

基于计算机视觉和 GA-SVM 的梭子蟹体重预测

唐杨捷, 胡海刚*, 张 刚, 唐 潮

(宁波大学 海运学院, 浙江 宁波 315211)

摘要: 以梭子蟹为研究对象, 利用计算机视觉技术对其进行测量. 通过 CCD 相机获取不同生长情况下的梭子蟹图像, 采用图像处理技术对图像进行分割处理, 计算获得的投影面积、全甲宽与甲长参数; 利用图像获取的尺寸参数对梭子蟹体重进行预测, 发现梭子蟹投影面积、全甲宽、甲长与体重具有正相关性; 并采用遗传优化(GA)的支持向量机(SVM)建立梭子蟹体重回归预测模型. 实测结果表明, 梭子蟹体重预测值与实测值平均绝对百分比误差(MAPE)为 2.23%, 均方根误差(RMSE)为 5.80 g, 优于 BP 神经网络和参数未优化的 SVM 预测. 证明基于计算机视觉与遗传优化支持向量机(GA-SVM)的梭子蟹体重预测方法能够达到梭子蟹体重测量要求.

关键词: 梭子蟹; 体重; 计算机视觉; 遗传支持向量机; 预测模型

中图分类号: TP183

文献标志码: A

文章编号: 1001-5132 (2019) 01-0032-06

发展海洋经济已成为 21 世纪世界各国加快经济发展、增强国际竞争力的战略选择之一, 水产养殖即是海洋经济的一个重要组成部分, 而梭子蟹是水产养殖产品中的一种^[1]. 目前工厂化单筐养殖逐渐成为梭子蟹养殖的主要模式, 这种养殖模式避免了个体之间的相互残杀, 使得养殖成活率显著上升^[2]. 而在养殖过程中, 对于梭子蟹的体重测量极为重要, 特别是在投饵方面, 投饵量的确定对于梭子蟹摄食与生长影响巨大. 目前, 投饵量主要依靠人工测量梭子蟹体重来确定, 而人工定期捕捞梭子蟹并测量存在着一定的弊端. 因此, 如何实现梭子蟹体重的无损精确测量是实现梭子蟹养殖科学、高效、无害的一个重要研究课题.

计算机视觉技术是一种无损的检测方法^[3-4]. 水产养殖领域已逐步开始运用计算机视觉技术来进行水产生物品种分类、品质分级、测量等研究^[5-6]. 而在水产生物体重预测方面, Lines 等^[7]将摄像机放置于水下拍摄获取鲑鱼鱼体图像, 通过鲑鱼的多种长度因子与体重相关性来进行体重估计. Liang 等^[8]采集了台湾罗非鱼的投影图像, 利用回归模型分别对其体长、体高、周长、面积和质量进行相关

性分析, 最终发现质量和面积的相关性最好. 刘伟等^[9]以白鲢切片建立质量分布模型, 利用计算机视觉预测白鲢质量. 张超等^[10]利用最小二乘法对梭子蟹面积和体重进行拟合, 其中二次多项式拟合系数能够达到 0.922 0, 但由于只采用面积参数进行体重估计, 随机误差较大.

鉴于上述问题和需求, 笔者提出了一种基于计算机视觉技术实现梭子蟹体重预测的研究. 利用计算机视觉技术获取梭子蟹静态图像, 通过图像处理来获得梭子蟹投影面积、全甲宽与甲长特征参数, 采用支持向量机(Support Vector Machine, SVM)回归预测梭子蟹体重, 并利用遗传算法(Genetic Algorithm, GA)优化 SVM 模型, 以提高回归预测精度, 从而获取准确的梭子蟹体重参数.

1 材料与方法

1.1 实验材料

实验对象为梭子蟹, 梭子蟹均采用单筐养殖, 实验周期为 60 d, 每半月抽取 1 次梭子蟹进行实验, 每批次选取生长状况良好且个体大小各有不同的共 40 只梭子蟹, 对每只梭子蟹依次进行拍摄.

收稿日期: 2018-03-15.

宁波大学学报(理工版)网址: <http://journalg.nbu.edu.cn/>

基金项目: 浙江省公益技术项目(2017C32014); 宁波市科技富民项目(2017C10006); 宁波市农业重大项目(2017C110007).

第一作者: 唐杨捷(1991—), 男, 浙江宁波人, 在读硕士研究生, 主要研究方向: 船舶自动化与机电控制. E-mail: 1242981685@qq.com

*通信作者: 胡海刚(1966—), 男, 浙江定海人, 高级实验师, 主要研究方向: 船舶工程安全技术与机电控制. E-mail: huhuigang@nbu.edu.cn

1.2 实验装置

实验系统装置结构如图1所示,实验主要在单个筐体内进行.图像采集装置由CCD摄像头、计算机、内框边长为37 cm×37 cm的养殖筐以及俯拍支架组成.采用CCD摄像头,最大像素2048×1536,通过USB接口传输,另外配备焦距为16 mm的定焦镜头,摄像头畸变率在0.3%以下.

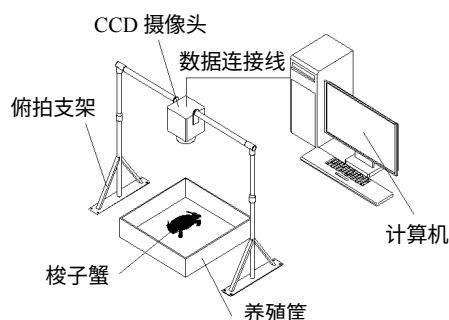


图1 实验装置结构图

将梭子蟹放置在养殖筐内,并且在养殖筐内注入一定量的养殖海水使梭子蟹能够在水中运动,摄像头安装在俯拍支架上,调整高度与水平角度,保证良好的拍摄效果.

另外,梭子蟹实际体重测量采用精度为0.1 g的电子秤,图像处理主要采用OpenCV实现,通过C++程序编译,体重预测估计通过Matlab实现.

1.3 方案设计原理

获取到的视频图像被传输到计算机上,将视频图像转化成静态图像,选取效果较好的若干张图像进行图像处理.图像处理计算梭子蟹投影面积、全甲宽与甲长等尺寸参数,通过SVM建立梭子蟹体重预测模型,并且利用遗传算法对SVM模型参数进一步优化,从而预测估计梭子蟹的体重,方案设计流程如图2所示.

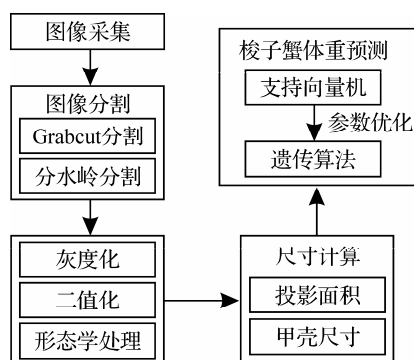


图2 方案设计原理

2 图像处理与参数获取

本研究通过摄像机俯拍获取图像后进行图像处理,进而计算梭子蟹投影面积和甲壳尺寸.本文主要甲壳尺寸如图3所示.

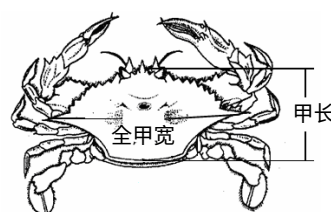


图3 甲壳尺寸^[11]

2.1 图像采集与预处理

为方便将图像尺寸换算成实际尺寸,选用白色正方形PVC板(37 cm×37 cm),将其放置于盛水养殖筐底部拍摄得到正方形板图像(图4(a)),再依次放入不同梭子蟹,拍摄获得目标图像(图4(b)).

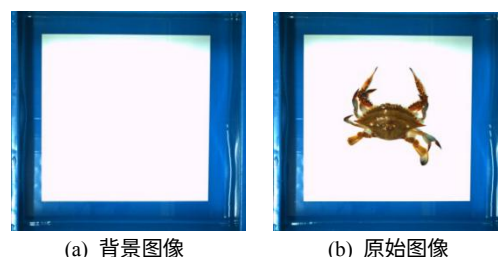


图4 图像采集

2.2 梭子蟹尺寸参数获取

针对梭子蟹投影面积提取,通过不同分割方法的试验对比,最终确定分割效果最好的Grabcut分割^[12].Grabcut分割后,得到梭子蟹彩色图像的分割效果(图5(a)),再进行灰度化处理.由于分割后的图像主要为彩色目标与白色背景,因此采取阈值分割能够达到较好的分割效果,且方法简单;图像在分割与二值化后还需要进行一定的形态学处理,包括膨胀、腐蚀、去除内部粒子与内孔,处理得到梭子蟹最后二值图像(图5(b)).分水岭分割是一种基于拓扑理论的数学形态学的分割方法^[13],采用分水岭分割方法对所有得到的Grabcut分割图像进行甲壳分割,分割效果良好,均能实现后续甲壳尺寸计算.因而,对于梭子蟹的全甲宽、甲长的参数获取主要采用此分割方法.其中,图5(c)即为对图5(a)得到的梭子蟹经分水岭分割后的效果,最后进行甲壳的长宽特征计算,得到图5(d).



图5 图像分割

2.3 尺寸计算

数字图像主要由像素点组成, 因此可通过像素点计算梭子蟹尺寸. 梭子蟹像素面积与像素长度能够利用 OpenCV 直接计算, 如图 5(b)、5(d)所示. 通过养殖筐底部的正方形板比例换算能够计算像素对应尺寸. 梭子蟹尺寸与换算公式如下:

$$L = L_0 \eta, S_z = S_x \eta^2, \eta = L_x / L_z, \quad (1)$$

式中, L 为标定正方形板实际尺寸长度; L_0 为标定正方形板的像素长度; η 为单个像素的实际长度; S_z 为梭子蟹投影面积; S_x 为梭子蟹像素面积; L_x 为目标像素长度; L_z 为目标实际长度.

在换算过程中, 由于梭子蟹面积无法测量, 因此需要根据梭子蟹全甲宽与甲长的尺寸测量来分析图像测量结果的准确性.

3 梭子蟹体重预测

3.1 支持向量机原理

支持向量机(SVM)由 Vapnik 等人于 20 世纪 90 年代提出的一种机器学习方法, 在分类与回归估计中已经成为一种有效的工具^[14]. SVM 具有较好的泛化能力, 能够对小样本、高维、非线性的系统精确拟合, 具有良好的预测效果^[15], 本文采用非线性回归支持向量机.

设训练样本集 $T = \{(x_i, y_i) | i = 1, 2, \dots, l\}$. 其中 $x_i \in \mathbf{R}^n$ 指第 i 个输入值, $y_i \in \mathbf{R}$ 为输出值, n 为向量维数, l 为样本个数. 通过非线性映射函数 $\psi(\cdot)$ 把 x 从原空间 $\mathbf{R}^n$ 映射到高维特征空间 $\mathbf{R}^g$ ($g > n$):

$$\psi(x) = (\psi(x_1), \psi(x_2), \dots, \psi(x_n)), x_i \in \mathbf{R}^g. \quad (2)$$

定义的不敏感损失函数 ε 为:

$$L(f(x), y, \varepsilon) = \begin{cases} 0, & |y - f(x)| \leq \varepsilon, \\ |y - f(x)| - \varepsilon, & |y - f(x)| > \varepsilon, \end{cases} \quad (3)$$

其中, $\varepsilon > 0$, f 为决策函数, 其表达式为:

$$f(x) = \mathbf{w}^T \cdot \psi(x) + b, \quad (4)$$

其中, $\mathbf{w}$ 为权值向量; b 为偏置量.

求解回归问题时应使目标函数 $\|\mathbf{w}\|^2 / 2$ 最小, 转化成最优化问题为:

$$\begin{aligned} \min \quad & \frac{1}{2} \|\mathbf{w}\|^2, \\ \text{s.t.} \quad & \begin{cases} y_i - \mathbf{w} \cdot \psi(x_i) - b \leq \varepsilon, \\ \mathbf{w} \cdot \psi(x_i) + b - y_i \leq \varepsilon, i = 1, 2, \dots, n. \end{cases} \end{aligned} \quad (5)$$

因为需要有拟合误差存在, 引入惩罚因子 C 和松弛变量 ξ_i, ξ_i^* , 则最优化问题表达式确定为:

$$\begin{aligned} \min \quad & \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*), \\ \text{s.t.} \quad & \begin{cases} y_i - \mathbf{w} \cdot \psi(x_i) - b \leq \varepsilon + \xi_i, \\ \mathbf{w} \cdot \psi(x_i) + b - y_i \leq \varepsilon + \xi_i^*, \\ \xi_i, \xi_i^* \geq 0, i = 1, 2, \dots, n. \end{cases} \end{aligned} \quad (6)$$

通过拉格朗日求解凸二次规划问题, 利用拉格朗日乘子求解法得到:

$$\begin{aligned} \min \quad & \left\{ \frac{1}{2} \sum_{i,j=1}^n (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) K(x_i, x_j) + \right. \\ & \left. \varepsilon \sum_{i=1}^n (\alpha_i + \alpha_i^*) - \sum_{i=1}^n y_i (\alpha_i - \alpha_i^*) \right\}, \\ \text{s.t.} \quad & \begin{cases} \sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0, \\ 0 \leq \alpha_i \leq C, 0 \leq \alpha_i^* \leq C. \end{cases} \end{aligned} \quad (7)$$

最终得到函数:

$$f(x) = \sum_{x_i \in \text{SV}} (\alpha_i - \alpha_i^*) K(x_i, x) + b, \quad (8)$$

式中, α_i, α_i^* 为非负拉格朗日乘子, $K(x_i, x)$ 为核函数. 核函数有多种形式^[16], 本文选择其中的径向基核函数(RBF):

$$K(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / \rho^2), \quad (9)$$

其中, ρ 为核函数参数.

3.2 基于 GA 的 SVM 参数优化

本文通过遗传算法^[17]来优化 SVM 参数: 惩罚因子 C 与 RBF 核函数参数 ρ . 利用遗传算法的全局搜索能力来确定最优参数, 从而提高模型的预测精度.

对惩罚因子 C 与核函数参数 ρ 进行二进制编码,

初始化生成由一定数量个体组成的初始种群, 每条染色体由参数组构成, 再把训练产生的误差作为目标函数对初始种群进行 SVM 训练, 计算得到适应度值, 通过给定的概率进行遗传操作, 包括选择、交叉、变异运算, 生成新种群^[18], 对新种群进行 SVM 训练, 当达到终止条件时停止操作, 最后解码, 将获得的参数最优解对预测样本进行建模预测. GA 优化 SVM 参数流程图如图 6 所示.

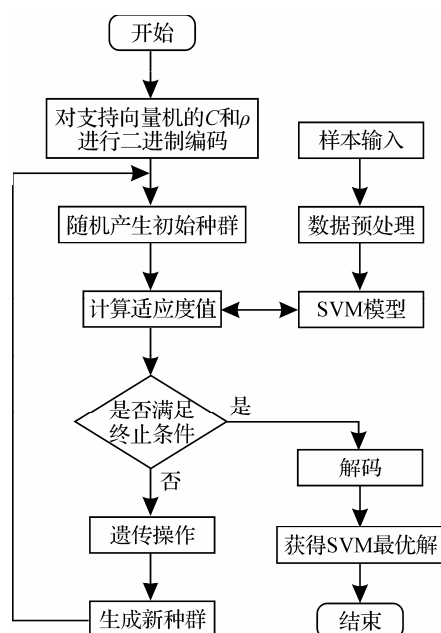


图6 GA 优化 SVM 参数流程

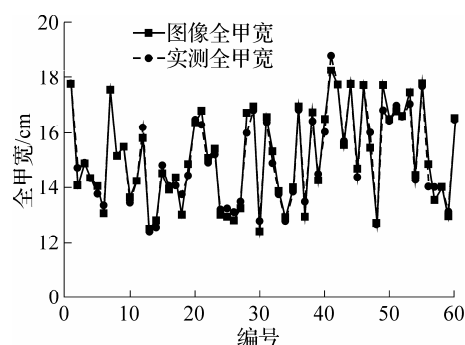
4 结果与分析

4.1 实验数据分析

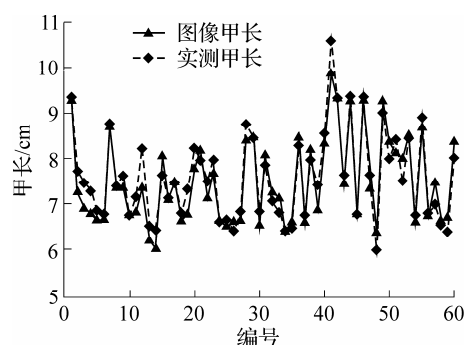
利用计算机视觉技术获取梭子蟹图像, 通过图像处理技术测量得到投影面积、全甲宽、甲长, 同时采用人工检测得到每只梭子蟹的甲长与全甲宽实测数据, 将实测尺寸与图像测量尺寸作比较. 图 7(a)为随机选取其中 60 只梭子蟹图像测量全甲宽与实测全甲宽的分析, 而图 7(b)表示梭子蟹图像测量甲长与实测甲长的分析. 数据经过误差分析得到, 全甲宽的平均绝对百分比误差为 1.74%; 甲长的平均绝对百分比误差为 3.33%, 结果表明能够满足测量精度要求.

4.2 GA-SVM 预测模型参数确定

利用遗传算法优选支持向量机惩罚因子 C 与核函数参数 ρ , 设定种群大小 $N=20$, 最大进化代数 $g_m=200$, 交叉概率 $p_c=0.4$, 变异概率 $p_n=0.01$, 惩罚



(a) 图像全甲宽与实测全甲宽对比



(b) 图像甲长与实测甲长对比

图7 图像测量与实测

因子参数 C 在 $(0,100]$ 内, 核函数参数 ρ 范围为 $(0, 1000]$. 通过计算机视觉技术共获取图像 150 张, 并对每张图像处理获得 150 组数据, 随机选取 132 组作为训练样本数据, 18 组作为测试样本数据.

通过遗传算法计算支持向量机的最优参数解, 其中惩罚因子 $C=0.9883$, 核函数参数 $\rho=2.4242$. 将这 2 个参数应用到 SVM 模型中, 对训练样本数据进行训练, 并将训练后的 GA-SVM 回归预测模型对测试样本进行预测, 预测结果见图 8 与表 1. 通过图 8 与表 1 分析计算得到, 预测相对误差绝对值最大为 5.66%, 最小为 0.11%, 平均绝对百分比误差为 2.23%.

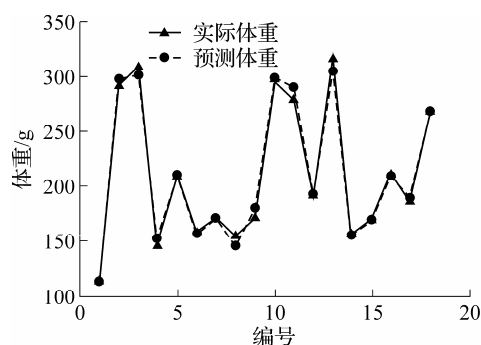


图8 GA-SVM 预测结果

表1 GA-SVM 预测误差分析

编号	投影面积/cm ²	图像全甲宽/cm	图像甲长/cm	预测体重/g	实测体重/g	相对误差/%
1	88.46	12.43	6.58	114.19	113.6	0.52
2	191.25	16.53	8.44	297.59	291.8	1.98
3	202.47	17.76	9.34	302.07	308.4	2.05
4	118.57	13.76	6.82	152.20	145.0	4.97
5	131.02	15.20	8.20	210.31	208.7	0.77
6	121.08	13.33	7.49	156.65	158.4	1.11
7	133.22	13.92	7.15	171.43	169.5	1.14
8	122.06	12.99	6.68	145.38	154.1	5.66
9	136.88	14.87	6.81	179.98	170.5	5.56
10	196.45	16.94	8.53	298.96	295.2	1.27
11	173.45	16.70	8.62	290.33	279.1	4.02
12	155.13	14.74	6.86	192.78	190.2	1.35
13	194.01	17.90	9.16	305.31	316.2	3.44
14	113.85	14.02	6.66	155.67	154.3	0.88
15	132.32	13.86	7.16	170.02	167.9	1.26
16	155.13	15.30	7.30	209.78	212.8	1.42
17	132.46	14.94	7.48	189.35	184.5	2.63
18	168.16	15.84	8.24	269.00	268.7	0.11

4.3 预测模型比较分析

为验证 GA-SVM 对于梭子蟹体重预测精确度, 采用 BP 神经网络、参数未优化的 SVM 与 GA-SVM 进行预测比较, 依旧选用测试样本的 18 组数据, 预测结果与实测值比较如图 9 所示。而 BP 神经网络、SVM 与 GA-SVM 模型误差比较数据见表 2。

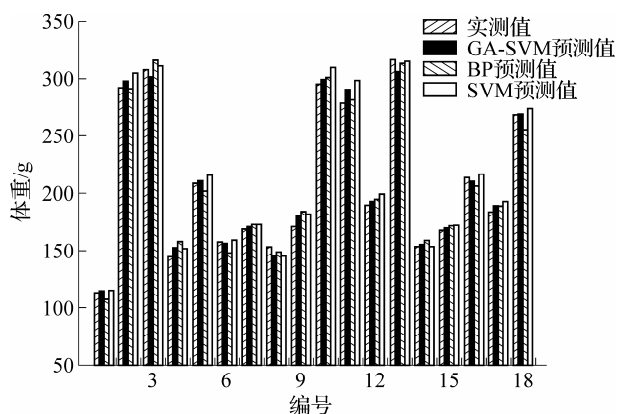


图9 预测结果比较

通过图 9 与表 2 可以看出, BP 神经网络、SVM 与 GA-SVM 对梭子蟹体重回归预测效果均较好, 3 种模型的预测平均绝对百分比误差均在 4% 以下, 而 GA-SVM 预测的平均绝对百分比误差为 2.23%,

表2 3 种模型预测误差比较

误差	BP	SVM	GA-SVM
最大相对误差/%	9.19	7.47	5.66
最小相对误差/%	0.11	0.37	0.11
平均绝对百分比误差/%	3.40	3.45	2.23
均方根误差/g	7.34	8.75	5.80

比 BP 神经网络和 SVM 分别低 1.17% 和 1.23%, 且 GA-SVM 最大相对误差低于 BP 神经网络, 最小相对误差相等。在均方根误差方面, GA-SVM 的均方根误差最小。总体相比较而言, GA-SVM 对于梭子蟹体重预测精度高于 BP 神经网络。而通过与 SVM 相比较, 不难发现 GA-SVM 各方面均优于传统的 SVM, 因此 GA 对于 SVM 的参数优化完全提高了 SVM 的预测性能。

5 结语

(1) 研究采用 CCD 相机获取梭子蟹图像, 通过计算机视觉技术测量计算得到梭子蟹投影面积、全甲宽与甲长, 将图像尺寸与实际尺寸测量值比较, 全甲宽平均绝对百分比误差为 1.74%, 甲长平均绝对百分比误差为 3.33%, 满足测量误差要求。

(2) 梭子蟹体重预测主要与梭子蟹投影面积、全甲宽与甲长相关, 将图像所测得的 3 个指标通过 GA-SVM 模型能够回归预测得到对应的梭子蟹体重, 基于 GA-SVM 的梭子蟹体重回归预测平均绝对百分比误差为 2.23%, 最大相对误差与最小相对误差为 5.66% 和 0.11%, 满足精度要求。

(3) 本文 SVM 选用 RBF 核函数, 利用 GA 对核函数参数 ρ 与惩罚因子 C 寻优, 确定最优参数组, 提出 GA-SVM 预测模型, 通过实验验证, 与 BP 神经网络和 SVM 进行比较分析, 预测误差均小于其他两种模型。因此, 基于计算机视觉与 GA-SVM 能够实现梭子蟹的体重预测, 具有较高应用性。

参考文献:

- [1] 穆占昆, 杨振国, 周玉, 等. 中国对虾和三疣梭子蟹混养试验[J]. 水产科学, 2001, 20(5):16-18.
- [2] 王春琳, 母昌考, 李荣华, 等. 三疣梭子蟹单体筐养高产高效生产技术[J]. 中国水产, 2013, 1(1):72-76.
- [3] 李长纓, 滕光辉, 赵春江, 等. 利用计算机视觉技术实现对温室植物生长的无损监测[J]. 农业工程学报, 2003, 19(3):140-143.

- [4] 徐建瑜, 崔绍荣, 苗香雯, 等. 计算机视觉技术在水产养殖中的应用与展望[J]. 农业工程学报, 2005, 21(8): 174-178.
- [5] Lee D J, Archibald J K, Schoenberger R B, et al. Contour matching for fish species recognition and migration monitoring[M]//Smolinski T G, Milanova M G, Hassanien A E. Applications of Computational Intelligence in Biology. Heidelberg: Springer, 2008:183-207.
- [6] Zion B, Alchanatis V, Ostrovsky V, et al. Classification of guppies' (*Poecilia reticulata*) gender by computer vision [J]. Aquacultural Engineering, 2008, 38(2):97-104.
- [7] Lines J A, Tillett R D, Ross L G, et al. An automatic image-based system for estimating the mass of free-swimming fish[J]. Computers and Electronics in Agriculture, 2001, 31(2):151-168.
- [8] Liang Y T, Chiou Y C. Machine vision-based automatic raw fish handling and weighing system of Taiwan Tilapia [M]//Chien B C, Hong T P, Chen S M, et al. Next-generation Applied Intelligence. Berlin: Springer, 2009: 711-720.
- [9] 刘伟, 谭鹤群, 黄丹, 等. 白鲢质量与截面积沿体长方向分布模型[J]. 农业工程学报, 2012, 28(12):288-292.
- [10] 张超, 徐建瑜, 王文静. 基于机器视觉的梭子蟹质量估计方法研究[J]. 宁波大学学报(理工版), 2014, 27(2):49-51.
- [11] 魏崇德, 陈永寿. 浙江动物志甲壳类[M]. 杭州: 浙江科学技术出版社, 1991:356.
- [12] Rother C, Kolmogorov V, Blake A. "GrabCut": Interactive foreground extraction using iterated graph cuts [J]. ACM Transactions on Graphics, 2004, 23(3):309-314.
- [13] 伊力哈木·亚尔买买提. 基于改进的自适应分水岭图像分割方法研究[J]. 计算机仿真, 2013, 30(2):373-377.
- [14] 方瑞明. 支持向量机理论及其应用分析[M]. 北京: 中国电力出版社, 2007.
- [15] 周辉仁, 郑丕谔, 赵春秀. 基于遗传算法的 LS-SVM 参数优选及其在经济预测中的应用[J]. 计算机应用, 2007, 27(6):1418-1419.
- [16] Gönen M, Alpaydin E. Localized multiple kernel learning [C]. Proceedings of the 25th International Conference on Machine Learning, ACM, 2008:352-359.
- [17] 雷英杰, 张善文. MATLAB 遗传算法工具箱及应用[M]. 西安: 西安电子科技大学出版社, 2014.
- [18] 张婉琳. 遗传算法优化支持向量机的交通流量预测[J]. 激光杂志, 2014(12):116-119.

Weight prediction of swimming crab using computer vision and GA-SVM

TANG Yang-jie, HU Hai-gang*, ZHANG Gang, TANG Chao

(Faculty of Maritime and Transportation, Ningbo University, Ningbo 315211, China)

Abstract: In the paper, computer vision technology is applied to make non-human-interference measurement of swimming crabs. First, the images of swimming crabs under different development stages are obtained by CCD camera. Then, the projected area of the swimming crab can be computed based on the image, so are the full carapace width and carapace length. By examining the positive correlation among the projected area, full carapace width, carapace length and weight of crab, the weight can hence be predicted. The weight prediction model of swimming crab based on Genetic Algorithm (GA) and Support Vector Machine (SVM) is established. The results suggest that the Mean Absolute Percent Error (MAPE) of body weight reads 2.23% , and the Mean Square Error (MSE) is found to be 5.80 g. Compared with BP neural network and SVM, this method shows a better forecasting precision. The experimental results indicate that the method based on computer vision and Genetic Algorithm-Support Vector Machine (GA-SVM) can better predict the body weight of the swimming crabs.

Key words: swimming crabs; weight; computer vision; Genetic Algorithm-Support Vector Machine (GA-SVM); prediction model

(责任编辑 章践立)