Infrared Physics & Technology 88 (2018) 200–205

Contents lists available at ScienceDirect

Infrared Physics & Technology

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

Regular article

Identification of blood species based on diffuse reflectance and transmission joint spectra with machine learning method

Hongxiao Li [a], Meixiu Sun [a,*], Zhiguang Xiang [b,1], Ling Lin [c], Chuan Qin [b], Yingxin Li [a]

[a] Institute of Biomedical Engineering, Chinese Academy of Medical Sciences & Peking Union Medical College, Tianjin 300192, China
[b] Institute of Laboratory Animal Sciences, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100021, China
[c] School of Precision Instrument and Opto-Electronics Engineering, Tianjin University, Tianjin 300072, China

HIGHLIGHTS

- Visible spectra more discrete than near-infrared for blood species identification.
- The joint spectra outperform the separate visible or near-infrared spectra.
- More information used to build the model, more accurate the predictions are.

ARTICLE INFO

Article history:
Received 25 July 2017
Revised 22 November 2017
Accepted 23 November 2017
Available online 24 November 2017

Keywords:
Blood species identification
Diffuse reflectance and transmission spectroscopy
Visible and near-infrared spectroscopy
Machine learning

ABSTRACT

The identification of blood species is of great importance for the forensic science and wildlife investigations. Our previous work has proved that the diffuse reflectance visible spectra and transmission near–infrared spectra are effective in differentiating the blood species non-contactedly. This paper compared these two spectra's abilities to classify the blood species, and proposed using these two spectra jointly to build the classification model. The machine learning method artificial neural network was the algorithm of model building. 1200 samples from five species were used to verify the recognition model. Results showed that the visible spectra outperform the near–infrared spectra, while the joint spectra perform better than the separate visible spectra. Therefore, the joint spectra of diffuse reflectance visible and transmission near–infrared provide more information of the blood components to distinguish the blood species more precisely.

1. Introduction

The identification of blood species is important in forensic science [1] and wildlife preservation. To achieve the aim of non-destructive measurement, Raman spectroscopy method was developed and proved to be an efficient technique to discriminate the blood species [2,3]. However, this method cannot detect a blood sample placed in a specific container, such as an anticoagulation tube, which is a common used container for blood transportation. Our group demonstrated that the diffuse reflectance visible spectra and transmission near–infrared (NIR) spectra are useful in identifying the species of blood samples placed in anticoagulation tubes [4,5], which could be very useful on screening blood samples for blood supervision.

The underlying reasoning of this method has two parts. Firstly, since Jöbsis [6] for the first time proposed the method of determining the blood components' contents by NIR spectroscopy in 1970s. Lots of studies on blood spectrum analysis were conducted to establish the prediction models between the blood spectra and the standard contents of blood components [7–10]. The latter is usually achieved by the biochemical methods. These previously published results indicate that the blood spectra can be used to predict the contents of the components in blood. Secondly, in the sexual creatures, the species is defined as a unified interbreeding group, consisting of populations with actually or potentially reproductive capacities. And they are reproductively isolated from other such groups [11]. Wang et al. [12] compared a number of blood physiological and biochemical indexes of cynomolgus monkeys and macaques, and found that there are significant differences in the blood components of these two groups. Ekser et al. [13] compared the hematologic parameters of pigs, baboons, cynomolgus monkeys, rhesus monkeys and humans. The published results indicate that the differences of blood components' contents among

* Corresponding author.
E-mail addresses: meixiu_sun@126.com (M. Sun), xiangzg@cnilas.org (Z. Xiang).
[1] Co-corresponding author.

https://doi.org/10.1016/j.infrared.2017.11.030
1350-4495/

different species are significant. Therefore, it is logically possible to predict the blood components' contents firstly, and then discriminate the blood species based on the predicted blood components' contents.

However, to logically infer a rule to describe the elements in the collection, we must have information about each element in the collection [14]. Which means we need to collect all the information of each blood component. This would be difficult in practice. Therefore, the machine learning methods are proposed to avoid this problem by offering only probabilistic laws, rather than the entirely certain laws used in purely logical reasoning. In this paper, the well-known classifier algorithm artificial neural network (ANN) was used to construct the spectra based blood species recognition models. The diffuse reflectance visible and transmission NIR spectra were jointly used to optimize the performance of the classification. A dataset consisting of 1200 samples was used to verify the proposed method.

2. Materials and methods

2.1. Blood samples

1200 blood samples of dog, goat, rhesus monkey, rat and human were collected by Institute of Laboratory Animal Sciences, CAMS & PUMC and refrigerated to deliver to the spectral detection laboratory. Each species has 240 blood samples. Each blood sample was provided by an individual donor with about 5 mL, and placed in a Greiner VACUETTE vacuum blood collection tube. Ethylene diamine detraacetic acid (EDTA) was added into the blood sample to prevent blood clotting. The blood samples were measured in their original containers. Each spectral data was obtained by averaging over five consecutive measurements. The human volunteers had given their consents for the experiments. All experiments performed were in compliance with the local laws and the guidelines of related institutions. All the institutes mentioned above had approved the experiments.

2.2. Measurement system

The measurement system includes four parts, as shown in Fig. 1. The super continuum source from Wuhan Yangtze Soton Laser has a spectral range of 450–2400 nm. The visible and NIR spectrometers from Avantes China have detection ranges of 294–1160 nm and 1021–1757 nm, respectively. The optical black box is a container for the fixture of the sample tube and the electric displacement platform (EDP), as shown in Fig. 2. During the measurement, the optical black box was closed to prevent the external stray light interfering the detection. Finally, the control system installed on a laptop is in charge of the manipulation of the devices above mentioned.

2.3. Data collection

The power of the light source is 800 mW. In our experiment, the output ratio is tuned to 30%, thus the output power is 240 mW. The process of the measurement is shown in Fig. 2. The fiber probe 1 collects the backscattering visible spectrum, and the fiber probe 2 collects the forward scattering NIR spectrum. To improve the performance, the NIR spectrum was collected from 10 vertically scanned equidistant sites [15]. The fiber probe 2 was moved by an EDP. The EDP's precision of movement is 1 μm. The distance of every two adjacent sites of the vertical move is 0.381 mm. In total eleven spectra collected from each blood sample, there are one visible spectrum and ten NIR spectra.

2.4. Data processing

All processing and statistical analysis of the data are performed with MATLAB 8.2.0. The spectral data were preprocessed before being used to build the recognition model. Firstly, the original spectral data were cleansed to eliminate the missing values. Then, the cleansed data were normalized. The averaged spectra of five species are shown in Fig. 3. To jointly utilize the visible and NIR spectra, the visible spectral data and the ten NIR spectra were

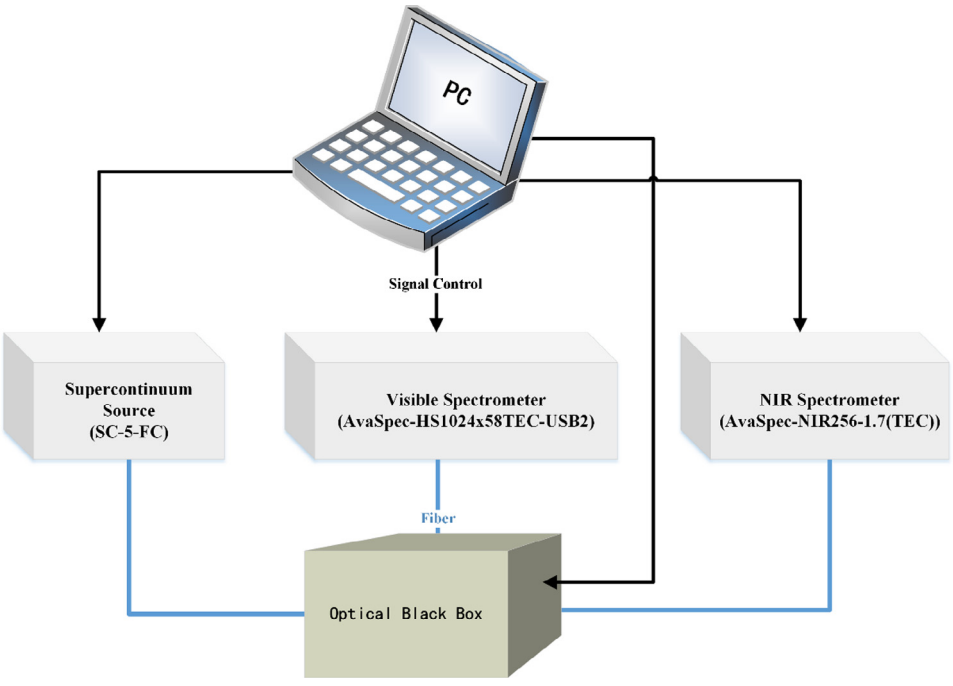

Fig. 1. The schematic diagram of the measurement system.