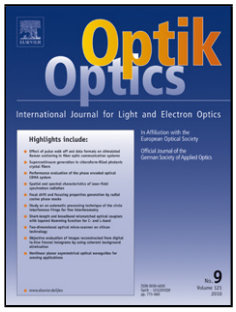

Title: Identification terahertz spectra for the dyestuffs based on principal component analysis and Savitzky-Golay filter

Authors: Haishun Liu, Zhenwei Zhang, Yuping Yang, Tao Chen, Cunlin Zhang

PII: S0030-4026(18)31051-9
DOI: https://doi.org/10.1016/j.ijleo.2018.07.079
Reference: IJLEO 61235

To appear in:

Received date: 18-5-2018
Accepted date: 20-7-2018

Please cite this article as: Liu H, Zhang Z, Yang Y, Chen T, Zhang C, Identification terahertz spectra for the dyestuffs based on principal component analysis and Savitzky-Golay filter, *Optik* (2018), https://doi.org/10.1016/j.ijleo.2018.07.079

Identification terahertz spectra for the dyestuffs based on principal component analysis and Savitzky-Golay filter

Haishun Liu[a], Zhenwei Zhang[a,*], Yuping Yang[b], Tao Chen[c], Cunlin Zhang[a]

[a]Beijing Key Laboratory for Terahertz Spectroscopy and Imaging, Key Laboratory of Terahertz Optoelectronics, Ministry of Education, and Beijing Advanced Innovation Center for Imaging Technology, Department of Physics, Capital Normal University, Beijing 100048,China

[b]School of Science, Minzu University of China, Beijing 100081, China

[c]School of Electronic Engineering and Automation, Guilin University of Electronic Technology, Guilin, Guangxi 541004, China

ARTICLE INFO

Keywords: Terahertz spectra
Savitzky-Golay smoothing
Principal component analysis
Clustering

ABSTRACT

Terahertz time-domain spectroscopy (THz-TDS) was employed to measure the absorption spectra for three types of dyestuff at the frequency range of 0.3~2.2THz. The raw spectra data were implemented dimensionality reduction by using principal component analysis (PCA). Depending on the weak cluster trend determined by the score plot, different levels of Savitzky-Golay (SG) smoothing combined with PCA processing was performed for elevating the recognition rate. The recognition effects were assessed by fuzzy c-means (FCM) and k-means clustering techniques, which consistently demonstrated the combination of polynomial order 1 and window size 5 for SG smoothing achieved the highest accuracy of 94.44%. For k-means and FCM clustering, the identification accuracies of raw spectra were 87.5% and 84.72% respectively, realizing the elevation recognition rate by using SG smoothing with polynomial order 1 and window size 5. Our results suggested SG smoothing coupled with PCA was a potent method to cope with THz spectra recognition for dyestuffs.

1. Introduction

Owing to the invention of synthetic dyestuffs, dyeing has become an indispensable step in making cultural relics, medicine, food, and textiles. For the precious cultural relics, it is necessary to maintain the hue to preserve the historical value by using the tiptop dyestuffs [1]. Recently, it is reported that superfluous poisonous dyestuffs were added into Chinese medicinal materials and decoction pieces by the greedy manufacturers, which creates a dangerous situation for patients. The similar issue exits in food and textiles as well. So it is urgent to screen the hazardous substances to guarantee the public health. However, some imperfection exists in the current detection methods. UV-visible and fluorescence spectra could hardly procure the information of complex substances. Raman and infrared spectra could solely get the vibrational spectra of functional group [2]. The vibrational modes of most biomolecules and weak intermolecular force lie within terahertz (THz) spectrum (0.1-10THz), so THz spectroscopy can be served as a potent technique to identify the different

* Corresponding author.
E-mail address: zhangzw@cnu.edu.cn