

DOI: 10.13652/j.spjx.1003.5788.2022.90058

基于多光谱特征分区的油桃品质分析算法

Nectarine quality analysis algorithm based on multi-spectral feature partition

王 杰

WANG Jie

(吉林建筑科技学院, 吉林 长春 130000)

(Jilin University of Architecture and Technology, Changchun, Jilin 130000, China)

摘要:目的:解决基于光谱图像识别技术易受相近类型干扰,致使对被测物的种类与品质识别产生偏差的问题。方法:设计了可见光与近红外光的独立双通道光谱采集系统,通过控制不同特征区域的光谱范围与光谱分辨率,实现对特征波长位置的吸光度快速采集;构建了品质参数,给出了其关于光谱变化与样品质量的函数表达形式;根据测试样品的光谱分布特性,选择合适的特征波长位置,并通过主成分分析给出了种类与品质参数的解算依据。结果:采用 CM-25D 分光仪和 FT-NIR 分光仪获取了 4 种常见的油桃样本的可见与红外光谱,建立了以吸光度值与吸光度比值作为判别因子的最小二乘权值分析法,与传统线性比例分析法进行比较,该算法种类识别率均值为 96.7%,归一化品质系数为 0.892,识别能力均有所增强。结论:采用双通道光谱采集硬件结构配合基于权值分配的偏最小二乘算法,可以对光谱特征相近的油桃品种进行更好的分类与识别。

关键词:光谱图像;多特征;种类识别;偏最小二乘权值分析法;主成分分析

Abstract: Objective: In order to solve the problem that the spectral image recognition technology is susceptible to interference from similar types, resulting in deviations in the identification of the type and quality of the measured object. **Methods:** An independent dual-channel spectral acquisition system for visible light and near-infrared light was designed. By controlling the spectral range and spectral resolution of different characteristic regions, the rapid acquisition of absorbance at the characteristic wavelength position was achieved. The function expression form of spectral change and sample quality was established. According to the spectral distribution characteristics of the test sample, the

appropriate characteristic wavelength position was selected, and the calculation basis of the species and quality parameters was given by principal component analysis. **Results:** The visible and infrared spectra of four common nectarine samples were obtained by CM-25D spectrometer and FT-NIR spectrometer. Compared with the traditional linear proportional analysis method, the average recognition rate of the algorithm was 96.7%, the normalized quality coefficient was 0.892, and the recognition ability was enhanced. **Conclusion:** The dual-channel spectrum acquisition hardware structure and the partial least squares algorithm based on weight assignment can better classify and identify nectarine varieties with similar spectral characteristics.

Keywords: spectral image; multiple features; category recognition; partial least square weight analysis (PLSWA); principal component analysis (PCA)

中国是油桃生产出口大国。油桃的种植、采摘与运输对产品质量的影响明显,而油桃的价格与其品质和口味密不可分^[1]。甜度、口感等的量化分析是优选油桃的重要依据。通过光谱分析对油桃的糖含量、酸度值等的无损检测成为研究热点。

为了提高系统工作的实用性,常常采用可见光或近红外光谱对果肉化学成分进行无损检测^[2]。Li 等^[3]对梨的果肉进行了光谱分析,完成了梨果肉的固形物含量、pH 值和硬度的定量分析。Li 等^[4]将模式识别技术应用于柚子种类分析,相比传统的图像识别而言,采用光谱分析的种类识别概率提高了近 1 倍。Lee 等^[5]将近红外光谱测试用于猕猴桃,完成了对磕碰猕猴桃的快速分拣。苗荣慧等^[6]利用光谱与可视化图像相结合,对 3 种不同类型油桃进行鉴别,识别率可达到 94.7%,具有很好的分类效果。武锦龙等^[7]利用神经网络算法对油桃表面的轻微损伤进行定量分析,提高了油桃品质,其算法主要针对可见光二维图像进行分类计算。由此可见,通过光谱、图像分析等方法对水果分类、状态检测的研究很多,

基金项目:国家自然科学基金青年项目(编号:61703056);吉林省优秀青年人才基金项目(编号:20190103154JH)

作者简介:王杰(1985—),女,吉林建筑科技学院讲师,硕士。

E-mail: wangjienuc@sina.com

收稿日期:2021-10-08

但大部分文献研究重点集中在分类、磕碰损伤等方面,对于不同品种的甜度及口感分析较少。

研究拟解决相似品种油桃的无损检测,以期在大量样本快速检测与识别过程中降低分类误差,从而实现不同品质油桃快速自动化分拣的目的。

1 计算模型设计

设置光谱振幅峰值、强度对比值、特征光谱谱段宽度等几个特征参数用于描述油桃的反射光谱的分布数据。为准确获取目标的光谱强度分布,对测试光信号进行标定,并测量对比用的白光光源。设任意样品 i 的光谱反射均值为 S_i ;有白光光源无测试目标时,获取的光谱分布为 S_{white} ;设当关闭光源并遮蔽光谱测试探头时获取的光谱分布为 S_{dark} 。则相对反射率为 R 有

$$R_i = \frac{1}{N} \sum_{n=1}^N \frac{S_i - S_{dark}}{S_{white} - S_{dark}} \quad (1)$$

模型中测试目标油桃的温度需要在测试环境中静置至与室温相同。其中样品上每个测试点为 n ,共计取 N 个点,然后计算平均光谱得到的。

采用 CIELAB 色彩空间^[8]对获取的光谱数据进行计算,通过光谱仪可以测量得到目标明度值为 L^* ,则其色度值 C^* 和色调值 h^* 可以表示为

$$\begin{cases} h^* = \arctan \frac{b^*}{a^*} \\ C^* = (a^{*2} + b^{*2})^{\frac{1}{2}} \end{cases}, \quad (2)$$

式中:

C^* —— 目标色度值;

h^* —— 目标色调值;

a^* —— 测试光谱中红/绿的强度值;

b^* —— 测试光谱中黄/蓝的强度值。

不同样品间的总色差^[9]可以表示为

$$\Delta E_{21} = \sqrt{(L_{i2}^* - L_{i1}^*)^2 + (a_{i2}^* - a_{i1}^*)^2 + (b_{i2}^* - b_{i1}^*)^2} \quad (3)$$

根据式(3)可以看出,对于不同的两种类型,可以通过获取其光谱分布完成对色差及光谱谱形的对比分析。当特征波长选择可见光和红外光中的几个特征位置时,不仅可以得到对油桃样品种类的区分信息,还可以利用

特征谱线位置分析其品质。

2 系统结构与光谱获取

2.1 系统结构

为了同时获取可见光与近红外的光谱分布,一方面保证光谱检测精度,采集特征波长位置上更精细的强度信号,另一方面,提高系统处理速度。系统采用了多通道数据采集的方式,并且在不同的光谱段采用不同的光谱分辨率。分辨率的选定由样品在对应区域上特征谱线的间距、峰值值及比例值决定。在可见光 380~650 nm 波段,虽然吸光度值较高,但相近品种之间的比例值较低,而占比范围相对宽,故适合采用分辨率低但覆盖范围宽的光谱采集形式,故设置光谱分辨率为 10.0 nm,而在近红外 650~1 600 nm 波段的比例值较高,有几个明显的特征吸收峰,所以采用窄光谱更合适,故采用 1.0 nm,从而在样品特征光谱位置上获得更好的细分效果。系统结构如图 1 所示。

由图 1 可知,系统除了数据处理模块之外主要分三大部分,分别是光谱分区采集模块、对比单元以及特征参数模块。光谱分区采集模块主要包括可见光通道及近红外光通道,可见光通道由数据采集卡、CM-25D 分光仪、准直透镜组以及可见光滤光片组成,近红外光通道由数据采集卡、FT-NIR 分光仪、准直透镜组以及近红外光滤光片组成。用于实现光谱分区采集,其特点是可以根据光谱特征位置而调节系统采样精度与分区位置;对比单元由参考白光与探测器构成,用于为照射样品的光源进行定标,测试得到背景光谱的辐射强度与光谱分布特征;特征参数模块包括特征数据库和分析算法,特征数据库中包含了所有类型油桃的光谱分布及其特征波长位置与吸光值,分析算法是用于完成分类、品质分析的数据模型。检测平台采用旋转结构,目的是采集目标多个位置的光谱数据进行平均,得到样品的平均光谱分布,提高品质分析的准确性。

2.2 光谱获取与预处理

为了同时获取可见光与近红外的光谱分布,提高光谱分辨精度及降低特征光谱之间的干扰,采用多通道分区获取的方式。由 CM-25D 型分光仪完成可见光光谱采

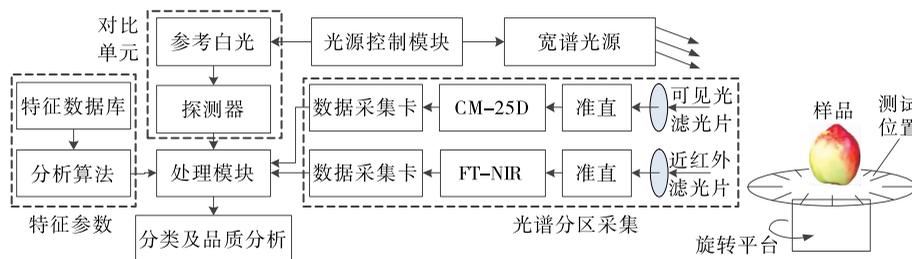


图 1 油桃种类及品质检测系统示意图

Figure 1 Schematic diagram of nectarine types and quality inspection system

集,由 FT-NIR 光谱仪完成近红外光谱采集,将分区光谱数据导入数据处理模块后利用光谱归一化^[10-11](用其每一个变量除以其相应的标准差),从而统一至同一范围区间。在此基础上,对光谱数据进行平滑处理,用以消除杂散噪声。

将每种样品的平均光谱曲线构建成矩阵,则形成4个表征对应油桃种类的矩阵,通过主成分分析提取光谱数据中的主要特征信息,再采用偏最小二乘^[12](PLS)完成4个光谱数据的回归模型。与此同时,定义样品的品质参数^[13](QP)为

$$P_Q = \frac{\ln(100 \times f_{\max} \times L^* \times h^*)}{C_{SS} \times C^*}, \quad (4)$$

式中:

P_Q ——样品的品质参数;

f_{\max} ——油桃撕裂强度,N;

C_{SS} ——可溶性固体含量,°Bx;

L^* ——目标明度值;

h^* ——目标色调值;

C^* ——目标色度值。

试验测试中,以样本集测试均值代入式(4), f_{\max} 为0.487 kN,SSC为3.19 °Bx,系数 L^* 为27.2,系数 h^* 为42.1,系数 C^* 为0.45。代入后,QP的均值为12.4,范围为11.2~13.6。

3 结果与分析

3.1 光谱获取

测试样品为瑞光5号(A)、金山早红(B)、世纪之星2号(C)和中油桃4号(D)4种油桃。每个品种选20个大小相近的样品,每5个样品通过平均计算获得一条光谱曲线。最终,通过在 Unscrambler 软件中将光谱以吸光度值进行表达,结果如图2所示。

3.2 光谱测试结果分析

由图2可知,4种样品的谱形趋势是相近的,但不同种类之间仍存在一些明显差异。对于A类而言,在400~500 nm的吸光度均值相比其他3种高,而在750~1100 nm的低吸光度平滑段,其吸光度均值要明显低于其他类型;对于B类而言,在543 nm附近的第1个弱吸收峰谷位置上,吸光度均值达0.413,具有较好的特征性,而同时,在1432 nm的强吸收峰峰位置上也存在吸光度均值偏弱的现象,而在750~1100 nm的低吸光度平滑段中其吸光度均值较高;对于C类而言,在400~483 nm的吸光度均值较其他3种类型最低,而在665 nm附近的第2个弱吸收峰峰位置上,吸光度均值达0.238,具有较好的识别性;对于D类而言,其第3个弱吸收峰谷位置的中心波长与其他3个存在明显差异,其均值中心波长为1268 nm,而其他3个的中心波长均介于1282~1296 nm,同时,在1412 nm的强吸收峰峰位置上,其吸

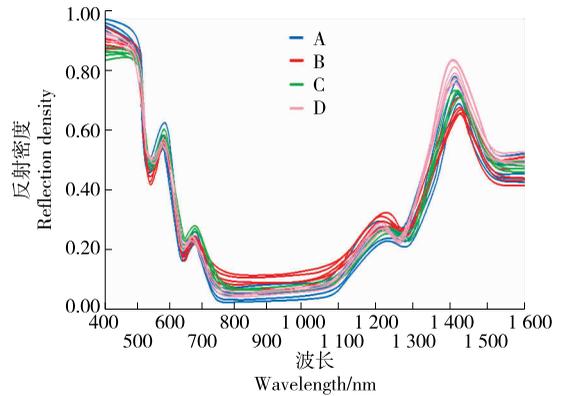


图2 4种油桃的可见光与近红外光的原始光谱分布
Figure 2 The original spectral distribution of visible and near-infrared light of four nectarines

光度均值达0.795,明显高于其他3种。分析完光谱数据中吸光度的特性后,还需要对比例值进行分析,因为单纯依靠吸光度绝对值的分析是不准确的,例如测试样本中存在杂质、浓度不均匀等问题。所以试验还引入了比例值的对比,将不同品种的相应特征位置或谱段的吸光度比值作为第2特征参量可以有效抑制单纯通过吸光度计算的干扰,系统中两种权重的比例采用1:1。由此可见,特征波长位置的选取只要能够覆盖对应种类的特征位置就能够有效地将其识别。对应区分度差或者单组数据混叠程度较高时,可以采用多个特征位置并设置合适的权重系数完成分析,该方法同样适用于样品品质、糖度等参数的分析。

综上所述,不同类型油桃样本的光谱数据具有各自的特性,而对其特性分布的分析也是选定特征光谱位置及参考权重的重要依据,也是样品种类、品质区分程度的重要标准。基于测试结果中不同样品光谱分布的特性,特征波段分别设置为400~430,530~560,670~700,1030~1060,1260~1290,1410~1440 nm。

3.3 主成分分布对比

将3.2中的特征波段区域作为数据分类计算波段,在保证种类与品质可求解的前提下,选取贡献程度高的特征波长位置,并在相应波长位置上以均值之差最大的两组样本进行比例计算,从而得到两个主成分的参数比。数据分析过程中,算法的预测样本与验证样本比例为2:1,即建模样本数为200个(4品种,每个品种50个样品),由此完成种类与品质的分类。每个样品存在两个特征区域,它们的对应关系有:①A和B的主成分分离采用450 nm和950 nm位置的比例系数;②B和C的主成分分离采用536 nm和671 nm位置的比例系数;③C和D的主成分分离采用671 nm和1411 nm位置的比例系数;④D和A的主成分分离采用450 nm和1411 nm位置的比例系数。当采用吸光度值与比例值作为主成分的两个

衡量参数(PC1 和 PC2,各 50%),则 4 种分类时的主成分样本映射如图 3 所示。吸光度值(PC1)是对光谱曲线中振幅量绝对值的描述,体现了获得的能量,但由于在整个光谱段上,4 种样品的光谱均存在混叠的问题,故采用了比例值(PC2)作为对比参数,即在满足吸光度振幅范围的同时符合两个样品振幅比值范围的才能够判定其类型与品质属性。为了提高识别概率每个样品分别选取了对应两种主成分的特征波长,所以在主成分分布图中均呈现两个区域。

由图 3 可以看出,AB 分类时,虽然在 450 nm 位置具有较强吸光度值,但是其比例值并不高,与 CD、DA 相近,而在 950 nm 位置上,虽然吸光度较弱,但比例值明显高于其他的组合方式,故增大 PC2 权重占比有利于区分 AB 的类别与品质参数;BC 分类时,在 536 nm 位置的振幅值具有较为明显的可分性,而 671 nm 处与 CD 的计算效果

相近,容易造成干扰,并且其分布重合度高,将 PC1 权重提高有利于区分 BC 的类别与品质参数;CD 分类时,相比 AB 和 BC 更为复杂,由于 400~430 nm 和 1 030~1 060 nm 两个波段位置均存在与另一种样品组混叠的可能,故需要同时符合两个特征的数据才能满足要求,即 PC1 和 PC2 的权重相近更合适;DA 分类时,由于其比例值均较低,仅 1 411 nm 波段的振幅值较高,故采用 PC1 权重大于 PC2 更合适,同时,采用 PC2 并不是求解 DA 的 PC2 值,而是用于对比 AB 的 PC2,从而在满足 PC1 符合条件且 PC2 很小时,判定为 DA 组合。至此,将所有情况分析完后就能够实现基于特征波长位置主成分分析的方法对样品种类及品质参数的定量分离。

3.4 分类概率与品质参数分析

根据主成分分析方法,结合样品测试获得的光谱数据,将权值分配引入偏最小二乘算法中,即偏最小二乘权值分析法^[14](PLSWA),并与常用的线性比例分析算法^[15](LRA)进行了比较,结果如表 1 所示。

由表 1 可知,谱段选择对分类识别概率具有明显的影响,当不区分特征谱段时,全谱段的均值效果会造成样品特征信噪比降低,从而导致样品种类识别概率降低。采用特征谱段或特征波长的 PLSWA 计算结果,4 种样品的种类识别率均值为 96.7%,归一化品质系数为 0.892,而 LRA 的仅为 79.1%,归一化品质系数为 0.803。由此可见,采用该算法对分类识别效果和品质系数测定都具有更高的准确性,即可以通过运算结果完成油桃类型和品质的分类。同时,采用特征分区的方式不但降低了光谱数据处理总量,减少了处理时间,还变相提高了信号权重,从而使其识别概率与归一化品质系数都得到了一定的提升。

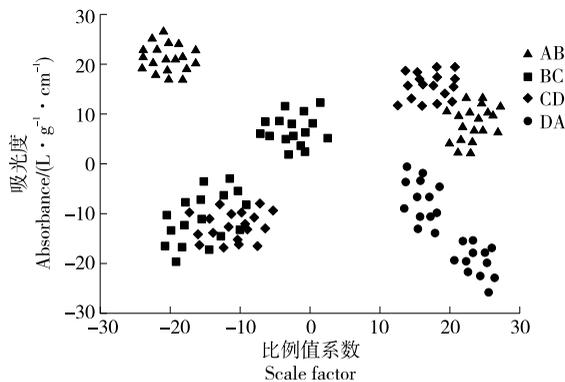


图 3 4 种油桃的测试点主成分分布图

Figure 3 Principal component distribution diagram of test points for four nectarines

表 1 不同谱段选择及不同算法条件下的识别概率与品质分析[†]

Table 1 Recognition probability and quality analysis by different spectrum and algorithms

算法	谱段选择	分类识别概率/%				归一化品质系数			
		A	B	C	D	A	B	C	D
PLSWA	FS	96.2	96.7	86.3	92.1	0.847	0.852	0.769	0.815
	CS	99.6	99.8	96.2	99.5	0.925	0.934	0.915	0.954
	WCS	99.5	99.7	95.7	99.6	0.937	0.912	0.904	0.948
LRA	FS	78.5	79.5	69.5	73.2	0.736	0.748	0.712	0.723
	CS	85.3	85.1	74.2	82.5	0.854	0.834	0.813	0.864
	WCS	84.6	83.5	72.3	81.2	0.843	0.836	0.825	0.857

[†] FS 表示采用整个光谱谱段;CS 表示采用特征谱段;WCS 表示采用特征段中的特征波长。

4 结论

针对相近品种之间光谱图像识别率低的问题,搭建了可见光与近红外光独立通道光谱图像采集系统,并设计了基于吸光度值与吸光度比值作为判别因子的最小二乘权值分析法。试验针对 4 种油桃进行种类及品质因子

分类测试,结果表明采用两个特征主成分调控的分类算法比仅依据原始光谱数据的线性数据分类的效果要好,分类识别概率与品质系数都更高。可见,该系统及其对应算法在诸如水果等相近样品的种类识别与品质分析方面具有更好的应用前景。

参考文献

- [1] 燕芳, 张俊林, 刘成毫, 等. 基于太赫兹辐射的甘草主成分光谱分析[J]. 红外技术, 2021, 43(3): 279-283.
YAN Fang, ZHANG Jun-lin, LIU Cheng-hao, et al. Principal component spectrum analysis of licorice based on terahertz radiation[J]. *Infrared Technology*, 2021, 43(3): 279-283.
- [2] 杨雨菲, 刘翠玲, 孙晓荣, 等. 基于 ATR-FTIR 光谱技术的原油含水率测量方法[J]. 传感器与微系统, 2020, 39(12): 134-136.
YANG Yu-fei, LIU Cui-ling, SUN Xiao-rong, et al. Crude oil moisture content measurement method based on ATR-FTIR spectroscopy[J]. *Transducer and Microsystem Technologies*, 2020, 39(12): 134-136.
- [3] LI J B, HUANG W Q, ZHAO C J, et al. A comparative study for the quantitative determination of soluble solids content, pH and firmness of pears by Vis/NIR spectroscopy[J]. *J Food Eng*, 2013, 116(2): 324-332.
- [4] LI X, YI S, HE S, et al. Identification of pummel cultivars by using Vis/NIR spectra and pattern recognition methods[J]. *Precis Agric*, 2016, 17: 365-374.
- [5] LEE J S, KIM S C, SEONG K C, et al. Quality prediction of kiwifruit based on near infrared spectroscopy[J]. *Kor J Horticult Sci Technol*, 2012, 30(6): 709-717.
- [6] 苗荣慧, 黄锋华, 杨华, 等. 基于光谱和 Gabor 纹理信息融合的油桃品种识别[J]. 江苏农业科学, 2019, 47(6): 174-178.
MIAO Rong-hui, HUANG Fen-hua, YANG Hua, et al. Nectarine variety recognition based on the fusion of spectrum and Gabor texture information[J]. *Jiangsu Agricultural Sciences*, 2019, 47(6): 174-178.
- [7] 武锦龙, 苗荣慧, 黄锋华, 等. 高光谱图像与卷积神经网络相结合的油桃轻微损伤检测[J]. 山西农业大学学报(自然科学版), 2019, 39(2): 79-85.
WU Jin-long, MIAO Rong-hui, HUANG Feng-hua, et al. Utilization of hyper-spectral image coupled with convolutional neural network on nectarine slight bruises detection[J]. *Journal of Shanxi Agricultural University (Natural Science Edition)*, 2019, 39(2): 79-85.
- [8] XAVIER S, COFIN A, OLSON D, et al. Remotely estimating beneficial arthropod populations: Implications of a low-cost small unmanned aerial system[J]. *Remote Sensing*, 2018, 10(9): 1 485-1 489.
- [9] KIM D W, YUN H E, JEONG S J, et al. Modeling and testing of grand status for Chinese cabbage and white radish with UAV-based RDB imagery[J]. *Remote Sensing*, 2018, 10(4): 563-569.
- [10] 曹念念, 刘强, 彭菁, 等. 基于近红外光谱技术的黄桃脆片可溶性固形物和硬度定量检测方法[J]. 食品与机械, 2021, 37(3): 51-57.
CAO Nian-nian, LIU Qiang, PENG Jing, et al. Study on quantitative detection of soluble solids and firmness of yellow peach chips by near-infrared spectroscopy[J]. *Food & Machinery*, 2021, 37(3): 51-57.
- [11] 马帅帅, 于慧春, 殷勇, 等. 黄瓜水分和硬度高光谱特征波长选择与预测模型构建[J]. 食品与机械, 2021, 37(2): 145-151.
MA Shuai-shuai, YU Hui-chun, YIN Yong, et al. Selection of hyperspectral characteristic wavelength and construction of prediction model for cucumber hardness and moisture[J]. *Food & Machinery*, 2021, 37(2): 145-151.
- [12] MA W, WEN Y Z, YU X M. Broadband metamaterial absorber at mid-infrared using multiplexed cross resonators[J]. *Opt Exp*, 2013, 21(25): 30 724-30 730.
- [13] NIKITIN A, GUINEA F, MARTIN-MORENO L. Resonant plasmonic effects in periodic graphene antidot arrays[J]. *Appl Phys Lett*, 2012, 101(15): 115-119.
- [14] 张珮, 王银红, 李高阳, 等. 基于近红外光谱对桃果实低温贮藏品质的定量检测[J]. 中国食品学报, 2021, 21(5): 355-362.
ZHANG Pei, WANG Yin-hong, LI Gao-yang, et al. Quantitative detection of the quality of peach fruit during low temperature storage based on near infrared spectroscopy[J]. *Journal of Chinese Institute of Food Science and Technology*, 2021, 21(5): 355-362.
- [15] 邵园园, 王永贤, 玄冠涛, 等. 基于高光谱成像的肥城桃品质可视化分析与成熟度检测[J]. 农业机械学报, 2020, 51(8): 344-350.
SHAO Yuan-yuan, WANG Yong-xian, XUAN Guan-tao, et al. Visual detection of SSC and firmness and maturity prediction for feicheng peach by using hyperspectral imaging[J]. *Transactions of the Chinese Society for Agricultural Machinery*, 2021, 21(5): 344-350.
- [24] ZHANG Xu-yun, LIU Chang, NEPAL S, et al. A hybrid approach for scalable sub-tree anonymization over big data using MapReduce on cloud [J]. *Journal of Computer and System Sciences*, 2014, 80(5): 1 008-1 020.
- [25] KOÇ M, KOÇ B, SUSYAL G, et al. Improving functionality of whole egg powder by the addition of gelatine, lactose, and pullulan[J]. *Journal of Food Science*, 2011, 76(9): S508-S515.
- [26] 叶贝贝. 白鲢鱼/金线鱼混合肌原纤维蛋白及肌球蛋白热聚集的作用机制[D]. 锦州: 渤海大学, 2019: 6-23.
YE Bei-bei. Mechanism of thermal aggregation of mixed myofibrillar protein and myosin from hypophthalmichthys molitrix and-nemipterus Virgatus[D]. Jinzhou: Bohai University, 2019: 6-23.
- [27] BENJAKUL S, VISESSANGUAN W, THONGKAEW C, et al. Effect of frozen storage on chemical and gel-forming properties of fish commonly used for surimi production in Thailand[J]. *Food Hydrocolloids*, 2005, 19(2): 197-207.
- [28] 张怡, 陈秉彦, 曾红亮, 等. 肌原纤维蛋白与鱼糜凝胶特性相关性概述[J]. 亚热带农业研究, 2016, 12(1): 13-24.
ZHANG Yi, CHEN Bing-yan, ZENG Hong-liang, et al. Research progress on the correlation between myofibrillar protein and gel properties in surimi[J]. *Subtropical Agriculture Research*, 2016, 12(1): 13-24.

(上接第 23 页)