

文章编号: 1005-5630(2023)02-0008-10

DOI: 10.3969/j.issn.1005-5630.2023.002.002

基于粒子群优化算法的水源微生物自动识别

闵新港, 黄邵祺, 游少杰, 戴 博

(上海理工大学 光电信息与计算机工程学院, 上海 200093)

摘要: 水源微生物检测在水源生物安全监测等方面具有非常重要的意义, 而传统的显微镜观测等方法存在效率低、需要专业人员操作等不足, 为此提出了一种水源微生物自动识别方法。采集水样, 并制作水源微生物图像集, 编写全自动与半自动两种图像分割算法用于提取目标微生物区域, 并提取6种图像特征。基于以上特征数据, 研究水源微生物识别模型的优化问题: 首先, 优化部分特征参数; 接着, 融合所有特征, 建立粒子群优化算法的支持向量机(support vector machine optimized by particle swarm optimization, PSO-SVM)微生物识别模型, 并与其他识别算法进行比较。结果表明, 相比于其他3种算法, PSO-SVM能更有效地识别各种微生物, 其平均识别率达到97.08%。

关键词: 微生物识别; 图像分割; 粒子群算法; 支持向量机

中图分类号: X 835 **文献标志码:** A

Automatic recognition of water source microorganisms based on particle swarm optimization algorithm

MIN Xingang, HUANG Shaoqi, YOU Shaojie, DAI Bo

(School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China)

Abstract: The detection of water source microorganisms is of great significance to the biosafety of water source and so on. However, the traditional methods such as microscopic observation are inefficient and need professional personnel. Therefore, an automatic recognition method of microorganisms in water source is proposed. Water samples were collected and a microorganisms image set was made. Automatic and semi-automatic image segmentation algorithms were proposed to extract the target microorganisms area, and 6 features were extracted. The model optimization problem of water microorganisms classification process was studied. First, the parameters of a few features were optimized. Then, all the features were fused, and a microorganisms recognition model of support vector machine optimized by particle swarm optimization (PSO-SVM) was established and compared with other recognition algorithms. The results show that, compared with the other 3 recognition algorithms, PSO-SVM can recognize different kinds of microorganisms more

收稿日期: 2022-11-18

基金项目: 国家重点研发计划专项(2020YFF01014503)

第一作者: 闵新港(1997—), 男, 硕士研究生, 研究方向为微生物检测。E-mail: 1967368661@qq.com

通信作者: 戴 博(1986—), 男, 教授, 研究方向为光学成像、超精密光学制造技术。E-mail: daibo@usst.edu.cn

effectively, with an average recognition rate of 97.08%.

Keywords: microorganisms recognition; image segmentation; particle swarm algorithm; support vector machine

引 言

水源微生物广泛分布于自然与人工水体环境中, 它们无法用肉眼直接观察, 例如水中的原生动物和真菌的大小通常介于 $0.1 \sim 100 \mu\text{m}$, 只能在显微镜下进行观察。作为水质的关键评价指标之一, 水源微生物的繁殖情况可以很好地反映水体的污染程度。因此, 推进水源微生物分类识别的相关研究对于水源的生物安全性监测、水环境治理等具有非常重要的意义。

考虑到传统机器学习方法对图形处理单元的要求较低, 能够以较低成本部署在设备中, 且所需的样本量小, 训练时间短, 许多学者仍在利用机器学习方法来提高微生物检测的准确性。Rahmayuna 等^[1]提出了一种利用支持向量机实现细菌种类的属级分类技术, 对 4 种细菌的预测准确率达到了 90.33%。Dhindsa 等^[2]利用像素聚类 and Kirsch 滤波分割图像中的微生物, 对不同分类器提取相应的特征, 再比较分类回归树等多种分类算法, 结果表明, 分类回归树算法最优, 对 10 种微生物的识别准确率达到了 98%。本研究通过边缘检测来提取图像中的微生物, 并提取多种图像特征并优化了部分特征参数, 基于多特征融合的方法制作微生物数据集, 采用粒子群优

化算法的支持向量机 (support vector machine optimized by particle swarm optimization, PSO-SVM) 来训练识别模型, 并将其与网格搜寻支持向量机 (grid search support vector machine, GS-SVM)^[3]、误差反向传播算法 (back propagation neural network, BP) 和集成支持向量机 (ensemble SVM)^[4] 比较, 建立了一种高效的微生物识别方法。

1 水样采集与微生物图像集制作

微生物采集于浦东威立雅自来水公司, 由于夏季是微生物繁殖的高峰期, 所以选择该时期进行采样工作, 采样类型包括过滤水和出厂水。首先将滤袋放置在出水管道上, 持续富集 3 d, 之后将采样带回实验室并转移到培养皿中, 在 Axio Vert.A1 型倒置显微镜 (德国卡尔·蔡司公司) 下观察, 照明光源为显微镜自带的明场光源, 显微物镜放大倍率为 $5\times$ 。使用 Canon EOS 80D 相机 (日本佳能株式会社) 对微生物进行拍照, ISO 速度为 800, 曝光时间为 $1/200 \text{ s}$, 图像分辨率为 4000×6000 。最终得到 8 种微生物, 共 640 张图像 (每种 80 张), 图 1 展示了这 8 种微生物的实例。采用每种微生物的 50 张图像用作训练集, 30 张图像用作测试集, 为减小计算

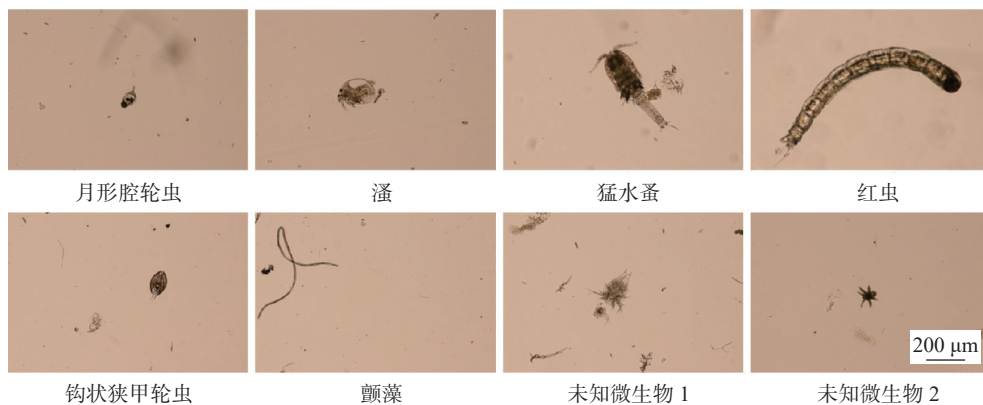


图 1 8 种微生物实例

Fig. 1 Examples of 8 species of microorganisms

量，提高后续的处理速率，图像尺寸统一调整为 600×900。用 Sefexa 软件来绘制评价标准图像。

2 水源微生物识别系统整体流程

提出了一种水源微生物自动识别系统，该系统的整体流程如图 2 所示。首先，输入制作完成的微生物图像集，通过自编写的图像分割算法对微生物图像执行图像增强、平滑滤波、边缘检测、形态学处理等操作，获得目标微生物的二值掩膜，并提取目标微生物所在区域，得到分割后的微生物图像。随后，对分割后的图像提取 6 种特征描述子，包括几何特征、内部结构直方图、傅里叶描述子、胡不变矩、灰度共生矩阵和旋转不变局部二值模式。由于 6 种特征的总维数不高，通过连接的方式进行特征融合。接下来，导入融合后的特征数据，通过粒子群算法优化 SVM 训练参数，建立优化后的微生物识别模型，最终由该模型输出微生物的识别结果。

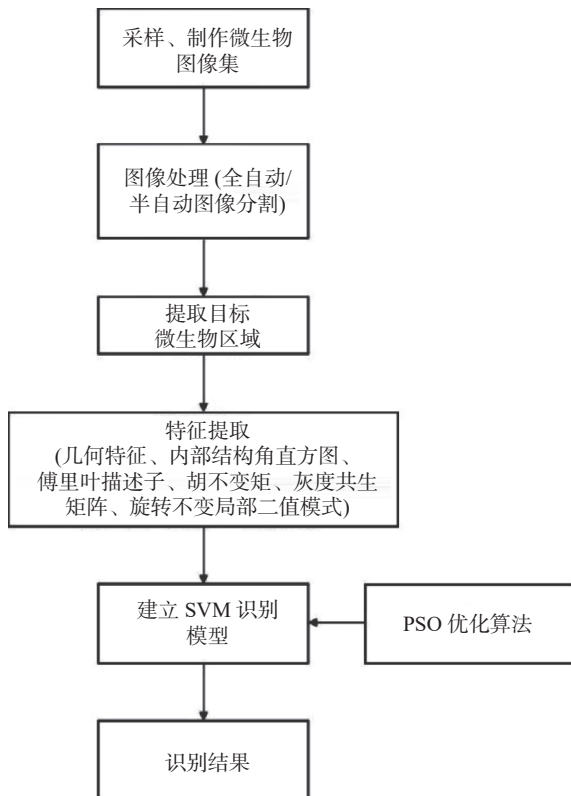


图 2 水源微生物识别系统整体流程

Fig. 2 Flow chart of the water source microorganisms recognition system

3 微生物图像分割与特征提取

3.1 微生物图像分割

设计了两种微生物图像分割方法，分别是全自动图像分割方法与半自动图像分割方法。半自动分割方法是在全自动分割方法基础上，新增了可以手动画出感兴趣的分割区域的功能，并只在该区域内进行后续的图像分割，这样可以在复杂环境中更加准确地分割出目标微生物。针对不同的环境复杂程度可采用不同的分割方法，例如：在流动的水体中，微生物和杂质密度较低，适合全自动分割方法；在富集的水样中，目标微生物会被大量杂质覆盖，适合半自动分割方法。

对于全自动分割方法，首先输入原始微生物图像，再将原始图像转化为灰度图，之后通过点运算增强对比度，高斯滤波去噪，随后用 Canny 边缘检测算子^[5]来检测微生物边界，通过形态学处理闭合边缘图像并获得最终的分割结果。对于半自动分割方法，其流程如图 3 所示：第 1 步，读取原始图像；第 2 步，用户通过光标指针手动画出感兴趣的区域；第 3 步，将原始图像转化为灰度图，高斯平滑处理，点运算增强对比度；第 4 步，裁剪出所画的感兴趣区域，并求出其他区域的灰度均值，用来填充黑色背景区域；第 5 步，用 Canny 双阈值边缘检测提取微生物边缘轮廓，并使用形态学处理方法闭合边界；最后，填充微生物区域内部以获得分割后的微生物图像。

Canny 边缘检测是最常用的边缘检测方法之一，它有着良好的抗噪声干扰性，对图像的弱边缘检测能力较好。基于梯度理论^[6]，假定 $f(x, y)$ 代表微生物图像中坐标 (x, y) 处的灰度值。则该坐标处的梯度被定义为

$$\nabla f = (G_x, G_y) = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right) \quad (1)$$

相应的，梯度的幅值大小为

$$M(\nabla f) = \sqrt{G_x^2 + G_y^2} = \sqrt{\left(\frac{\partial f}{\partial x} \right)^2 + \left(\frac{\partial f}{\partial y} \right)^2} \quad (2)$$

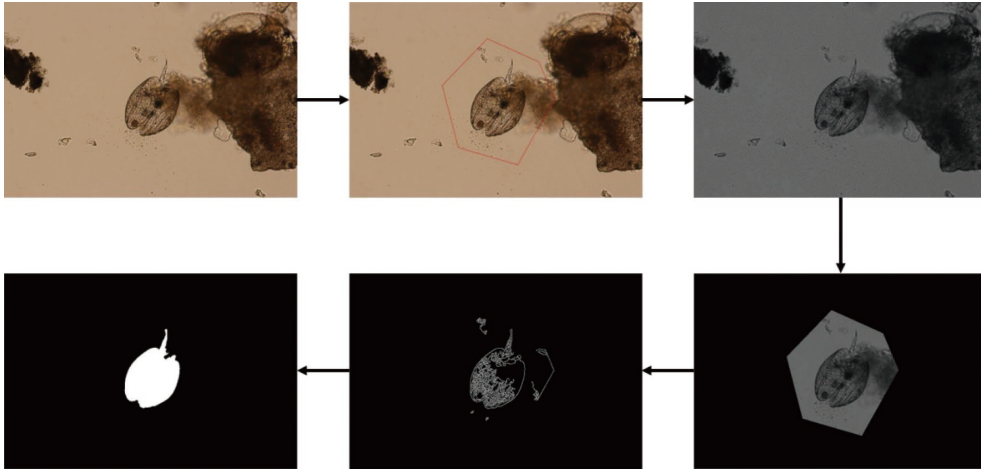


图 3 微生物图像半自动分割流程

Fig. 3 Flow chart of semi-automatic microorganisms image segmentation

该坐标处的梯度方向为

$$\alpha(x, y) = \arctan\left(\frac{G_y}{G_x}\right) \quad (3)$$

式中, G_x 和 G_y 分别为横轴 x 方向和纵轴 y 方向的梯度分量。根据式(2)和(3), 可以得到整幅图像所有像素点的梯度幅值和方向。随后 Canny 算子在 0° 、 45° 、 90° 和 135° 4 个梯度方向上对数据执行非极大值抑制, 对于中心像素点 (x, y) , 选定其周围的 8 邻域像素范围, 得到该点的梯度方向所在直线与邻域相交的两个端点。由于实际图像是离散二维矩阵, 端点不一定真实存在, 需要进行插值。将点 (x, y) 与两个端点比较, 若中心点梯度幅值是 3 个点中的最大值则保留, 反之置 0。随后设定适当的低阈值与和高阈值, 若像素点梯度幅值小于低阈值则舍弃; 若梯度幅值大于高阈值则保留; 若梯度幅值介于两者之间, 则求出该点的 8 邻域像素的梯度幅值, 如果梯度幅值高于高阈值, 则保留该点, 否则舍弃。

3.2 图像分割评价

对于以上两种图像分割方法, 采用相似度 (similarity)、灵敏度 (sensitivity) 以及特异度 (specificity) 3 个指标进行评价, 评价时的分割图像与评价标准图像均为二值图像, 1 代表感兴趣像素, 0 代表背景像素。其相似度式定义分别为

$$V_{\text{sim}} = \frac{T_p}{T_p + F_p + F_n} \quad (4)$$

式中: V_{sim} 表示分割图像与评价标准图像的相似度; T_p 表示被正确标记的像素总数; F_p 表示被错误标记为感兴趣像素的背景像素总数; F_n 表示被错误标记为背景像素的感兴趣像素总数。

$$V_{\text{sen}} = \frac{T_p}{T_p + F_n} \quad (5)$$

式中: V_{sen} 表示灵敏度, 即分割图像被正确标记的感兴趣像素总数与评价标准图像中的实际感兴趣像素总数之比。

$$V_{\text{spe}} = \frac{T_n}{T_n + F_p} \quad (6)$$

式中: V_{spe} 表示特异度, 即分割图像被正确标记的背景像素总数与实际背景像素总数之比; T_n 表示被正确标记的背景像素总数。

3.3 微生物特征提取

从分割后的微生物图像中提取几何特征、内部结构直方图特征、傅里叶描述子、胡不变矩、灰度共生矩阵和旋转不变局部二值模式, 共 6 种图像特征。

几何特征 (geometrical feature, GF) 包括: (1) 微生物的面积 A ; (2) 微生物边界轮廓的周长 P ; (3) 微生物的圆度 (circularity) $4\pi A/P^2$; (4) 与微生物区域具有相同归一化二阶中心矩的椭圆长轴长度 l_1 和短轴长度 l_2 。

内部结构直方图 (internal structure histogram, ISH) 特征: 如图 4 所示, 在微生物边界轮廓上

标记 K 个等距分布的采样点，将任意 3 个点组合，计算它们的内部结构角(例如图 4 所示的角 θ)，并将所有能得到的内切角按照一定的角度范围进行计数，由此可以得到 ISH 特征。

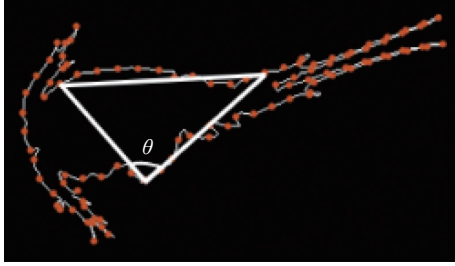


图 4 内部结构角示意图

Fig. 4 Internal structure angle

傅里叶描述子(Fourier descriptor, FD): 在微生物边界轮廓上标记 M 个等距分布的采样点，计算所有采样点对应的中心点坐标，并求出每个采样点到中心点的欧几里德距离，按顺序排列为 $\{d_0, d_1, \dots, d_m\}$ ，并通过离散傅里叶变换提取它们的频域分量作为傅里叶描述子

$$a_s = \frac{1}{M} \sum_{t=0}^{M-1} d_s \exp\left(-\frac{j2\pi st}{M}\right) \quad (7)$$

式中: a_s 表示第 s 个复系数, $s = 0, 1, \dots, M-1$; j 表示单位虚部; t 表示第 t 个采样点到中心点的欧几里德距离。为了降低噪声、平移和旋转对系数 a_s 的影响, 对其采取内部标准化

$$b_s = \frac{a_s}{a_0} \quad (8)$$

由于微生物边界轮廓可表示为实函数, 所以 b_s 有共轭对称性, 最终得到一组 $M/2$ 维特征向量 $(b_0, b_1, \dots, b_{\frac{M}{2}-1})$ 。

胡不变矩(Hu moments, Hu): Hu 是一种具有平移、旋转、尺度不变性的特征描述子^[7]。在离散条件下, Chen^[8]对其进行了改进, 得到 8 个不变矩特征, 使得提取的图像形状特征更加全面。前 7 个参数在以往的文献中已提及, 不再赘述, 第 8 个不变矩计算式为

$$\Phi_8 = 2\eta_{11} \left[(\eta_{30} + \eta_{12})^2 - (\eta_{03} + \eta_{21})^2 \right] - 2(\eta_{20} - \eta_{02})(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \quad (9)$$

式中, η_{pq} 表示图像的 $p+q$ 阶归一化中心矩。

灰度共生矩阵(gray level co-occurrence matrix,

GLCM): GLCM 是一种纹理特征, 它能反映图像中任意两像素点的灰度值的空间相关性^[9]。为了减小计算量, 把原始图像的灰度级压缩并量化为 16 级。在 0° 、 45° 、 90° 和 135° 4 个方向上计算共生矩阵, 对它们分别归一化处理, 再对每个方向的共生矩阵计算 6 个统计量, 即能量、熵、对比度、惯性矩、相关度以及逆差分矩, 最终得到一组 24 维的特征向量。

旋转不变局部二值模式(rotation invariant local binary pattern, RI-LBP): RI-LBP 特征是原始 LBP 特征的一种改进型, 具有计算简单、可适应不同尺度、灰度和旋转不变性的优点^[10]。通过将灰度图像分块, 然后在每个块内的每一个像素点的圆形邻域内进行编码采样, 之后对采样得到的 LBP 值进行循环移位操作, 得到若干个不同的值, 取其中最小的一个作为最终的 LBP 值, 最后对每个块的 LBP 值进行直方图统计得到旋转不变 LBP 特征。

假设以灰度图的某一点 (x_c, y_c) 为中心, 在其半径为 R 的圆形邻域内, 采样点个数为 Q , 则 RI-LBP 的计算方法如下

$$LBP_{Q,R}(x_c, y_c) = \sum_{q=1}^Q s(I(q) - I(c)) 2^{q-1} \quad (10)$$

$$RI-LBP_{Q,R}^{rot} = \min\{ROR(LBP_{Q,R}, i) \mid i = 0, \dots, P-1\} \quad (11)$$

式中: ROR 表示沿顺时针方向旋转 LBP 算子 i 次; $I(q)$ 和 $I(c)$ 分别表示第 q 个采样点和中心点 (x_c, y_c) 的灰度值; $s(z)$ 为门限函数。由于采样区域是圆形, 采样点可能并非实际存在的点, 要通过线性插值来完成采样点的估计。

3.4 特征参数优化与特征融合

调整 3 种特征参数, 分别是 ISH 的采样点个数和角度统计范围、FD 的采样点个数、RI-LBP 的邻域半径与采样点个数。随后用 SVM 训练不同参数条件下提取的所有微生物特征数据并测试, 为了降低结果的偶然性, 每组实验重复 10 次, 每次实验的训练集和测试集均为随机取样, 以 10 次实验的平均准确率作为评价标准, 选择准确率最优的特征参数用于后续实验。准确

率的定义为^[11]

$$V_{\text{acc}} = \frac{T_p + T_n}{T_p + F_p + F_n + T_n} \quad (12)$$

式中: T_p 表示把实际为正样本的微生物分类为正; T_n 表示把实际为负样本的微生物分类为负; F_p 表示把实际为负样本的微生物分类为正; F_n 表示把实际为正样本的微生物分类为负。

之后, 提取参数优化后的特征, 将所有 6 种特征通过连接的方式融合起来。

4 PSO-SVM 水源微生物识别方法

4.1 粒子群优化算法

粒子群优化算法(PSO)是通过模拟鸟类和鱼类种群的觅食行为而发展起来的一种基于群体协作的全局随机搜索算法^[12]。其原理是通过随机初始化粒子, 然后迭代更新粒子的速度和位置来搜寻最优解, 每个粒子沿着个体最优与全局最优的方向移动, 在多维空间的搜寻过程中, 每个粒子更新各自的速度和位置的式分别为

$$v_{id}^k = wv_{id}^{k-1} + c_1 \text{rand}_1(p_{id}^* - x_{id}^{k-1}) + c_2 \text{rand}_2(g_d^* - x_{id}^{k-1}) \quad (13)$$

$$x_{id}^k = x_{id}^{k-1} + v_{id}^k \quad (14)$$

式中: id 表示在第 d 维空间中的粒子 i ; k 为迭代次数; w 为惯性权重; c_1 和 c_2 为两个加速度常量; rand_1 与 rand_2 是在 $(0, 1)$ 中均匀分布的两个随机数; p^* 与 g^* 分别表示个体最优位置与全局最优位置。

4.2 支持向量机

支持向量机(SVM)是一种基于统计学习理论的监督学习算法。它在求解过程中加入正则项来优化结构风险, 适合处理小样本问题, 广泛应用于分类和回归分析等领域^[13]。因此, 对于本研究有限的微生物图像数据, 选择 SVM 来进行分类识别, 由于径向基函数(radial basis function, RBF)具有普适性^[14], 使用它作为核函数。

SVM 的决策函数可表示为^[15-16]

$$f(x) = \text{sign}\left(\sum_{i=1}^N \alpha_i y_i K(x, x_i)\right) + b \quad (15)$$

式中: α_i 为拉格朗日乘子; N 为用于模型训练的样本数; b 为常数偏置项; x_i 为训练样本的特征数据; y_i 为训练样本的类标签; K 为 RBF 核函数

$$K(x, x_i) = \exp(-\gamma \|x - x_i\|^2) \quad (16)$$

式中, $\gamma (\gamma > 0)$ 为核参数。上述优化问题可以用拉格朗日泛函表述为

$$\max_{\alpha} \left(\sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j K(x_i, x_j) \right),$$

$$\alpha = (\alpha_1, \dots, \alpha_i, \dots, \alpha_N); \sum_{i=1}^N \alpha_i y_i = 0; 0 \leq \alpha_i \leq C \quad (17)$$

式中: y_j 表示序号为 j 的单独样本的标签; x_i 和 x_j 分别是对应于 α_i 和 α_j 的支持向量; C 是惩罚因子。

4.3 粒子群优化算法的支持向量机(PSO-SVM)

在此项研究中, 使用 PSO 算法优化支持向量机的核参数 γ 以及惩罚因子 C , 建立 PSO-SVM 识别模型。其流程如图 5 所示, 具体步骤为: (1) 输入微生物的特征数据集; (2) PSO 算法参数设定, 包括粒子群的粒子总数、最大迭代次数、惯性权重、局部和全局学习因子、待优化参数 C 和 γ 的限制范围; (3) 初始化粒子群, 随机初始化粒子的速度和位置, 迭代次数置零; (4) 循环执行步骤 6 到步骤 9, 每循环一次则迭代次数加 1; (5) 利用 PSO 优化的 SVM 参数 C 和 γ 建立微生物图像特征的 SVM 分类模型, 并评估测试集的分类准确率; (6) 计算每个粒子的适应度; (7) 计算粒子的个体和全局优化适应度, 并根据式(13)和式(14)更新每个粒子的速度和位置; (8) 判断循环是否终止, 如果达到两代之间的最小误差或者最大迭代次数, 则终止循环, 反之则返回步骤 5; (9) 输出最优的参数 C 和 γ ; (10) 建立优化的 SVM 模型; (11) 输出微生物分类识别结果。

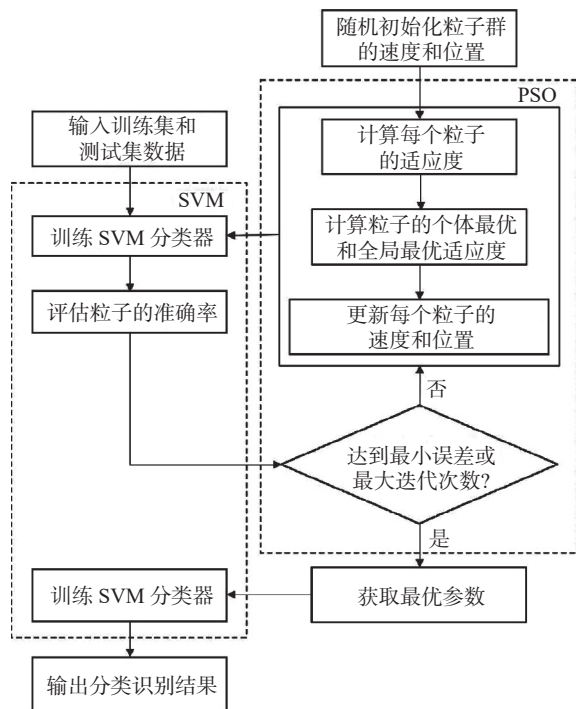


图 5 PSO-SVM 识别模型的建立流程

Fig. 5 Flow chart of establishment of the PSO-SVM recognition model

5 结果与讨论

5.1 特征参数优化

为了获得客观的评估结果，使用评价标准图像进行特征参数的优化。表 1 展示了在不同采样点数量以及不同角度统计范围下提取的 ISH 特征的识别准确率，结果表明，采样点为 100 个，角度分区范围为 $\{0\sim 15^\circ, \dots, 165^\circ\sim 180^\circ\}$ 时的准确率最高，达到了 79.41%，此时的特征维数为 12 维。表 2 展示了在不同采样点数量下提取的 FD 特征的识别准确率，结果表明，当采样点数量为 150 个时，准确率最高，达到 88.50%，此时的特征维数为 75 维。表 3 展示了不同邻域半径和采样点个数条件下 RI-LBP 特征的识别准确率，当邻域半径为 1，采样点为 8 个时，准确率最高，达到 72.16%，此时特征维数为 256 维。

5.2 图像分割评价

图 6 展示了两种分割方法的实际分割效果以及评价标准图，第 1 行是原始图像，第 2 行是评价标准图像，第 3 行是全自动分割图像，第 4 行

表 1 ISH 特征参数优化结果
Tab. 1 Results of parameter optimization of ISH feature

采样点个数	角度分区范围/(°)	特征向量维数	识别准确率/%
100	0~30, ..., 150~180	6	74.16
150	0~30, ..., 150~180	6	75.25
100	0~15, ..., 165~180	12	79.41
150	0~15, ..., 165~180	12	77.72

表 2 FD 特征参数优化结果
Tab. 2 Results of parameter optimization of FD feature

采样点个数	特征向量维数	识别准确率/%
50	25	83.16
100	50	86.41
150	75	88.50
200	100	86.97

表 3 RI-LBP 特征参数优化结果
Tab. 3 Results of parameter optimization of RI-LBP feature

采样点个数	邻域半径/像素	特征向量维数	识别准确率/%
4	1	16	60.94
4	2	16	46.84
8	1	256	72.16
8	2	256	66.97

是半自动分割图像。原始图像中目标微生物的周围有不同程度的杂质干扰，通过半自动分割方法可以将目标微生物非常精准地分割出来，分割后的图像形态与评价标准图像非常接近，而全自动分割方法无法去除与目标微生物紧挨着的大体积杂质，分割效果较差。

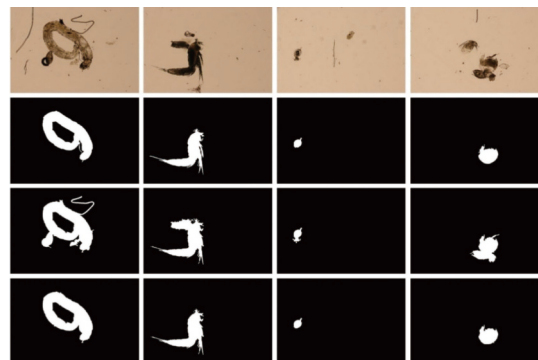


图 6 不同分割方法效果图

Fig. 6 Results of different segmentation methods

由式(4)—(6)可计算得到全自动分割图像与半自动分割图像相较于评价标准图像的相似度、灵敏度和特异度。评价结果如表 4 所示: 半自动分割图像与评价标准图的相似度为 90.94%, 比全自动分割方法提升了 4.96%; 半自

动分割图像的灵敏度为 97.24%, 比全自动分割方法提升了 0.70%; 半自动分割图像的特异度为 99.89%, 比全自动分割方法提升了 0.27%。因此, 半自动分割方法可以更好地分割微生物图像。

表 4 全自动分割方法与半自动分割方法图像评价结果
Tab. 4 Results of image evaluation of full-automatic segmentation method and semi-automatic segmentation method

评价指标	微生物类别								平均
	月形腔轮虫	溞	猛水蚤	红虫	钩状狭甲轮虫	颤藻	未知微生物1	未知微生物2	
相似度/%(全自动)	90.89	84.73	86.53	81.49	90.39	84.47	87.95	81.36	85.98
灵敏度/%(全自动)	99.42	96.10	94.00	98.14	98.25	99.19	95.94	91.27	96.54
特异度/%(全自动)	99.90	99.34	99.73	98.77	99.76	99.84	99.80	99.84	99.62
相似度/%(半自动)	94.13	91.67	90.94	94.17	95.68	87.18	89.30	84.44	90.94
灵敏度/%(半自动)	99.40	96.99	96.52	98.25	98.43	99.12	97.01	92.19	97.24
特异度/%(半自动)	99.97	99.84	99.85	99.80	99.97	99.91	99.84	99.91	99.89

为了进一步评估分割性能, 将以上两种不同分割方法分割后的微生物图像制成特征数据集, 利用 SVM 训练和测试, 求出单类别特征的识别准确率, 训练过程采用 5 折交叉验证^[17], 实验重复 10 次, 取平均值。结果如图 7 所示, 除 RI-LBP 特征以外, 半自动分割方法的其他所有特征的识别准确率均高于全自动方法, 其中 GF 的识别结果最优, 其准确率为 90.84%。与评价标准图相比, 半自动分割图像的 GF、Hu、GLCM 3 种特征的识别准确率差异均在 1% 以内, ISH 的准确率差异约 3%, RI-LBP 的准确率差异约 4%。此外, 对于半自动分割图像, 其所

有 6 种特征的平均准确率为 77.35%, 与评价标准图的相应值相比仅低了 2.62%; 全自动分割方法的所有特征平均准确率与评价标准图的相应值比较, 低了 7.35%。因此, 对拍摄于富集后的水体微生物图像, 适用于半自动分割方法。因为大量的微生物与杂质堆积在一起, 分割难度非常大, 而通过手动选择感兴趣区域, 可以有效降低图像噪声的影响, 实现对目标微生物的精准分割。对于拍摄于流动水体的微生物图像, 则适合全自动图像分割方法, 因为该方法无需人工干预, 能实时处理图像, 且微生物和杂质的密度较低, 大多数微生物能被正确地分割。

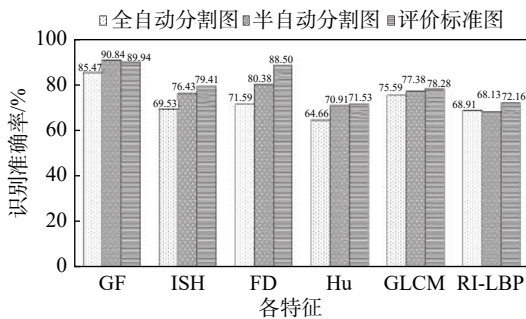


图 7 半自动分割图像、全自动分割图像与评价标准图像的单类别特征识别结果

Fig. 7 Single feature recognition results of semi-automatic segmented image, automatic segmented image and evaluation standard image

5.3 PSO-SVM 微生物识别模型

评价标准图提取的微生物的 6 种特征数据, 经过特征融合后作为特征数据集。利用 PSO 算法优化 SVM 的核参数 γ 以及惩罚因子 C , 得到 PSO-SVM 微生物识别模型。PSO 算法的参数设置如下: 粒子的种群大小为 24; 加速常数 C_1 与 C_2 均为 1.5; 惯性权重为 0.6; 最大迭代次数为 100。在 PSO 优化过程中, 使用 SVM 模型的识别准确率作为粒子适应度。

利用 PSO 优化 SVM 参数过程中的适应度曲线如图 8 所示, 适应度随着粒子的迭代次数增加而逐渐收敛, 在第 72 代时达到最大值, 由此获

得最优的 SVM 参数 $C = 21.1652$, $\gamma = 0.025504$ 。随后利用训练完成的 PSO-SVM 模型对 8 种微生物的测试集进行分类识别, 采用 5 折交叉验证, 测试 10 次, 取平均值作为最终测试结果, 并将其与 GS-SVM、BP、Ensemble SVM 3 种识别方法的测试结果比较。

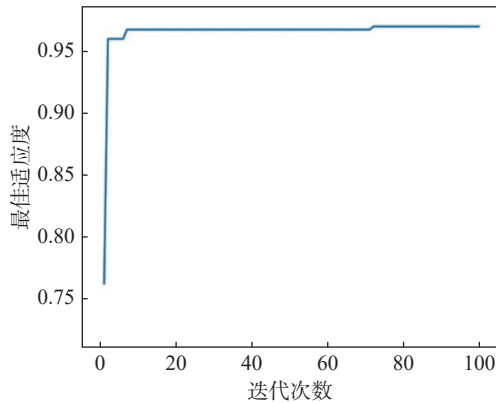


图 8 PSO 适应度曲线

Fig. 8 The fitness of PSO

结果如图 9 所示, PSO-SVM 识别方法对于此 8 种微生物具有最好的分类性能, 平均识别率为 98.60%, 其中: 月形腔轮虫的识别率为 99.19%; 溇的识别率为 98.13%; 猛水蚤的识别率为 98.77%; 钩状狭甲轮虫的识别率为 97.17%; 未知微生物 1 的识别率为 97.92%; 未知微生物 2 的识别率为 97.60%; 红虫和颤藻的识别率最高, 均为 100%。GS-SVM 模型的平均识别率为 96.28%, 比 PSO-SVM 低了 2.32%; Ensemble SVM 的识别率为 91.57%, 比 PSO-SVM 低了 7.03%; BP 神经网络的平均识别率为 81.33%, 在 4 种算法中最低, 这是因为微生物样本数量较少所致。

实验结果表明, PSO-SVM 识别方法的识别性能优于其他 3 种分类算法, 更适用于水源微生物识别。利用全自动分割图像提取的特征数据建立的 PSO-SVM 模型平均识别率为 92.25%, 相较于评价标准图低了 6.35%, 这是由于本文中的原始微生物图像均采集于富集后的水样, 其微生物与杂质比较密集, 微生物图像的全自动分割效果较差所致。利用半自动分割图像提取的特征数据建立的 PSO-SVM 模型, 其平均识别率为 97.08%, 仅比评价标准图的识别率低 1.52%。

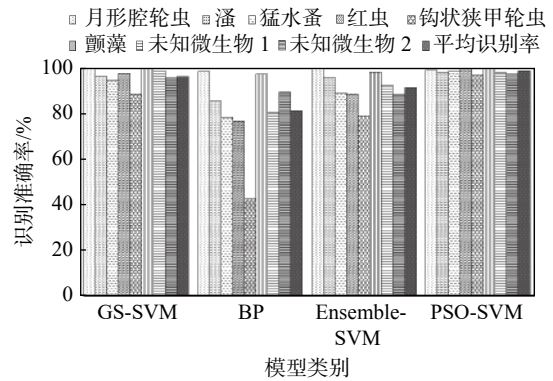


图 9 4 种算法微生物识别结果

Fig. 9 Microorganisms recognition results of four algorithms

6 结 论

提出了一种基于多特征融合和 PSO-SVM 的水源微生物自动识别方法。考虑到水源环境的不同复杂程度, 设计了全自动和半自动分割两种微生物图像分割方法, 分别适用于流动水体中拍摄的微生物图像以及水体富集后拍摄的微生物图像。针对 8 种微生物优化了 ISH、FD、RI-LBP 的特征参数, 通过连接的方式将所有特征融合起来, 最后利用该特征数据建立了 PSO 算法优化的 SVM 模型, 实现对 8 种水源微生物的分类和识别。实验结果表明, 该模型能够准确地识别其中的每一种微生物, 平均识别准确率达到 97.08%。因此, 该技术有望用于水源微生物的高效自动检测, 以提高居民用水的安全保障和水源环境监测效率。

参考文献:

- [1] RAHMAYUNA N, RAHARDWIKA D S, SARI C A, et al. Pathogenic bacteria genus classification using support vector machine[C]//Proceedings of 2018 International Seminar on Research of Information Technology and Intelligent Systems. Yogyakarta: IEEE, 2018: 23 - 27.
- [2] DHINDSA A, BHATIA S, AGRAWAL S, et al. An efficient microbes detection system using microscopic images via morphological and correlation based features[J]. *Biomedical and Pharmacology Journal*, 2020, 13(3): 1113 - 1124.
- [3] 黄挺, 王元庆, 张自豪. 基于 GS-SVM 的彩色图像分

- 割算法 [J]. *电子测量技术*, 2017, 40(7): 105 – 108,112.
- [4] 黄永明, 章国宝, 董飞, 等. 基于 Gabor、Fisher 脸多特征提取及集成 SVM 的人脸表情识别 [J]. *计算机应用研究*, 2011, 28(4): 1536 – 1539,1543.
- [5] RAMNARAYAN R, SAKLANI N, VERMA V. A review on edge detection technique “canny edge detection” [J]. *International Journal of Computer Applications*, 2019, 178(10): 28 – 30.
- [6] GONZALEZ R C, WOODS R E. *Digital image processing*[M]. 3rd ed. New Jersey: Prentice Hall, 2008: 728.
- [7] ANAMI B S, BHANDAGE V A. Combined Hu moments, orientation knowledge, and grid intersections feature based identification of Bharatanatyam mudra images[J]. *Pattern Analysis and Applications*, 2019, 22(4): 1439 – 1454.
- [8] CHEN C C. Improved moment invariants for shape discrimination[J]. *Pattern Recognition*, 1993, 26(5): 683 – 686.
- [9] HARALICK R M, SHANMUGAM K, DINSTEN I H. Textural features for image classification[J]. *IEEE Transactions on Systems, Man, and Cybernetics*, 1973, SMC-3(6): 610 – 621.
- [10] LIU P Z, GUO J M, CHAMNONGTHAI K, et al. Fusion of color histogram and LBP-based features for texture image retrieval and classification[J]. *Information Sciences*, 2017, 390: 95 – 111.
- [11] CHEN Y J, WU Z N, ZHAO B, et al. Weed and corn seedling detection in field based on multi feature fusion and support vector machine[J]. *Sensors*, 2021, 21(1): 212.
- [12] WANG D S, TAN D P, LIU L. Particle swarm optimization algorithm: an overview[J]. *Soft Computing*, 2018, 22(2): 387 – 408.
- [13] BURGESS C J C. A tutorial on support vector machines for pattern recognition[J]. *Data Mining and Knowledge Discovery*, 1998, 2(2): 121 – 167.
- [14] CHANG C C, LIN C J. LIBSVM: a library for support vector machines[J]. *ACM Transactions on Intelligent Systems and Technology*, 2011, 2(3): 27.
- [15] YOU D Y, GAO X D, KATAYAMA S. WPD-PCA-based laser welding process monitoring and defects diagnosis by using FNN and SVM[J]. *IEEE Transactions on Industrial Electronics*, 2015, 62(1): 628 – 636.
- [16] XU Q S. Identification and compensation of piezoelectric hysteresis without modeling hysteresis inverse[J]. *IEEE Transactions on Industrial Electronics*, 2013, 60(9): 3927 – 3937.
- [17] BERGMEIR C, BENÍTEZ J M. On the use of cross-validation for time series predictor evaluation[J]. *Information Sciences*, 2012, 191: 192 – 213.
- (编辑: 李晓莉)