measurement using remote sensing spectroscopy is an emerging research field. Also, most analysis techniques in remote sensing spectroscopy are based on the use of an intensity threshold which introduces indeterminacies in the number of detected specimens. In this manuscript, we investigated the possibility of analyzing passive remote sensing spectroscopy measurement data using the maximum noise fraction method. The results obtained show that this analysis technique can help to overcome the measurement of background noise in spectroscopic measurements.

Keywords: Remote sensing; spectroscopy; blind source separation; unsupervised clustering; insects

2. Introduction

Remote sensing spectroscopy is an indirect measurement technique used to characterize aerial fauna and atmospheric particles. This measurement technique is done through instrument generally composed by reflector to collect backscattered light by object of interest, refractor to extend incident beam and sensors to measure backscattered light [1]. Several remote sensing spectroscopy instruments are been developed in recent decades to measure the spectral footprint [2,3], wing beat frequency [4], optical cross section of flies species [5,6]. These instruments are calibrated sometimes to measure fly direction [4] and detection range of insects [7,8]. These measurement instrument do not capture simultaneously image and identification parameters of the specimens studied, namely spectral footprint, wing beat frequency and optical cross section. Also, backscatter intensities due to insects are extracted using intensity threshold which sometimes omits intensities backscattered by small specimens [2]. Intensity threshold derived in conjunction of median intensities, the interquartile range of intensities and the choice of an arbitrary constant related to the signal to noise ratio of each data file has been proposed [9]. The arbitrary choice of this constant is subjective and leads to loss of information. Blind source separation methods seem to be an attractive alternative [10–12]. These information extraction techniques have been widely used in biomedical engineering, speech recognition systems and telecommunications. An extension of this technique has been proposed by some authors under the name of maximum noise fraction (MNF) [13]. In order to overcome the shortcomings of conventional indirect measurement systems, we are investigating a technique for simultaneous indirect measurement of the spectral footprint and image capture of flying specimens and analysis techniques based on the MNF for feature extraction. An unsupervised cluster analysis was done to identify insects' species.

3. Experimental set up and synchronous acquisition of footprint and image

The equipment is composed of two receivers a reflector (Skywatcher telescope ; D=200 mm ; F=1000 mm ; coated optic) and a refractor (Bresser Messier ; AR 90 Achromatic refractor telescope 90/900 f/10.0 ; coated optics) , mounted in parallel on an aluminum bar with a separation distance L (L=500 mm). The whole was fixed on an equatorial mount. The position of the counterweight is adjusted to determine the equilibrium position in the xy plane and in the yz plane. It is often necessary to move on the position of the telescope to adjust the balance in the yz plane. The determination of the equilibrium point allows flexibility in the handling of the instrument. This also allows the instrument to remain fixed at low atmospheric pressures, which minimizes the margins of error in measurements (Fig. 1).

Figure 1: Panoramic view of the setup. (a) Passive remote sensing equipment consisting with two receivers, a reflector to measure the spectral footprint and the reflector for imaging. (b) Dark box to reduce the background backscatter and avoid saturation of spectrometer. (c) Image of a Lambertian sphere falling in front of the dark box to measure the reference intensity. (d) A schematic diagram of the setup.

Before any measurement, the reflector and the refractor are aligned on the black box located 1 meter above the ground, about 60 m from the experimental device (Fig. 1). This alignment phase consists of first, using an eyepiece, then a CCD camera connected to the computer to observe the orifice of the black box. The alignment phase of the telescope and the telescope allows the spectral and spatial information of the flying specimen to be acquired synchronously. Once the alignment phase is complete, the CCD camera is connected to the telescope then to the computer to record the image and the spectrometer on the telescope then to the computer to record the spectral imprint of the insect which is introduced into the volume defined by the intersection of the solid angle of the refractor and the reflector. The spectrometer and the camera are controlled by a computer program that we have developed using MATLAB for automated acquisition. The reference measurements are carried out by free-falling white Lambertian spheres in the focal plane of the black box. The release time is noted for future identification of the reference data stored on the computer. The integration time of the spectrometer has been set at 20 ms and the number of samples acquired is 1000.

4. Method for extracting and determining the spectral footprint

Raw data, $I_{raw,t}$, recorded is a 16 bits files, containing 3648 pixels x 1000 samples. Backscattered intensities due to dark box reflection and electronic noise intensities generated by electronic chain of spectrometer are add up to different measurements at each instant. These intensities represent the static intensities, $I_{static,t}$. Interquartile range intensity is insensitive to intensity flows backscattered by insects, atmospheric particles, plant leaves, aerosols and other undesirable factors such as smoke entering probe volume. This intensity has been calculated and represent $I_{static,t}$. Then, $I_{raw,t}$ was centered around $I_{static,t}$ to get the intensity I_t . Then, derived intensity I_t is supposed to contain backscattered intensity by an insect $I_{insect,t}$ and intensities backscattered in the volume probed by unknown objects considered as noise, $I_{noisy,t}$, at time t . $I_{insect,t}$ and $I_{noisy,t}$ are assumed to be independent and orthogonal and with zero means. We can write:

$$I_t = I_{insect,t} + I_{noisy,t} \quad (1)$$

The covariance of I_t , $I_t^T I_t$, is express as a function of the covariance of $I_{insect,t}$ and the covariance of $I_{noisy,t}$ [14]:

$$I_t^T I_t = I_{insect,t}^T I_{insect,t} + I_{noisy,t}^T I_{noisy,t} \quad (2)$$

The principle of the maximum noise fraction consists in finding a basic vector, ϕ , such that the maximum noise fraction, f_ϕ , define by equation (3) is optimal [13].

$$f_\phi = \frac{\|I_{noisy,t} \phi\|}{\|I_t \phi\|} \quad (3)$$

To determine the value of ϕ which makes optimal f_ϕ it is necessary to set to zero the derivative of f_ϕ with respect to ϕ . The calculus led to equation (4). This equation is the generalized eigenvalues problem.

$$I_{noisy,t}^T I_{noisy,t} = f_\phi I_t^T I_t \quad (4)$$

The solution of this equation to generalized eigenvalues leads to obtaining the values of the linear transformation matrix, B . This matrix separates and orders the non-correlated spectra by decreasing values of the maximum noise fraction, f_ϕ . The equation of this linear transformation links the uncorrelated spectral intensity, Y_t , with measured backscattered intensity, I_t :

$$Y_t = B^T I_t \quad (5)$$

Knowledge of background noise, $I_{noisy,t}$, in this methodology is required to derive B . Several existing techniques allow to estimate $I_{noisy,t}$. The method of differentiating neighboring pixels is generally used in image processing. $I_{noisy,t}$ was evaluated, by determining the difference of the intensities of the neighboring samples.

$$I_{noisy,t} = I_t - I_{t+1} \quad (6)$$

After extracting the spectra of interest, the spectral footprint, R_{insect} , characteristic of each species of specimen, was calculated based on the spectrum of the specimen, I_{insect} , and the spectrum of the Lambertian sphere, I_{ref} , according equation (7). In this equation, the subtraction of the intensity of the background noise does not appear, because the background noise was dissociated from the spectrum of interest during the extraction of the spectrum of interest (Fig. 2).

$$R_{insect} = \frac{I_{insect}}{I_{ref}} \quad (7)$$

Figure 2: Examples of spectral of two events (a,b) Raw data centered around static intensity. (a) corresponds to reference spectra and (b) insect spectrum passing in the probe volume. (c,d) Spectra obtained after application of the maximum noise fraction method. The spectra were ordered according to the decreasing signal to noise ratio (SNR). The spectrum of interest is the high intensity spectrum (blue curves). In (c), highest intensity spectrum is sun spectrum which is used as illumination light; in (d), highest intensity spectrum is the spectrum of an insect detected by the acquisition system. Lower intensity spectra are considered to be background noise.

5. Identification of insects by their spectral footprints and their associated images by unsupervised classification

The spectral fingerprints determined were normalized and then grouped together. The Euclidean distance was then calculated. The decision criterion which allowed the determination of the number of classes (clusters) is a method based on the elbow detection called L-method [15–17]. The L-method consists of dividing the data relating to the associated distances into two batches and finding the best linear fit of the two batches (Fig. 3a) which minimizes the total mean square error, $RMSE_T$, (Fig. 3b) calculate according to the mean square error of the adjustment of each of the lots on the left (Fig. 3a, red curve), $rmse_L$, and right (Fig. 3a, green curve), $rmse_R$, the number of clusters, a , the maximum number of clusters, b , according equation (8).

$$RMSE_T = \frac{a-1}{b-1}rmse_L + \frac{b-a}{b-1}rmse_R \quad (8)$$

Figure 3: (a) Associated Euclidean distance vs number of classes. The intersection of the regression line in red and the regression line in green determines the decision boundary for the number of classes. (b) the minimum value of the total mean square error of regression of the lines in red and in green gives the value of the class number. The number of classes in this study was estimated at 14.

After determining the number of classes, the calculated Euclidean distances are used to construct the different classes according to their similarity. The determination of the corresponding image was carried out manually using the time of detection of the insect by the measurement system.

Figure Fig. 4 shows the central spectra of the 14 classes of insects identified and Fig. 5 shows the image associated with the central spectrum of each of the classes.

Figure 4: Central spectral fingerprints of the 14 species of flying specimens detected by the measurement system. In each sub-figure, the class of each species of flying specimens is indicated by C1, C2,..., The number of flying specimens in each class is indicated by N. On the far right at the top of the top sub-figures the time at which the flying specimen was detected is indicated and in the lower figures this time is written at the bottom right. The shaded curves indicate the interquartile range within a group.

Figure 5: Raw images of flying specimens associated with different central spectral imprints. The captured specimens are identified on each sub-figure by a rectangle in orange. The images are saturated and do not identify the species detected. Certain fast flapping specimens have passed without being detected by the camera, for example in C2, C3, C10, etc. Others have passed far from the object focal plane. The image obtained in this case is blurred. This is the case of the images in C4, C6, C14, ..

6. Conclusion

We have implemented a device for synchronous acquisition of the image and the spectral imprint of flying specimens by passive remote sensing. We used the maximum noise fraction method to extract the spectra backscattered by the flying specimens in a set of 1000 samples containing the background noises. An unsupervised classification was used over a short measurement period to group the insect species detected by our system and thus identify each class of insect by an acquired image. As the camera used was not fast, we obtained saturated images. The simultaneous acquisition of the image and the spectral imprint being a major stake, this study can be deepened by using a fast camera.

Acknowledgments

The authors acknowledge Mikkel Brydegaard and Samuel Jansson for their technical support.

Funding Statement

General financial support from Swedish International Development Cooperation Agency (SIDA) through International Science Program (ISP), Uppsala University as well as the Third World Academy of Sciences (TWAS).

Data Accessibility

This article has no additional data.

Competing Interests

We have no competing interests.

References

1. Brydegaard M, Merdasa A, Gebru A, Jayaweera H, Svanberg S. 2016 Realistic Instrumentation Platform for Active and Passive Optical Remote Sensing. *Appl. Spectrosc.* (doi:10.1177/0003702815620564)
2. Runemark A, Wellenreuther M, Jayaweera HHE, Svanberg S, Brydegaard M. 2012 Rare events in remote dark-field spectroscopy: An ecological case study of insects. *IEEE J. Sel. Top. Quantum Electron.* (doi:10.1109/JSTQE.2012.2184528)
3. Brydegaard M, Jansson S. 2019 Advances in entomological laser radar. *J. Eng.* (doi:10.1049/joe.2019.0598)
4. Gebru A, Rohwer E, Neethling P, Brydegaard M. 2014 Investigation of atmospheric insect wing-beat frequencies and iridescence features using a multi-spectral kHz remote detection system. In *Remote Sensing and Modeling of Ecosystems for Sustainability XI*, (doi:10.1117/12.2060827)
5. Brydegaard M. 2015 Towards quantitative optical cross sections in entomological laser radar - Potential of temporal and spherical parameterizations for identifying atmospheric fauna. *PLoS One* (doi:10.1371/journal.pone.0135231)
6. Kirkeby C, Wellenreuther M, Brydegaard M. 2016 Observations of movement dynamics of flying insects using high resolution lidar. *Sci. Rep.* (doi:10.1038/srep29083)
7. Mei L, Brydegaard M. 2015 Atmospheric aerosol monitoring by an elastic Scheimpflug lidar system. *Opt. Express* (doi:10.1364/oe.23.0a1613)
8. Jansson S, Brydegaard M. 2018 Passive kHz lidar for the quantification of insect activity and dispersal. *Anim. Biotelemetry* (doi:10.1186/s40317-018-0151-5)
9. Malmqvist E, Jansson S, Torok S, Brydegaard M. 2016 Effective parameterization of laser radar observations of atmospheric fauna. *IEEE J. Sel. Top. Quantum Electron.* (doi:10.1109/JSTQE.2015.2506616)
10. Parra L, Sajda P. 2004 Blind source separation via generalized eigenvalue decomposition. In *Journal of Machine Learning Research*, (doi:10.1162/jmlr.2003.4.7-8.1261)
11. Pham DT, Cardoso JF. 2001 Blind separation of instantaneous mixtures of nonstationary sources. *IEEE Trans. Signal Process.* (doi:10.1109/78.942614)
12. Jutten C, Herault J. 1991 Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. *Signal Processing* (doi:10.1016/0165-1684(91)90079-X)
13. Hundley DR, Kirby MJ, Anderle M. 2002 Blind source separation using the maximum signal fraction approach. *Signal Processing* (doi:10.1016/s0165-1684(02)00342-0)
14. Sun L, Rieger J, Hinrichs H. 2009 Maximum noise fraction (MNF) transformation to remove ballistocardiographic artifacts in EEG signals recorded during fMRI scanning. *Neuroimage* (doi:10.1016/j.neuroimage.2009.01.059)
15. Zagouras A, Inman RH, Coimbra CFM. 2014 On the determination of coherent solar microclimates for utility planning and operations. *Sol. Energy* (doi:10.1016/j.solener.2014.01.021)
16. Zagouras A, Kazantzidis A, Nikitidou E, Argiriou AA. 2013 Determination of measuring sites for solar irradiance, based on cluster analysis of satellite-derived cloud estimations. *Sol. Energy* (doi:10.1016/j.solener.2013.08.005)
17. Kouakou BK, Jansson S, Brydegaard M, Zoueu JT. 2020 Entomological Scheimpflug lidar for estimating unique insect classes in-situ field test from Ivory Coast. *OSA Contin.* (doi:10.1364/osac.387727)