Sensors and Actuators B 268 (2018) 299–309

Contents lists available at ScienceDirect

Sensors and Actuators B: Chemical

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

High-speed sex identification and sorting of living silkworm pupae using near-infrared spectroscopy combined with chemometrics

Zhiqiang Zhu[a], Hongfu Yuan[a,*], Chunfeng Song[a,*], Xiaoyu Li[a], Dewen Fang[b], Zhaoyi Guo[b], Xinye Zhu[c], Wei Liu[c], Gendong Yan[b]

[a] College of Materials Science and Engineering, Beijing University of Chemical Technology, Beijing 100029, China
[b] East-wide Silkworm Group, Shandong 262500, China
[c] CSEPAT (Beijing) Technology Co., Ltd., Beijing 100029, China

ARTICLE INFO

Article history:
Received 3 October 2017
Received in revised form 18 April 2018
Accepted 19 April 2018
Available online 22 April 2018

Keywords:
Sex sorting
Silkworm chrysalis
Near infrared spectroscopy
Pattern identification
SIMCA

ABSTRACT

An identification method of near infrared spectroscopy in combination with multivariate analysis and a high-speed automated sorting system for living silkworm chrysalis were developed. First, the factors affecting sex discrimination were examined using static spectra, including the varieties and positions of pupae. Second, modeling using dynamic spectra was investigated. The dynamic spectra contain much more noise than that of static spectra. The significant noise dominated by optical path difference (OPD) has a strong impact on the identification of male and female pupae. High-speed sex discrimination of living silkworm pupae was successfully realized by combining soft independent modeling of class analogy (SIMCA) and a preprocessing method involving angle spectra. The results indicated that this method achieved a correct identification rate of 98.0%. Finally, various varieties of living silkworm pupae were sorted with respect to sex using the high-speed sorting device. The sorting rate exceeded 7.7 pupae per second and the error rate could be controlled within 2.5%. The device has been applied successfully to sort living silkworm pupae about 1.2 tons per day in different seasons and regions, and can exert a significant effect on the development of modern sericulture.

1. Introduction

Sericulture is a unique industry that has not only made a great contribution to the civilization of the ancient world, but also become a model industry in the contemporary era with regard to its high utilization of resources. Silk is a high-quality natural protein fiber. Silkworm chrysalises and silkworm moths have been used as raw materials for developing new foods. Because of its high chlorophyll content, silkworm excrement is used for producing chlorophyllin copper, which is a raw material for pharmaceuticals. China is the world's oldest silkworm cocoon producer. The Silk Road, an ancient network of trade routes between the Mediterranean and China, made substantial contributions to the economic development and cultural prosperity of the ancient world. China remains the world's largest producer and exporter of mulberry silk. Nevertheless, in recent years, the sericulture industry has grown rapidly in countries including India, Myanmar, Vietnam, and Kazakhstan. Thus, the sericulture industry continues to have a significant influence on the global economy and culture.

High-quality silkworm eggs are very important to the development of the sericulture industry. In the modern sericulture industry, hybridization is used to enhance the yield and quality of silk. Crossbreeding is a key process in sericulture, and this process entails sorting silkworm chrysalises into male and female pupae. To date, sorting work has been performed manually through visual inspection of pupa organs due to the lack of automatic techniques for sex sorting. Silkworm chrysalises develop into silkworm moths in just over 1 week, which creates a very heavy workload for large-scale producers. The Chinese firm East-wide Silkworm Group is the largest producer of silkworm eggs, with more than 2000 highly skilled and well-trained workers who sort approximately 8 tons of pupae per day in production seasons. Sex identification through visual inspection of pupa organs is not only inefficient but also has a high sorting error rate. Additionally, the workforce is aging and recruiting new workers is difficult because the sorting of pupa sex is temporary seasonal work and labor resources are limited. Therefore, labor shortage and the rising cost of hiring have become a

* Corresponding author.
E-mail addresses: zhu5319@126.com (Z. Zhu), hfyuan@mail.buct.edu.cn (H. Yuan), cfsong@mail.buct.edu.cn (C. Song), lixy@mail.buct.edu.cn (X. Li), gtfdw@126.com (D. Fang), gtjtzjs@126.com (Z. Guo), xinye_zhu@sinosepat.com (X. Zhu), wei_liu@sinosepat.com (W. Liu), sdygd001@163.com (G. Yan).

https://doi.org/10.1016/j.snb.2018.04.093
0925-4005/

serious challenge to the development of the sericulture industry. Clearly, the male/female pupa sorting technique has been become a bottleneck. Therefore, the research and development of new sorting methods and high-speed sorting systems for live silkworm chrysalises are crucial for silkworm crossbreeding.

Research on mechanizing and automating the sex separation of silkworm pupae has been conducted for more than 50 years. Several sex-separation techniques have been proposed, including DNA analysis [1–3], mass and shape analysis [4–6], magnetic resonance imaging (MRI) [7], X-ray imaging [8], and spectral imaging [9–12]. Although DNA analysis is extremely accurate, this method is time-consuming and destructive, and its application is limited to laboratory genetic studies. Hu et al. found through statistical analysis that the weight and size of female chrysalises are significantly larger than those of male ones [6]. A sorting machine for determining the sex of chrysalises according to weight and size was designed. However, the size and weight of silkworm chrysalis of the same sex can change substantially depending on factors such as species, breeding, growing conditions, and season. The machine is unable to adapt to these changes and has not been applied in practice. Cai et al. [8]. applied X-ray imaging technology combined with chemometric methods to cocoons for real-time sex identification. Four types of algorithm, namely the K nearest neighbors, linear discriminant analysis (LDA), back-propagation artificial neural network, and support vector machine algorithms, were compared; LDA was found to be the best algorithm. Liu C et al. [7] proposed a method in which T2-weighted imaging of MRI is used for sex identification of silkworm cocoons. However, the disadvantages of X-ray imaging and MRI methods include complicated operation, expense, and causing mutation of pupae. Kamtongdee [9,12] found that green or red light can pass through the body of a silkworm chrysalis, but blue and near-infrared (NIR) light cannot. In particular, red light can clearly illuminate an important organ of females called the “chitin gland”, enabling the sex of silkworms to be identified with high accuracy. On the basis of this finding, a sorting machine was designed that achieves a speed of 30 silkworm chrysalises per minute and has an accuracy exceeding 95%. However, the low sorting efficiency of this system is such that it cannot meet the demands of high sorting speed in actual silkworm chrysalis processing.

By virtue of being nondestructive and rapid, NIR analysis has been widely applied to agriculture [13,14], materials [15–17], pharmaceutical [18], petrochemical [19], life sciences [20] and wildlife and biodiversity [21]. NIR diffuse reflectance spectroscopy combined with derivative and Bayesian classification has been used to identify the sex of silkworm pupae and cocoons in laboratory settings [22,23]. The results showed very promising accuracy levels of 98.7% and 95.7% for identifying the sex of silkworm chrysalises and cocoons, respectively. However, this identification accuracy was obtained under precisely designed measuring conditions in which the posture of a chrysalis and the position measured must be artificially controlled in laboratory. Moreover, collecting the average spectrum of several spectra to improve measurement precision is a time-consuming process. Obtaining a high-quality spectrum in a sufficiently short time for the method to be economically feasible is extremely challenging. Although this method is not applicable in practice, the research results reveal that NIR spectroscopy has potential use in realizing the high-speed sorting of silkworm pupae. However, numerous challenges remain for achieving high-speed sorting of silkworm chrysalises, most of which concern measurement precision and pretreatment methods regarding spectra.

Two measuring modes can be applied for determining the sex of silkworm chrysalises using NIR, namely transmission and diffuse reflectance. For the transmission mode, light passing through a silkworm chrysalis must have a sufficiently high power source, which may harm the living silkworm chrysalis. In practice, the diffuse reflectance mode is more suitable because light need not penetrate the silkworm pupa and a weak light source can meet the requirements, thereby avoiding harm to the silkworm pupa. However, the positions of the collecting probe and light source are fixed when using the diffuse reflectance mode. Consequently, when pupae of different sizes pass through the measuring spot, the distance between the end surface of the collecting probe and the body surface of the living silkworm chrysalis changes substantially. The change of distance is referred to as the optical path difference (OPD). This results in poor spectral measurement precision [24,25]. Moreover, achieving a spectrum of high quality in a sufficiently short time for the cost to be feasible is very difficult.

Compared with the static spectra of living silkworm chrysalises collected using bench-top instruments under precisely designed measuring conditions, the dynamic spectra collected online during high-speed sorting contain more noise, which has substantial adverse effects on the identification of male and female pupae. In general, those effects are eliminated or reduced by spectral pretreatment. Therefore, spectral signal pretreating methods are crucial for reducing noise and thereby realize high-speed sorting. Commonly used methods for diffuse reflectance spectra include standard normal variate (SNV) and multivariate scatter correction (MSC), as well as derivative and centralization methods [26–33]. Both SNV and MSC can eliminate the spectrum scattering effect [29]. SNV assumes that all wavelengths of the spectrum have the same standard deviation [32]. The prerequisites of MSC are that information concerning drift and spectrum scattering is considerably greater than that regarding component information or chemical properties [31]. The derivative method [33] can eliminate spectral shifting: The first-order derivative (1′st) eliminates the linear shift independent of the spectrum, and the second derivative (2′nd) does the wavelength-dependent linear shift [34]. These methods are widely used in NIR quantitative and qualitative techniques because of their unique advantages under certain conditions. However, for the dynamic spectra of silkworm chrysalises collected during high-speed sorting, these methods are not capable of eliminating the impact on identification caused by the OPD.

In this study, a high-speed sorting system was designed using NIR and the spectra were transformed into angle vectors to reduce the error produced by the OPD. A new identification method combining angle vectors with the soft independent modeling of class analogy (SIMCA) method was proposed and identification models of living silkworm pupae were constructed. The system has been successfully applied in production to realize the high-speed sorting of male and female pupae with an accuracy of more than 97.5% at a speed of 7.7 pupae per second.

2. Theory

The Beer–Lambert law and Kubelka–Munk equation [35] are theories of spectral analysis that indicate that absorbance and diffuse reflectance change linearly with the concentration of the component under investigation, and they are used to conduct quantitative and qualitative analysis. The problem is that the absorbance or reflectance intensity is affected by variations in the light path length, which are often caused by irregular shapes of solid samples or changes in the density of liquids at different temperatures. In this paper, the method for eliminating the effect of the OPD is proposed as follows.

The spectrum of a mixture containing n compounds can be represented by the linear addition of the spectra of all the components:

$$\boldsymbol{d} = \sum_{i=1}^{n} \boldsymbol{s_i} \boldsymbol{c_i} \quad (1)$$

where $\boldsymbol{d}$ is the spectrum of the mixture, and $\boldsymbol{s}_i$ and $\boldsymbol{c}_i$ are the spectrum and concentration of the ith component, respectively.