

文章编号: 1005-5630(2023)06-0001-07

DOI: 10.3969/j.issn.1005-5630.202301300006

利用数据库技术的近红外消光法颗粒粒度测量

王鸿钦, 郑泽希, 项华中, 李祥鹏

(上海理工大学上海市介入医疗器械工程研究中心, 上海 200093)

摘要: 为了实现在近红外波段测量介质中颗粒的粒径, 提出了一种基于消光光谱结合数据库技术的颗粒粒径反演方法。首先通过全散射消光的 Lambert-Beer 定律计算生成与颗粒粒径大小及分布相关的消光光谱数据库, 然后用数据库处理算法比较数据库中数据与实测颗粒消光光谱数据, 两者最佳吻合所对应的数据库中颗粒粒度及分布参数即为所求值。分别对颗粒的单峰和双峰粒度分布进行数值仿真, 仿真结果的体积频率分布的相对均方根误差小于 10%。利用 3 种不同粒度的国家标准颗粒进行实验验证, 实测得到粒径 D50 的相对误差均在 10% 以内, 仿真及实验对所提算法的验证得到了满意的结果。研究表明, 将数据库技术及算法应用于近红外窄波段消光法颗粒粒度的测量是可行的和有效的。

关键词: 消光法; 光谱; 数据库; 近红外; 粒径测量

中图分类号: O 433 **文献标志码:** A

Particle size measurement by near-infrared light extinction method using database technologies

WANG Hongqin, ZHENG Zexi, XIANG Huazhong, LI Xiangpeng

(Shanghai Engineering Research Center of Interventional Medical Device, University of Shanghai for Science and Technology, Shanghai 200093, China)

Abstract: In order to measure the particle size in the medium using near-infrared light source, a particle size measurement method based on light extinction method combined with database technology is proposed. The extinction spectrum database related to different particle size and distribution is first calculated and generated by the total scattering extinction Lambert-Beer law. The database processing algorithm is employed to compare the data in the database with the measured particle extinction spectrum data. Then the particle size and distribution parameters of the measured samples are determined from the calculated values of the best match in the database. Next, numerical simulations of both the unimodal and bimodal particle size distribution are carried out, and the relative root mean square error of the volume frequency distribution of the simulation results is less than 10%. Three kinds of national standard particles with different particle sizes are

收稿日期: 2023-01-30

基金项目: 国家自然科学基金青年科学基金(52206102)

第一作者: 王鸿钦(1998—), 男, 硕士研究生, 研究方向为生物医学工程。E-mail: wanghongqin66@163.com

通信作者: 郑泽希(1989—), 男, 博士后助理研究员, 研究方向为超精密加工与测试和微纳机械。E-mail: zexizheng@outlook.com

used for experimental verification, and the relative errors of the measured particle size D50 are all less than 10%. Thus the validity of the proposed method and algorithm is demonstrated through simulations and experiments with satisfactory results. The research in this paper shows that the application of database technology and algorithm to the measurement of particle size by near-infrared narrow-band extinction method is feasible and effective.

Keywords: light extinction method; spectrum; database; near-infrared; particle size measurement

引 言

颗粒粒径的测量在环境监测^[1]、生物医学^[2]、制药、化工^[3]等众多领域有着重要的意义。光散射法的先进性促成了其在粒径测量中的广泛使用^[4]，而消光法的原理及测量相较于其他光散射法更为简单^[5]，对于目前使用较多的粒度在线测定具有较大的应用空间。消光法利用经颗粒散射后的连续透射光谱或离散多波长光谱进行反演运算 j 进而求得颗粒粒度与尺寸分布^[6-7]。

求颗粒粒径的反演方法一般有两类。一类为独立模式法，即不使用某种函数关系来近似描述颗粒粒径分布，而直接通过求解相关离散方程计算出粒径。另一类为非独立模式反演算法，假设所求的粒径分布符合某个函数规律^[8]，然后利用优化算法比较理论计算值和实测值，搜寻出最佳分布。非独立模式法涉及更少的自变量，其运算复杂度较独立模式法更低。国内外对颗粒粒度反演的各种算法进行了诸多探索。孙晓刚等^[9]利用全局性的遗传算法同时按照三类分布函数来反演，获取了粒径近似分布规律的同时也避免了搜寻时困于局部极值的情况。赵蓉等^[10]通过非线性最小二乘法反演小于 10 μm 的粒径，再通过图像法反演超过 10 μm 的粒径，达到了测量跨微米混合粒径的目的。Ping 等^[11]用主成分分析探究消光数据与波长间关系以筛选出最佳波长，并结合 BP 神经网络对粒径分布进行计算，以提高计算收敛的速度和计算结果的稳定性。然而，在采用近红外窄波段光源的消光法粒径测量中，仍较难找到一种粒径反演准确性较好的算法。李祥鹏等^[12]采用近红外窄带光源，通过人工蜂群算法对透射消光光谱反演粒径，取得一定效

果。本文则引入新颖的数据库技术，相较于传统的反演算法，该方法可预先计算及建立仪器测量范围内的含颗粒粒度及尺寸分布的数据库，并利用光谱特征结合数据库高效的管理与查询能力来减小搜寻范围及复杂度，实现较高效率的全局搜寻，同时降低易困于局部极值的风险。

目前的社会生活与大数据的联系愈发紧密^[13-14]，数据量的剧增促使数据库技术的兴起与成熟。数据库技术是一种能够对已有的海量数据进行有效存储、管理及分析的技术手段，已在模式(指纹、人脸等)识别等众多领域得到广泛应用^[15-17]。本实验采用在医学无创检验如血氧饱和度、血糖等测量中^[18-19]使用的近红外窄波段光源。相比消光法颗粒粒度测量中常用的可见光宽波段光源，近红外窄波段的消光光谱法反演难度更大，但本文提出的基于数据库技术的求颗粒粒径的反演算法取得了较为满意的实验结果。因为数据库是预先建立的，在实测时本文方法的反演速度较快，精准度高，且对双峰分布也同样有较高的精度。

1 颗粒粒度的求解方法

1.1 消光法

本文以散射消光的 Lambert-Beer 定律为基础。设一平行光束的波长为 λ 、光强为 I_0 ，使它穿过待测区域。待测区域里悬浮在介质中的颗粒会散射光束，则经过待测区后的光强 I 会被削弱^[20]。当介质内悬浮物为多分散粒子群时，出射光强与入射光强的关系表示为^[9]

$$\ln(I/I_0)_\lambda = Q \int_{D_{\min}}^{D_{\max}} K_{\text{ext}}(\lambda, m, D) f(D) / D dD \quad (1)$$

式中: I/I_0 为反映光强削弱度的消光值; λ 、 m 及 D 分别为波长、相对折射率及粒子尺寸; K_{ext} 为消光系数, 利用 Mie 理论以及函数关系算得^[21]; 常量 $Q = -3LN_D/2$, L 为光束穿过待测区时经过的光程, N_D 为待测区所含颗粒总和; $f(D)$ 为待求的描述不同粒径所占体积频度比例的颗粒尺寸分布函数^[9]。式(1)属于 Fredholm 第一类方程, 通常依靠数值积分法求解, 经此法离散后得^[9]

$$E = Bf_B \quad (2)$$

式中: $E = [\ln(I/I_0)_{\lambda_1}, \ln(I/I_0)_{\lambda_2}, \dots, \ln(I/I_0)_{\lambda_l}]^T$, 选用 l 个波长采样; $B = [B_{ij}]$, $B_{ij} = Qc_j K_{\text{ext}}(\lambda_i, m, D_j)/D_j$, c_j 代表数值积分系数, $i = 1, 2, \dots, l$, $j = 1, 2, \dots, M$, M 及 D_j 分别代表颗粒大小在 $[D_{\min}, D_{\max}]$ 区间的粒径划为 M 档及各档对应的等效直径; $f_B = [f(D_1), f(D_2), \dots, f(D_M)]^T$ 。依据前述展开式(2)得

$$\ln(I/I_0)_{\lambda_i} = Q \sum_{j=1}^M c_j K_{\text{ext}}(\lambda_i, m, D_j) f(D_j) / D_j \quad (3)$$

出射光强 I 可由式(3)进行变换得^[12]

$$I_{\lambda_i} = I_{0\lambda_i} \exp \left(Q \sum_{j=1}^M c_j K_{\text{ext}}(\lambda_i, m, D_j) f(D_j) / D_j \right) \quad (4)$$

通过式(3)和式(4)可以发现, 不同的 $f(D_j)$ 对应不同的一组消光值 $\ln(I/I_0)_{\lambda_i}$ 或透射光强 I_{λ_i} , 在间隔小、数量多的不同波长 λ 下多个消光值或透射光强可看作光谱曲线。因此, 不同的 $f(D_j)$ 对应不同的光谱曲线, 众多的光谱曲线构成一个数据库。通过数据库算法找出与实测光谱曲线最佳吻合的理论光谱曲线, 相应的 $f(D_j)$ 便是所求的最优颗粒粒径分布解。

1.2 数据库的构建

数据库存储 ID 和相应的分布函数、光谱数据、特征数据这 4 部分内容, ID 为标识每条数据的编号。

根据设定的尺寸分布和式(3)或式(4), 在 MATLAB 中算出对应的由消光值或透射光强构成的光谱曲线, 接着算出在光谱曲线上的几个特征值(如斜率、距离、均值等)。连接数据库软件 PostgreSQL, 将上述尺寸分布及光谱数据、

特征数据存入 PostgreSQL。PostgreSQL 是目前所有开源数据库中功能最为强大的数据库管理软件, 可支持不同语言自行编程如计算欧氏距离等程序。

特征数据用于缩小计算范围, 减少计算时间, 否则实测光谱数据将不得不与库中海量光谱数据逐一运算比较。光谱具有形状、数据等方面的特征, 其中数据方面主要有描述离散度、衡量集中度及表征变化趋势等方面的特征。由于实验所用的近红外波段较窄, 每组光谱数值之间差别小, 光谱离散性小且光谱间形状差异小, 光谱对离散程度及形状方面的特征不敏感, 因此本文从数据集中程度和变化趋势两个区分度更显著的方面选取平均值、中位数及斜率作为光谱曲线的特征参数。其中, 斜率特征 S 设定为

$$S = \sum_{i=1}^l E_i / \lambda_i \quad (5)$$

式中: E_i 是每条光谱数据; λ_i 为各光谱数据对应的波长。特征数据之间一般都存在较大甚至不同数量级的差异, 难以用一个确定的阈值找出库中与实测光谱特征最相似的光谱曲线。所以本文采用对特征相似度 T 进行排名的方法并提取排名靠前的数据。特征相似度 T 设定为

$$T = |(t_1 - t_2) / t_1| \quad (6)$$

式中 t_1 和 t_2 分别代表实测的光谱和理论算得的光谱的特征数据。

1.3 数据库的运行

在 PostgreSQL 中, 输入实测的光谱数据、特征数据。数据库分别按照 3 种特征相似度 T 的升序排名重新排列所有数据, 并分别提取在设定的排名指标内的数据集合 a_x , $x \in \{1, 2, 3\}$ 。得到数据集合 a_x 后, 提取 3 个集合中 ID 的交集及 ID 后面的数据, 得到数据集合 b , 则 b 中光谱满足各个特征都与实测光谱特征最相似。若出现 ID 交集为空的情况, 则逐步倍数扩大排名指标, 重新进行上述步骤。最后对 b 所有数据按照每条光谱数据与实测光谱数据的目标函数值的升序进行重新排列, b 中第一条数据的分布参数便是所求的尺寸参数。目标函数为^[10]

$$F = \sum_{i=1}^l (E_{1,i} - E_{2,i})^2 \quad (7)$$

式中： E_1 为实测的光谱数据； E_2 为理论计算的光谱数据； l 为波长数量。

由于排序时大量数据无法一次性放进内存，故使用多路归并排序(multipath merge sort, MMS)，其将待排序的数据集在磁盘中分为多个体积小于内存的子集合，对每个子集合进行快速排序(quick sort, QS)并将各自最小的数存入内存，对内存的数进行快速排序并将最小的数存入磁盘，接着将此数所在的有序子集合的下一个数替补到内存，继续将内存中最小的数存入磁盘，重复以上步骤直到排完数据集。图 1 为排序的流程图。

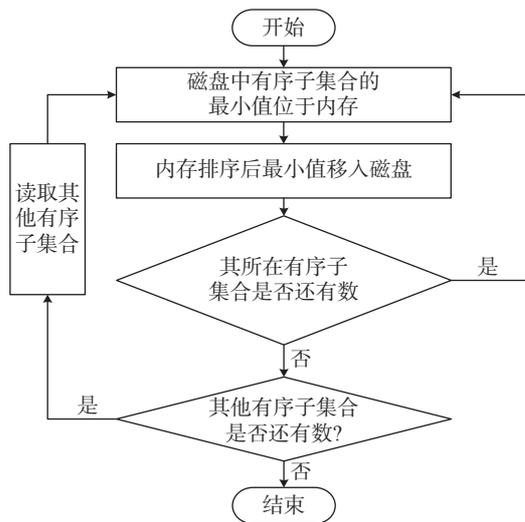


图 1 排序的流程图

Fig. 1 Flow chart of sorting

2 仿真和实验

2.1 仿真

近红外光源中心波长为 850 nm，取波段 835~865 nm，颗粒相对于水的折射率 m 为 1.591/1.33，粒径范围为 0.1~30 μm 。由于实际的大多数颗粒粒径分布满足 Rosin-Rammler 分布^[8]，本文使用 R - R 函数作为 $f(D)$ 。其中单峰分布为

$$f(D) = \left(\frac{D}{\bar{D}}\right)^{k-1} \exp\left[-\left(\frac{D}{\bar{D}}\right)^k\right] k/D \quad (8)$$

双峰分布设为

$$f(D) = n f_1(D) + (1-n) f_2(D) \quad (9)$$

式中： \bar{D} 和 k 分别为反映粒径大小和分布宽窄的两个参数； $n \in [0, 1]$ ^[10]。用相对均方根误差(relative root mean square error, RRMSE)来评价反演结果，其表达式为

$$R = \sqrt{\frac{\sum_{i=1}^M [f_{in}(D_i) - f_{out}(D_i)]^2}{\sum_{i=1}^M [f_{in}(D_i)]^2}} \quad (10)$$

式中： $f_{in}(D)$ 为理论上的分布； $f_{out}(D)$ 为反演得到的分布； R 为相对均方根误差。数据库中单峰分布参数区间设定为： $\bar{D} \in [1, 30]$ ， $k \in [1, 15]$ ，步长均为 0.01，库中约 400 万条数据。双峰分布的几个参数也作类似设定。排名指标设定为前 2.5%。

表 1 为单峰分布仿真结果，表 2 为双峰分布仿真结果。

表 1 单峰分布仿真结果

Tab. 1 Simulation results of unimodal distribution

设定值 (\bar{D}, k)	反演值 (\bar{D}, k)	R
(8.155, 9.349)	(8.16, 9.28)	0.0076
(12.734, 3.352)	(12.9, 3.24)	0.0412

表 2 双峰分布仿真结果

Tab. 2 Simulation results of bimodal distribution

设定值 $(\bar{D}_1, k_1, n, \bar{D}_2, k_2)$	反演值 $(\bar{D}_1, k_1, n, \bar{D}_2, k_2)$	RRMSE
(3.358, 11.71, 0.5, 10.242, 11.452)	(3.36, 11.65, 0.5, 10.24, 11.50)	0.0021
(3.164, 11.52, 0.7, 10.195, 11.853)	(3.16, 11.85, 0.7, 10.14, 11.75)	0.0638

单峰和双峰分布仿真结果的误差都小于 10%，算法的仿真效果较为令人满意。

为研究本方法对噪声的抗干扰性，对上述分布参数的光谱数据多次加入不同大小的随机噪声，以 10 次运算结果的平均值作为反演值，表 3 为仿真结果，图 2 为体积频率分布曲线。

表 3 加噪后仿真结果
Tab. 3 Simulation results after adding noise

设定值	随机噪声	反演值	R
(8.155, 9.349)	1%	(8.199, 9.242)	0.035 1
	3%	(8.24, 9.484)	0.064 6
(3.358, 11.71, 0.5, 10.242, 11.452)	1%	(3.388 0, 11.310 0, 0.500 0, 10.144 0, 11.265 0)	0.040 7
	3%	(3.342 0, 11.650 0, 0.500 0, 10.390 0, 11.795 0)	0.061 2

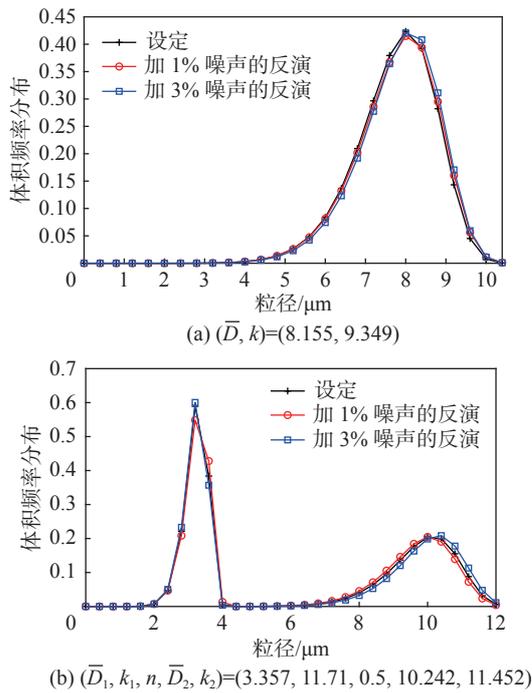


图 2 仿真结果的体积频率分布曲线
Fig. 2 Volume frequency distribution curve of simulation results

单峰和双峰分布仿真结果的误差随着随机噪声的增加而增加, 但误差仍在 10% 内, 说明算法对噪声具有一定的抗干扰能力。

2.2 标准颗粒实验

使用编号为 GBW120134、GBW120024、GBW120041 的国家标准颗粒进行实验来验证算法的有效性。

三者的 D50 粒径标称值为 3.44 μm、9.40 μm、14.62 μm, 实验的参数条件设置同仿真部分。

进行单峰分布实验时, 分别测量 3 种颗粒的入射光强和透射光强, 按照第 1.2 节建立相关数据库, 输入实测的光谱数据、特征数据进行运算, 表 4 为实验结果。从表 4 可以看出, 反演实

表 4 标准颗粒单峰分布实验结果
Tab. 4 Experimental results of unimodal distribution of standard particles

样本编号	D50 标称值/μm	D50 实测值/μm	相对误差
GBW120134	3.44	3.718	0.080 8
GBW120024	9.40	9.661	0.027 8
GBW120041	14.62	15.68	0.072 5

测值与 D50 标称值的相对误差均在 10% 以内。

进行双峰分布实验时, 将 D50 粒径分别为 3.44 μm、9.40 μm 的颗粒混合于蒸馏水中, 后续实验过程同前述单峰测量部分, 表 5 为实验结果。图 3 分别给出了部分实测的单峰及双峰分布颗粒的体积频率分布曲线。从表 5 可看出, 标准颗粒 D50 标称值与实测值的相对误差均在 10% 内。

表 5 标准颗粒双峰分布实验结果
Tab. 5 Experimental results of bimodal distribution of standard particles

标称值 D50/μm	实测值 D50/μm	相对误差
(3.44, 9.40)	(3.201, 10.180)	(0.069 5, 0.083 0)

3 结 论

本文提出了一种基于数据库技术来实现近红外窄波段反演颗粒尺寸的计算方法。预先建库存放与颗粒尺寸及分布参数相关的光谱数据, 并结合光谱特征与数据库高效的管理与查询能力来减少复杂运算、提升运算效率并降低易困于局部极值的风险, 做到较准确的全局搜寻。本文对提出的算法进行了数值仿真和标准颗粒的实测, 测量准确度均在 10% 内, 取得了满意的结果, 从而

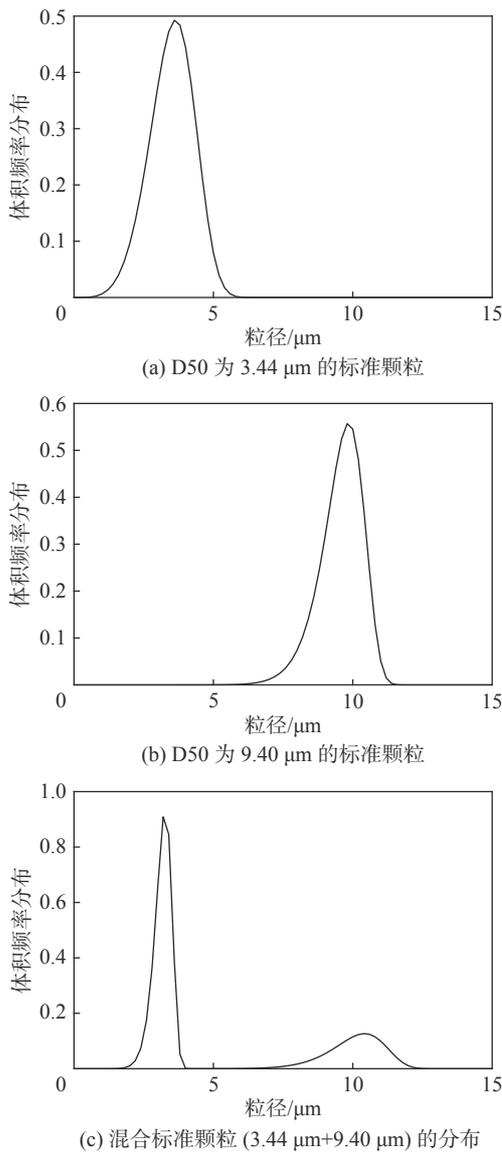


图 3 实验结果的体积频率分布曲线

Fig. 3 Volume frequency distribution curve of experimental results

验证了将数据库技术用于近红外窄波段消光法的颗粒测量是可行的和有效的,也给生物医学中利用近红外光进行体外以及体内的无创血糖检测等提供了一个新的思路。

参考文献:

- [1] 关泽, 范晨阳, 王站成, 等. 柴油发动机超细颗粒物数量检测方法研究进展[J]. 仪器仪表学报, 2022, 43(7): 1-16.
- [2] 严心涛, 王策, 王耀, 等. 空气辅助雾化的细胞递送参数的优化[J]. 光学精密工程, 2022, 30(14): 1725-1737.
- [3] QIAN C C, HU K H, WANG H Y, et al. The effect of particle size distribution on the microstructure and properties of Al_2O_3 ceramics formed by stereolithography[J]. *Ceramics International*, 2022, 48(15): 21600-21609.
- [4] WANG L, HUANG Q H, WANG S, et al. Research on bat algorithms for inversion of particle size distribution in spectral extinction method[J]. *IOP Conference Series: Materials Science and Engineering*, 2019, 569(5): 052087.
- [5] JIANG M, LIU X W, HAN J K, et al. Influence of particle properties on measuring a low-particulate-mass concentration by light extinction method[J]. *Fuel*, 2021, 286: 119460.
- [6] 杨斌, 赵蓉, 王继, 等. 消光光谱颗粒粒径测量方法影响因素实验研究[J]. 光散射学报, 2020, 32(4): 355-360.
- [7] 肖新宇, 熊兵, 陈军, 等. 发散光束对消光法颗粒粒径测量影响[J]. 光子学报, 2022, 51(5): 0551314.
- [8] 唐红. 光全散射法颗粒粒径分布反演算法的研究[D]. 哈尔滨: 哈尔滨工业大学, 2008: 1-100.
- [9] 孙晓刚, 唐红, 原桂彬. 非独立模式算法下粒径分布反演及分类的研究[J]. 光谱学与光谱分析, 2008, 28(5): 1111-1114.
- [10] 赵蓉, 潘科玮, 杨斌, 等. 跨微米尺度混合颗粒粒径的同步测量方法[J]. 光学学报, 2020, 40(7): 0712001.
- [11] PING L, ZHAO R, YANG B, et al. Inversion of particle size distribution in spectral extinction measurements using PCA and BP neural network algorithm[J]. *Spectroscopy and Spectral Analysis*, 2021, 41(11): 3639-3647.
- [12] 李祥鹏, 项华中, 郑刚, 等. 基于近红外窄波段光源的消光光谱粒径测量[J]. 激光与光电子学进展, 2022, 59(21): 2129002.
- [13] 方文辉, 胡竹林, 朱欣娟. 基于遗传算法的用户行为数据生成[J]. 国外电子测量技术, 2021, 40(9): 154-159.
- [14] 朱小栋, 徐怡, 魏紫钰. 云环境中大数据挖掘的有效花费研究[J]. 上海理工大学学报, 2020, 42(3): 247-252.
- [15] 刘好伟, 吴志勇, 吴佳彬, 等. 大视场阵列探测器归心算法研究[J]. 光电工程, 2021, 48(6): 210039.
- [16] SHI L, LAN S, GUI H, et al. A novel 2D contactless fingerprint matching method[J]. *Neurocomputing*, 2022, 500: 547-555.

- [17] CHETANA K, KALLIMANI J S. A study on different face recognition algorithms on identical twins database[J]. *Journal of Computational Intelligence in Bioinformatics*, 2021, 14(1): 1 – 14.
- [18] HAN G, CHEN S Q, WANG X Y, et al. Noninvasive blood glucose sensing by near-infrared spectroscopy based on PLSR combines SAE deep neural network approach[J]. *Infrared Physics & Technology*, 2021, 113: 103620.
- [19] WANG Z Y, ZHOU L H, LIU T Q, et al. Development of non-invasive blood glucose regression based on near-infrared spectroscopy combined with a deep-learning method[J]. *Journal of Physics D:Applied Physics*, 2022, 55(21): 215401.
- [20] 郑刚, 卫敬明, 王乃宁. 用多波长消光法测量大颗粒的尺寸分布 [J]. *光学学报*, 1993, 13(2): 165 – 169.
- [21] ZHOU H, LI L. Experimental research on size distribution of suspended particles in water based on Mie scattering theory[J]. *IOP Conference Series:Earth and Environmental Science*, 2021, 769(4): 042063.

(编辑: 张 磊)