

基于流式细胞摄像技术(FlowCAM) 的赤潮藻类识别分析初探*

余肖翰^{1,2}, 曾松福^{1*}, 曹宇峰¹, 陈瑶¹, 谢杰镇², 郑少平²

(1. 国家海洋局 厦门海洋环境监测中心站, 福建 厦门 361008;

2. 厦门大学 信息科学与技术学院, 福建 厦门 361005)

摘要:采用流式细胞摄像系统(Flow Cytometer And Microscope, FlowCAM)技术, 构建了福建南部海域的赤潮生物图谱数据库, 并利用该数据库分别用 VisualSpread 软件自动识别法和描述子分析算法(MATLAB)来鉴定样品, 对两种方法的识别精确度进行了比较。VisualSpreadsheet 专业软件分析能力达到 98.2% 的平均精度, 自主开发的特征描述子分类器达到 94.1% 的平均精度, 拓展了实验设备的应用范围。

关键词:FlowCAM; 图谱数据库; VisualSpread; 形状特征描述子

中图分类号:X835

文献标识码:A

文章编号:1671-6647(2013)04-0515-11

赤潮是在特定的环境条件下, 海水中某些浮游植物、原生动物或细菌爆发性增殖或高度聚集而引起水体变色的一种有害生态现象。近年来, 随着经济和社会的发展, 我国沿海地区赤潮爆发的频率和规模都在增加。赤潮已经成为制约我国沿海经济可持续发展的一个重要因素, 引起了政府和社会各界的广泛关注^[1-2]。如何准确地进行赤潮的监测, 采取有效的防治措施, 减少赤潮造成的危害和损失, 已经成为海洋环保部门的当务之急, 而这些都离不开对浮游生物的鉴定。

流式细胞摄像系统(Flow Cytometer And Microscope, FlowCAM)技术是将流式细胞分析技术与显微成像技术相结合的新兴技术, 它可以在分析流体的同时高分辨率地实时显示和存储每个微粒的数字彩色图像, 通过专业软件分析、筛选、分类、查看相关信息, 快速分辨出水样中藻类, 直观地看到结果。其高效率、高灵敏度、高分辨率和高精度等特点, 使得试验人员可以对赤潮生物细胞进行同步、迅速并且多参数的定量和定性分析, 这对赤潮生物鉴定、浮游生物多样性的研究有着重要意义。

本文章采用流式细胞摄像系统技术采集赤潮藻类样本, 构建福建南部海域的赤潮生物图谱数据库, 在该图谱数据库的基础上, 使用 VisualSpread 软件自动识别法对赤潮藻类样品进行鉴定; 此外还采用另一种分析方式即通过 Matlab 开发的运算描述子来对样品进行二次分析归类, 并对两种方法的识别精确度进行比较。

1 设备与原理

1.1 FlowCAM 系统组成

FlowCAM 流式细胞摄像系统是一套将流式细胞仪和显微镜的功能以及成像和荧光技术结合到一起的

* 收稿日期: 2012-07-01

资助项目: 海洋赤潮灾害立体监测技术与应用国家海洋局重点实验室开放基金——福建省南部海域赤潮生物图谱数据库构建 (MATHAB200901)

作者简介: 余肖翰(1982-), 男, 福建厦门人, 助理工程师, 主要从事计算机赤潮藻类图像识别处理方面研究. E-mail: showhand13@126.com

* 通讯作者: E-mail: zsf6371@126.com

(杜素兰 编辑)

新型的数字式流式细胞分析系统,由三个子系统构成:流控系统,光学系统和电学系统(图1)。

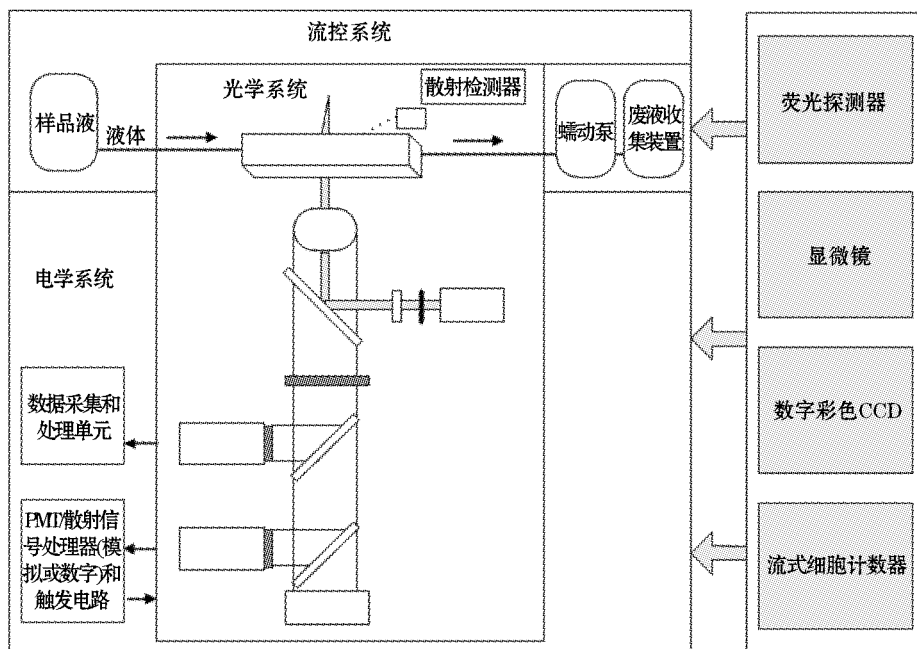


图1 FlowCAM 体系结构示意图

Fig. 1 A sketch map of FlowCAM structure

1.2 技术路线

通过不断摸索,实验人员总结出一套流式细胞摄像系统与显微镜鉴定比对技术路线,以此建立图谱库并对识别率进行计算分析(图2)。

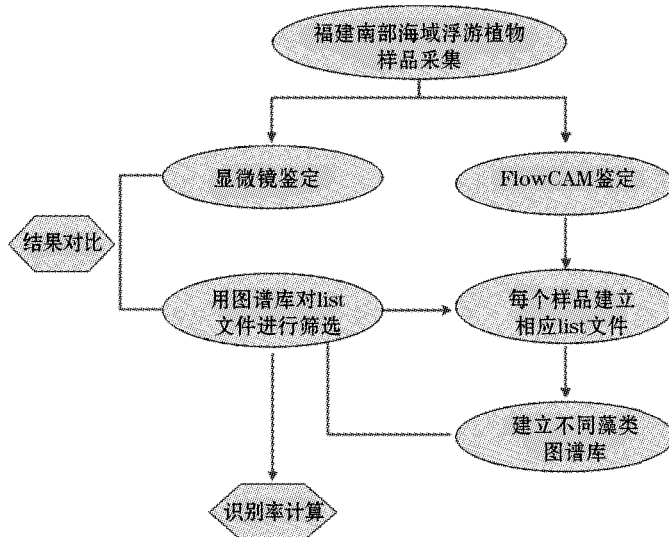


图2 FlowCAM 技术路线图

Fig. 2 Technology Scheme of the FlowCAM

2 福建南部海域赤潮生物图谱数据库的构建

FlowCAM 配置多个放大倍数的显微物镜,根据分析颗粒物的大小范围,针对性选择安装其中某一个,并使用与所安装的物镜相对应的 VisualSpreadsheet 采集软件(表 1)。

表 1 显微物镜和流通池的配对

Table 1 Matching between microobjective and flow cell

编号	流通池深度	放大倍数	监测范围
FC50	50 μm	10 倍/20 倍	3~50 μm
FC100	100 μm	10 倍/20 倍	15~100 μm
FC300	300 μm	4 倍	30~300 μm
FC600	600 μm	4 倍	50~600 μm
FC800	800 μm	2 倍/4 倍	100~800 μm
FC1000	1mm	2 倍	200~1 000 μm

采集模式使用自动拍摄模式见图 3。

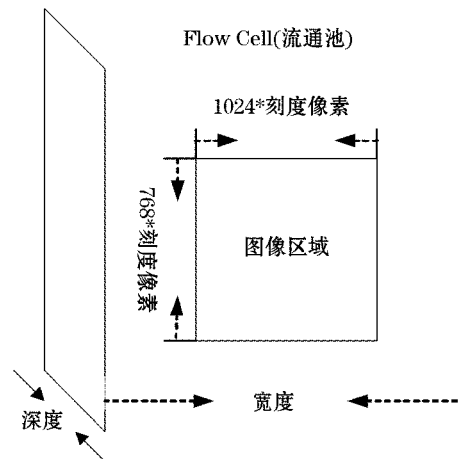


图 3 自动拍摄模式示意图

Fig. 3 Automatically shooting mode

浓度与颗粒采集计算式为

$$M = C / [R \times (E - S) \times (1024 \times F) \times (768 \times F) \times D] \quad (1)$$

式中, M 为粒子浓度; C 为颗粒数; R 为图像采集率; E 为结束时间; S 为开始时间; F 为刻度像素; D 为宽度。

实验组采用课题研究与其它监测任务相结合的方式,采集海洋浮游植物网样与水样用于图谱库的分析和建立。未爆发赤潮时则通过分析其它监测任务所采集的样品进行图片收集工作;赤潮爆发时则迅速派专人到现场采集样品进行分析;赤潮已消退较快而无法派专人现场采集的,则通过福建省其他相关海洋部门收集样品。

以 2010-06-14 日深沪湾爆发的米氏凯伦藻赤潮样品测试为例,当采用 100 μm 孔径的筛绢过滤样品所获得滤液,选择 100 μm 流通池和 20 倍显微物镜,打开相应倍数的软件并设置,进行图片采集(图 4)。具体实验过程中针对不同水样采用了 50,100,150,200,250,300 和 600 μm 等各种孔径的筛绢过滤后的滤液进行分批次鉴别。

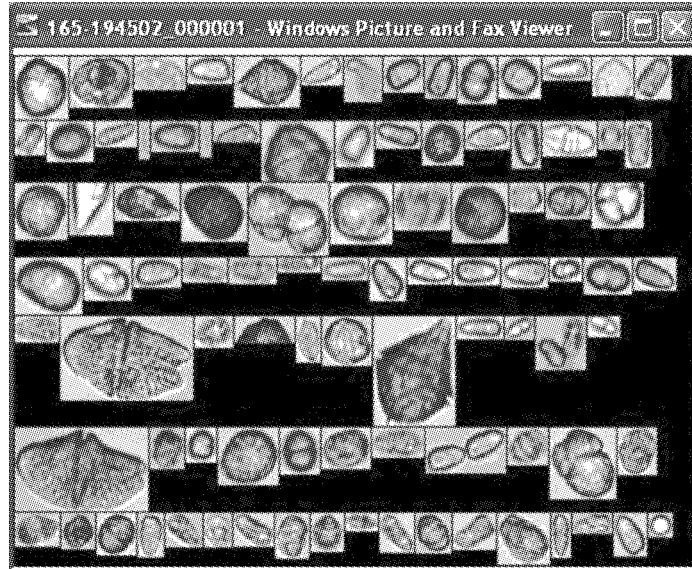


图 4 TIF 拼图文件

Fig. 4 TIF puzzle file

使用 VisualSpread 将采集的藻类进行排列,并在拼图页内显示,将选取图片归到 Libraries(库文件)中。

实验人员选择默认参数作为筛选变量,其中主要关注 ESD、PPC、Aspect Ratio 等重要参数,需进行进一步详细设置(图 5)。

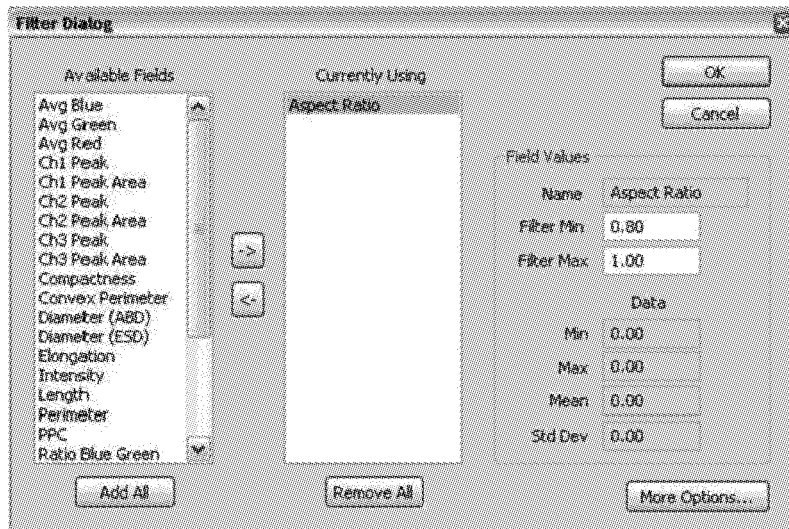


图 5 筛选参数设置

Fig. 5 Settings of screening parameters

根据本次发生赤潮的米氏凯伦藻进行筛选,建立米氏凯伦藻库文件存入图谱数据库中(图 6)。

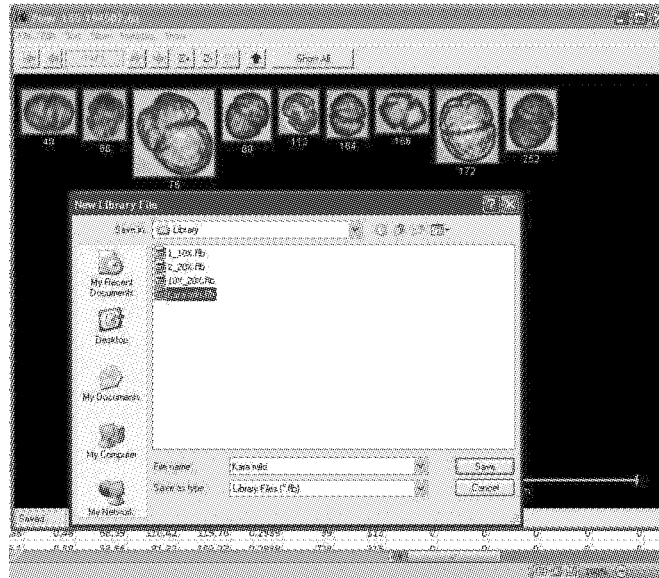


图 6 藻类入库截图

Fig. 6 Storing screenshots of algae

3 描述子算法样本分析

使用 Matlab2009 对数据的多重处理进行进一步的深入开发,通过对藻类样本图像的特征提取并结合 SVM 算法作分类计算^[3-7]。

本研究不对图像预处理(如灰度处理、图像去噪、平滑与锐化的滤波、膨胀与腐蚀等)进行赘述,直接采用处理后的切割图片进行特征分析识别。

3.1 特征提取

使用不变矩特征采用 Hu^[8]构造 7 个经典的不变矩,特征的维数为 7 维。

使用傅里叶描述子进行轮廓特征提取,特征的维数为 64。

使用形态谱算法形态谱提供了 40 个形态特征^[9]。

7 个自定义的形态特征:包括细胞图像的长轴长度(MajorAxisLength)、短轴长度(MinorAxisLength)、周长(Perimeter)、面积(Area)、伸长度(MajorAxisLength/ MinorAxisLength)、矩形度(Area/(MajorAxisLength×MinorAxisLength))和圆度($4 \times \text{Pi} \times \text{Area} / \text{Perimeter}^2$)^[10]。

3.2 藻类图像特征降维

1)通过流式细胞摄像系统采集了各种藻类照片 5 000 余张。目前为止,收集到的可以建库的福建南部典型的赤潮种类共计 15 种,如中肋骨条藻、夜光藻、东海原甲藻、米氏凯伦藻、血红哈卡藻(又名血红阿卡藻,红色赤潮藻^[11])和多纹膝沟藻等(图 7)。

2)对这 6 种福建南部典型赤潮藻类的 1 500 个样本进行特征提取,得到 118 维的组合特征向量。由于不同的特征提取算法,得到的特征向量长度差异较大,实验组又通过归一化多级主特征向量评估准则进行特征降维(图 8)。

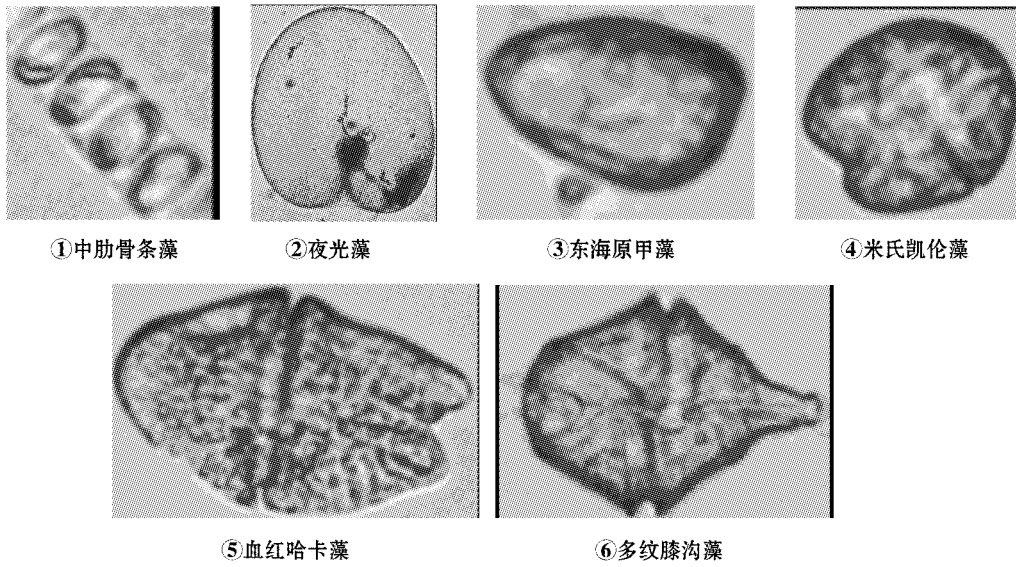


图 7 6 种福建南部海域典型赤潮种类

Fig. 7 Six typical species of red tide algae in the sea area south of Fujian

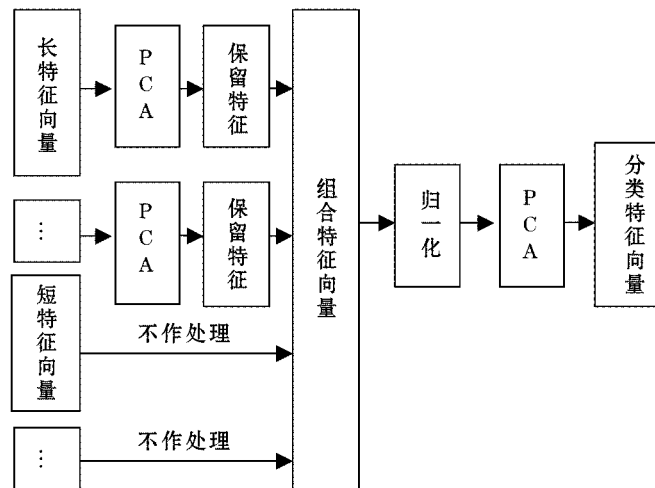


图 8 归一化多级主特征向量评估示意图

Fig. 8 A sketch map of normalized multi-level main eigenvector assess

各特征向量降维前后的长度变化情况见表 2:

表 2 降维处理后各特征向量维数变化表

Table 2 Changes in dimensions of eigenvectors after dimensionality reduction

特征名称	原长度/个	PCA 后长度/个	组合特征向量长度/个	归一化后长度/个	再次 PCA 后长度/个	最终特征向量/个
Hu 矩	7	不处理				
傅里叶描述子	64	7	28	28	10	10
形态谱	40	7				
自定义形态特征	7	不处理				

3.3 SVM-KNN 多值分类器

支持向量机(SVM)是从线性可分情况下的最优分类发展而来,其基本原理可以用二维线性可分的分类问题进行解释。如图 9 所示,实心点和空心点代表两类样本, H 为分类线, H_1 、 H_2 分别为经过各类中离分类线最近的样本点且平行于分类线的直线,它们之间的距离叫做分类间隔。所谓最优分类线就是要求分类线不但能将两类正确分开,而且使分类间隔最大^[12]。

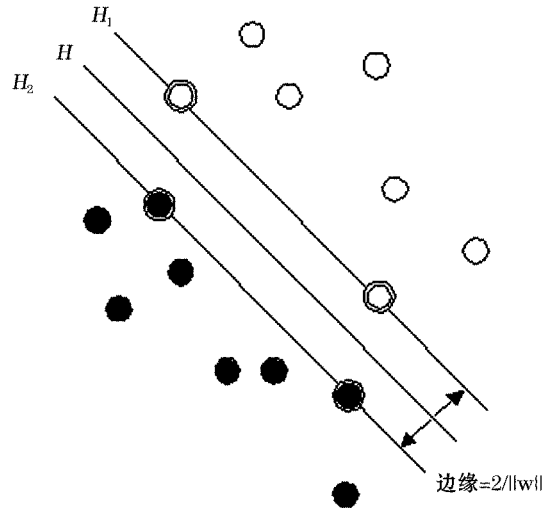


图 9 二维线性可分情况下的最优分类线

Fig. 9 The optimal classification line under the two-dimensional linear separable case

构造方式采取一对一方式,共训练了 6 个 SVM 二值子分类器,所取得函数都是径向基(RBF)函数见式(2):

$$k(x_i, x_j) = e(-r \|x_i - x_j\|^2), r > 0 \quad (2)$$

KNN 算法中 k 取值为 10, ϵ 取值为 0.75, r 为 SVM 的参数 g , c 为惩罚系数。

以径向基(RBF)为核函数的支持向量机作为识别器,参数 C 、 γ 采用上面介绍的网格搜索方法确定,分别取 $C=0.5$, $\gamma=0.0625$ 对 6 种藻类进行分类识别实验。

4 算法比对结果

4.1 VisualSpreadsheet 自动识别结果

VisualSpreadsheet 是 FlowCAM 配套的一套专业控制工作和分析颗粒物的软件,从 1999 年问世至今不断完善着各类自定义粒子成像。由于本课题是使用 VisualSpreadsheet 本身作为图谱数据库的采集、编辑、建库工具,因此图谱数据库中已入库的 1500 个样本不能直接作为比对标准。而在比对过程中,只能使用已有水样并从中取 1500 个清晰图片样本进行分析,其中不可避免因为入库样本与测试水样中藻类颗粒的相似度较高,造成 VisualSpreadsheet 结果表现较好(表 3)。

表 3 VisualSpreadsheet 实验精度表

Table 3 Experimental accuracy of VisualSpreadsheet

项 目	藻类类别/个						精 度/%
	第 1 种	第 2 种	第 3 种	第 4 种	第 5 种	第 6 种	
第 1 次分析	250	0	0	0	0	0	100.0
第 2 次分析	0	250	0	0	0	0	100.0
第 3 次分析	0	0	248	0	1	1	99.2
第 4 次分析	0	0	0	246	1	3	98.4
第 5 次分析	0	0	0	2	241	7	96.4
第 6 次分析	0	0	0	1	11	238	95.2
平均精度							98.2

另外值得注意的是,FlowCAM 系统采用的是多光学镜头配合相应流通池进行图片采集,因此实验过程中仅采取单一物镜取得的数据进行识别分析,避免了因不同物镜造成的同类藻形状分辨率大小差异带来的误差。

4.2 描述子与 SVM 分类器实验结果

实验过程中,对每个样本用 3 种算法从不同的角度描述形状特征,得到的特征向量之间可能存在信息冗余。但之所以使用 3 种方法更重要目的是将特征向量间的信息缺陷进行互补磨平。

通过混淆矩阵可以得到如下精度表(表 4):

表 4 描述子与 SVM 分类器实验精度表

Table 4 Experimental accuracy of descriptor and SVM

项 目	藻类类别/个						精 度/%
	第 1 种	第 2 种	第 3 种	第 4 种	第 5 种	第 6 种	
第 1 次分析	250	0	0	0	0	0	100.0
第 2 次分析	0	238	1	5	4	2	95.2
第 3 次分析	0	0	247	1	0	0	98.8
第 4 次分析	0	0	2	229	11	8	91.6
第 5 次分析	0	0	1	9	227	13	90.8
第 6 次分析	0	1	2	9	18	220	88.0
平均精度							94.1

5 水样测试结果

5.1 图谱库识别精度表

通过已建立的赤潮藻类图谱数据库,对福建南部海域 4 个站位的浮游植物网样进行筛选归类,每个样品选择 300 个清晰图片样本进行分析(不到 300 个样本数则全选)。得出识别精度结果如表 5 所示,可看出特征形态明显的藻类识别精度较高,反之较低。

表 5 图谱库识别精度表
Table 5 Accuracy of Library Recognition

样本号	测试编号	识别精度/%						备注		
		佛氏海毛藻	钟状中鼓藻	叉状角藻	梭角藻	三角角藻	夜光藻	种类	识别个数	样本个数
C1D1120821 植 7	301-182659	69.7	16.7	N/A	N/A	N/A	N/A	佛氏海毛藻	136	195
								钟状中鼓藻	32	192
	301-183150	92.0	17.0	N/A	N/A	N/A	N/A	佛氏海毛藻	276	300
								钟状中鼓藻	51	300
	301-185623	95.3	14.3	N/A	N/A	N/A	N/A	佛氏海毛藻	286	300
								钟状中鼓藻	43	300
DQ210901 植 7	301-163036	84.3	N/A	N/A	N/A	N/A	N/A	佛氏海毛藻	253	300
YY040523 植 7	301-170802	N/A	N/A	98.7	100.00	99.7	N/A	叉状角藻	296	300
								梭角藻	300	300
								三角角藻	1299	300
YY120523 植 7	334-214421	N/A	N/A	98.3	100.00	91.7	78.4	叉状角藻	295	300
								梭角藻	300	300
								三角角藻	275	300
								夜光藻	145	185
YY120523 植 7	334-214730-1	N/A	N/A	97.7	100.00	98.7	87.0	叉状角藻	293	300
								梭角藻	300	300
								三角角藻	296	300
								夜光藻	261	300
YY120523 植 7	334-223645-1	N/A	N/A	97.7	100.00	96.7	91.0	叉状角藻	293	300
								梭角藻	300	300
								三角角藻	290	300
								夜光藻	273	300
YY040523 植 7	302-110918	N/A	N/A	100.00	100.00	100.00	N/A	叉状角藻	96	96
								梭角藻	15	15
								三角角藻	4	4
YY120523 植 7	302-124435	N/A	N/A	100.00	100.00	N/A	N/A	叉状角藻	66	66
								梭角藻	17	17
平均精度		85.4	16.0	98.7	100.0	97.3	85.5	—		

注:“样本个数”是从图谱库筛选的自动识别结果中随机挑选的清晰图片的数量;“识别个数”是对所选取的样本进行人工确认为该种类的图片数量;“识别精度”是“识别个数”与“样本个数”的百分比。

5.2 图谱库筛选结果与显微镜结果比对

选取 2 个福建南部海域站位的浮游植物网样,每个样品分别测试 3 次,用图谱库对几种相应的浮游植物优势种进行筛选归类输出图谱库筛选结果,与相应的显微镜结果进行比对(表 6)。

图谱库筛选结果与显微镜结果的比值变化范围在 32.2%~96.2%之间,可见采用流式细胞摄像系统在计数方面存在较大瓶颈,对浮游植物样品进行筛选以计算其密度还需要进一步的优化和改进。

表 6 图谱库筛选结果与显微镜结果比对
Table 6 Comparison of Library and Microscopy

样本号	种 类	密 度/个·m ⁻³		图谱库结果/ 显微镜结果/%	
		图谱库筛选结果	显微镜结果		
C1D1120821 植 7	佛氏海毛藻	3.90E+05	4.30E+05	90.7	
		2.98E+05	4.30E+05	69.1	
		3.77E+05	4.30E+05	87.5	
	钟形中鼓藻	1.03E+05	3.13E+05	33.1	
		1.93E+05	3.13E+05	61.7	
		3.01E+05	3.13E+05	96.2	
YY120523 植 7	夜光藻	2.67E+04	5.14E+04	51.9	
		2.75E+04	5.14E+04	53.5	
		4.36E+04	5.14E+04	84.8	
	叉状角藻	4.36E+04	1.16E+05	37.7	
		7.20E+04	1.16E+05	62.2	
		9.43E+04	1.16E+05	81.5	
	梭角藻	1.05E+05	2.13E+05	49.3	
		9.23E+04	2.13E+05	43.3	
		1.01E+05	2.13E+05	47.2	
		4.28E+04	1.33E+05	32.2	
		三角角藻	4.56E+04	1.33E+05	34.4
			8.24E+04	1.33E+05	62.0

6 结 论

1) 本项目利用流式细胞摄像系统,初步建立了一种能快速检测和鉴定赤潮生物及其他浮游植物种类的方法,经文献资料与实验收集建立了福建南部海域赤潮生物档案。

使用新建立的检测方法及构建的福建南部海域赤潮生物图谱数据库,能达到快速鉴定的典型赤潮生物至少 5 种,常见赤潮生物 10 种,其他浮游植物 9 种,合计 24 种基本涵盖了福建历史上爆发过的赤潮藻类和常规监测中容易产生优势种的藻类。

这种新的检测方法可以作为福建南部海域赤潮生物鉴定的辅助手段,可在一定程度上减少传统显微镜检的工作量,并能对所得数据加以分析和总结,有助于提高赤潮应急工作效率,对福建南部海域赤潮灾害应急监测与预警预报工作将起到积极的作用。

在图片资源的基础上进一步探讨了形状特征描述子,从不同角度描述了藻类的形状特征,为后续系统的扩展算法提供另一种有效途径。

2) 对于直径 20 μm 以下的,和/或无法与其他藻类明显区分开的(如一个属内的圆筛藻、角毛藻等),最多只能定性到属,而无法到种;由于同属种类形态特征相似无法区分识别,故在福建南部海域曾发生过赤潮的种类——角毛藻属(聚生角毛藻、柔弱角毛藻、旋链角毛藻等)与海链藻属(诺氏海链藻、细弱海链藻等)等藻种已建库备用,暂未投入应用分析。

3) 目前的赤潮图谱数据库还有待完善,截止课题完成,图谱库中 24 种藻种的有效样本合计已达 4 万余个,初步满足赤潮藻类分析的要求。但通过流式细胞摄像技术来识别赤潮藻类只能作为辅助技术,要将该技术业务化还需今后大量的分析研究;需要继续增加藻类水样的分析以获得更多的图谱丰富数据库,并在此基础上广泛全面的开发其他的形状特征描述子,如:Sobel 算子在像素加权的特点、Canny 算子在边缘分割的优势、Laplacian 算子增强高频又能保持低频不变等,这些都是很有特点且行之有效的图形分析方法;而近年自主学习算法的实际应用(如谷歌的虚拟脑技术)更是将图形分析上升到一个新的高度。

如何有效避免目前已有技术的局限性,寻求更全面的解决方案应用于赤潮藻类识别分析,开发出更准确的运算描述方法是实验室未来的发展方向。

致谢:本研究的完成得到导师及厦门大学信息与技术学院王博亮教授的指导,并提出宝贵意见,在此表示真挚的感谢。

参考文献(References):

- [1] ZHAO D Z, ZHAO D L, ZHANG F S, Type of Formation, Distribution And Temporal Trend of Red Tides Occurred In The China Sea [J]. Marine Environmental Science, ,2003, 22(3):7-11. 赵冬至,赵冬玲,张丰收. 我国海域赤潮灾害的类型、分布与变化趋势[J]. 海洋环境科学,2003, 22(3):7-11.
- [2] CULVERHOUSE P, WILLIAMS R, REGUERA B, et al, Expert and machine discrimination of marine flora: a comparison of recognition accuracy of field collected phytoplankton[J]. IEEE International Conference on Visual Information Engineering, 2003, 7:177-181.
- [3] LIU F, CAO G H, TANG Z M, Real-time Human's Shape Monitoring & Recognition System Using Moment Invariants[J]. Computer Engineering, 2002, 28(9):80-83. 刘菲,曹冠华,唐振民. 基于不变矩的实时人形识别研究[J]. 计算机工程,2002, 28(9):80-83.
- [4] ZAHN C T, ROSKIES R Z. Fourier Descriptors for Plane Closed Curves[J]. IEEE Transactions On Computers,1972, C-21(3):269-281.
- [5] RUI Y, SHE A C, HUANG T S. Modified Fourier Descriptors for Shape Representation-A Practical Approach. [C]. [s. l.]:[s. n.], 1996:47-12.
- [6] DUAN L J, A Survey of Shape Feature[J]. Computer Science,2007. 34(8): 215-218,247. 段立娟. 形状特征的编码描述研究综述[J]. 计算机科学,2007. 34(8): 215-218,247.
- [7] GONZALEZ R C, WOODS R E. Digital Image Processing, Second Edition. [M]. Beijing: PTPRSS,2003.
- [8] HU M K. Visual pattern recognition by moment invariants[J]. IRE Transaction on Information Theory,1962,8(2):179-187.
- [9] TANG X. Texture Information in Run-Length Matrices[J]. IEEE Transaction on Image Processing, 1998, 7(11):1602-1609.
- [10] ZHENG S P, XIE J Z, WANG B L, et al. Red Tide Dominant Species Recognition Research Using Shape Descriptors[J]. Journal of Xiamen University (Nature Science), 2009, 48(5):676-679. 郑少平,谢杰镇,王博亮,等. 基于形状特征描述子的赤潮优势种识别研究[J]. 厦门大学学报(自然科学版),2009,48(5):676-679.
- [11] HUANG Z G, LIN M, China Marine species with atlas[M]. Beijing: Ocean Press, 2012. 黄宗国,林茂. 中国海洋物种和图集[M]. 北京:海洋出版社,2012.
- [12] LI R, YE S W, SHI Z B. SVM-KNN Classifier——A New Method of Improving the Accuracy of SVM Classifier[J]. Acta Electronica Sinica, 2002,30(5):745-748. 李蓉,叶世伟,史忠柏. SVM-KNN 分离器——一种提高 SVM 分类精度的新方法[J]. 电子学报, 2002,30(5):745-748.

Identification of Red Tide Algae Based on FlowCAM

YU Xiao-han^{1,2}, ZENG Song-fu¹, CAO Yu-feng¹, CHEN Yao¹, XIE Jie-zhen², ZHENG Shao-ping²

(1. *Xiamen Marine Environmental Monitoring Central Station of SOA*, Xiamen 361008, China;

2. *College of Science and Technology, Xiamen University*, Xiamen 361005, China)

Abstract: An image database of red tide algae in the sea area south of Fujian is constructed by means of Flow Cytometer and Microscope (FlowCAM), with which the samples are identified by using two methods: the automatically identifying by applying VisualSpreadsheet software and the feature extracting by using a character descriptor classifier which was self-developed through MATLAB. The comparison between the two methods indicates that the application of the VisualSpreadsheet professional software can reach to an average accuracy of 98.2%, and the self-developed character descriptor classifier can achieve an average accuracy of 94.1%. These results have expanded the application scope of the experimental equipment.

Key words: FlowCAM, image database, VisualSpreadsheet, shape-character descriptor

Received: July 1, 2012