

学校编码: 10384
学号: 23020061152423

分类号____密级____
UDC_____

厦 门 大 学

硕 士 学 位 论 文

赤潮生物图像分类识别技术研究

The Research of Red Tide Algae Images Classification and
Recognition Technology

刘君君

指导教师姓名: 王博亮教授

专业名称: 计算机系统结构

论文提交日期: 2009年4月

论文答辩时间: 2009年6月

学位授予日期: 2009年 月

答辩委员会主席: _____

评 阅 人: _____

2009年4月

厦门大学博硕士学位论文摘要库

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果,均在文中以适当方式明确标明,并符合法律规范和《厦门大学研究生学术活动规范(试行)》。

另外,该学位论文为()课题(组)的研究成果,获得()课题(组)经费或实验室的资助,在(ATR)实验室完成。(请在以上括号内填写课题或课题组负责人或实验室名称,未有此项声明内容的,可以不作特别声明。)

声明人(签名):

年 月 日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文(包括纸质版和电子版)，允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

()1. 经厦门大学保密委员会审查核定的保密学位论文，于
年 月 日解密，解密后适用上述授权。

()2. 不保密，适用上述授权。

(请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。)

摘 要

赤潮是我国近海常见的重要灾害之一，不仅造成了重大的经济损失，而且对海洋生态环境、资源和公众健康构成了严重威胁。目前只能通过监测和预报的手段来减少赤潮造成的损失，因此建立赤潮生物的有效监测方法成为摆在我们面前的亟待解决的课题。传统的赤潮监测方法是通过显微镜进行人工辨认和计数。这种方法不仅存在劳动强度大、效率低等问题，而且赤潮生物由于形态相近难于分辨，因此需要经验丰富的专家才能进行分类识别。这些因素都严重影响了对赤潮灾害预测的反应时间，十分不利于赤潮减灾防灾。

本文的研究工作在于针对赤潮生物提出具有较高准确率的实时自动分类方法。首先对赤潮生物图像原始数据集进行特征分析，并在此基础上，对原始特征集进行特征选择以去除特征集中的无关特征和冗余特征，得到最优特征子集，然后分别讨论和分析了 SVM 和 KNN 两种分类器在最优特征子集上的分类效果，最后提出了使用 SVM-KNN 分类器来进行赤潮生物图像的识别分类。

本文研究工作的主要内容及创新包括以下几点：

(1) 在对数据原始特征集分析的基础上，提出了将 ReliefF 算法与顺序后向搜索（SBS）策略相结合进行特征选择的方法。使用该方法可以有效的去除原始特征集中的无关特征和冗余特征，减少它们对分类器分类精度的影响。并通过实验对比了 SVM 和 KNN 两种分类器特征选择前后的分类效果。

(2) 将经过特征选择后的 4 类样本数据集和 7 类样本数据集，分别用 SVM 和 KNN 两种分类器进行分类实验，针对实验结果对两种分类器分类的特点和性能进行深入的讨论和分析。

(3) 对 SVM 分类器和 KNN 分类器的理论进行了研究，发现了两种分类器的理论上的共通点，找到了将两种分类器结合起来的理论依据，提出了使用 SVM-KNN 分类器来改善赤潮生物图像的分类效果，并通过实验证明 SVM-KNN 分类器确实可以有效的提高分类性能。

关键词： 赤潮；特征选择；支持向量机；K 近邻分类器

Abstract

The Red Tide is one of the serious familiar disasters along the offing, which not only causes great loss financially, but holds severe threat against oceanic environment and its resources as well as public health. While the loss led by the Red Tide could only be lessened through inspecting and forecast at present. Therefore, it is urgent for us to have an effective inspecting method for the Red Tide Algae. The traditional inspecting method is to employ microscopes to make identification and take count, by which there are not only intense labor and inefficiency, but also it is hard to differentiate due to the close forms of the Red Tide. So the task of classification and identification could be accomplished by experienced experts only. All of these factors thus would have severe impact upon the responsive time to the forecast of the Red Tide, quite unfavorable to the lessening and preventing of the disaster of the Red Tide.

This paper mainly studies to bring forward a method of a veracious, real-time automatic classification for the Red Tide Algae. Firstly it makes a feature analysis of the original data of Red Tide Algae Images and receives the optimal feature subset by a feature selection among the feature set with the intension to remove those unrelated and redundant features. Different classifying effects of SVM and KNN on the optimal feature subset are discussed and analyzed respectively afterwards. Finally the paper claims to use the SVM-KNN classifier to make the identification and classification of Red Tide Algae Images.

The main points as well as the creative ones in the paper are included in the following:

(1) Based on the data analysis of the original feature set, the paper brings forward the feature selection method by combining the ReliefF algorithm and sequential backward search strategy, which can remove the unrelated and redundant features in the original feature set and lessen the impact upon the precision of classifiers. Besides, the experiment data show the comparison of the classifying effects both before and after classifying by SVM and KNN.

(2) SVM and KNN are employed respectively to conduct classificatory experiments on four types of sample data set and seven types of sample data set received respectively after feature selection. Features and properties of the two

classifiers are later discussed and analyzed according to the results of the experiments.

(3) The paper conducts study of the theories of SVM and KNN classifiers and its theoretically common points have been discovered. Hence the theoretical foundation of combining both classifiers is obtained and so the paper claims to adopt the SVM-KNN classifier to improve the classifying effect on the red tide algae image. Furthermore, experiment data prove the truth that the SVM-KNN classifier can enhance the classifying performance effectively.

Key words: Red Tide; Feature Selection; Support Vector Machine; K-Nearest Neighbors Classifier

厦门大学博硕士学位论文摘要

厦门大学博硕士学位论文摘要库

目 录

第一章 绪论	1
1.1 课题的背景和意义	1
1.2 赤潮生物图像分类识别技术的国内外研究现状	2
1.3 本课题的研究工作及创新之处	4
1.4 本文的组织结构	5
第二章 赤潮生物图像的获取与特征提取	7
2.1 赤潮生物物流式图像采集控制系统简介	7
2.1.1 流路控制系统	7
2.1.2 图像采集系统	8
2.1.3 荧光采集系统	9
2.1.4 赤潮生物物流式图像采集控制系统结构框图	10
2.2 赤潮生物图像的检测	11
2.2.1 灰度信息	11
2.2.2 纹理信息	12
2.2.3 尺度信息	12
2.3 赤潮生物图像的特征提取	12
2.3.1 基于矩的特征	12
2.3.2 纹理特征	13
2.3.3 多分辨率高斯差特征	13
2.3.4 基本形状特征	14
2.4 实验选取的赤潮生物图像	14
2.5 本章小结	15
第三章 赤潮生物图像的特征选择	17
3.1 特征选择问题的提出	17
3.2 特征选择理论	18
3.2.1 特征降维概述	18

3.2.2 特征选择的三种方式.....	19
3.3 特征选择方法.....	21
3.3.1 ReliefF 算法.....	21
3.3.2 顺序后向搜索 (SBS) 策略.....	24
3.3.3 ReliefF-SBS 算法.....	24
3.3.4 BW 算法.....	26
3.4 赤潮生物图像的特征选择结果与分析.....	26
3.4.1 特征选择实验数据.....	26
3.4.2 ReliefF-SBS 特征选择结果与分析.....	27
3.4.3 BW 特征选择结果与分析.....	30
3.4.4 特征选择对 SVM 分类器的影响.....	31
3.5 本章小结.....	32
第四章 赤潮生物图像的分类识别.....	33
4.1 模式识别概述.....	33
4.2 支持向量机(SVM).....	34
4.2.1 支持向量机基本理论.....	34
4.2.2 多分类支持向量机.....	38
4.2.3 分类识别的实验数据与 SVM 软件包.....	40
4.2.4 多分类支持向量机的选择.....	41
4.3 K 近邻理论.....	48
4.3.1 最近邻原理和 K 近邻算法.....	48
4.3.2 KNN 分类结果与分析.....	49
4.4 SVM-KNN 分类器.....	51
4.4.1 SVM-KNN 分类器的原理.....	52
4.4.2 SVM-KNN 分类器算法.....	53
4.4.3 SVM-KNN 分类结果与分析.....	53
4.5 本章小结.....	55
第五章 总结与展望.....	57
参考文献.....	59

硕士期间所参与的科研课题与发表论文	63
致 谢	64

厦门大学博硕士论文摘要库

Contents

Chapter1 Introduction	错误！未定义书签。
1.1 Background	错误！未定义书签。
1.2 Research Status at Home and Abroad	错误！未定义书签。
1.3 Research and Innovation	错误！未定义书签。
1.4 Structure of Thesis	错误！未定义书签。
Chapter2 Capture of Red Tide Algae Images and Feature Extraction	错误！未定义书签。
2.1 Introduction to The Fluorescence Detection and Analysis System for the Flowing Red Tide Algae	错误！未定义书签。
2.1.1 The Flow Control System.....	错误！未定义书签。
2.1.2 The Image Capture System.....	错误！未定义书签。
2.1.3 The Fluorescence Detection System.....	错误！未定义书签。
2.1.4 Structure Diagram of The Fluorescence Detection and Analysis System for the Flowing Red Tide Algae.....	错误！未定义书签。
2.2 Detection of Red Tide Algae Images	错误！未定义书签。
2.2.1 Gray Information.....	错误！未定义书签。
2.2.2 Texture Information.....	错误！未定义书签。
2.2.3 Scale Information.....	错误！未定义书签。
2.3 Feature Extraction of Red Tide Algae Images	错误！未定义书签。
2.3.1 Moment Invariant Feature.....	错误！未定义书签。
2.3.2 Texture Feature.....	错误！未定义书签。
2.3.3 Multi-resolution Gaussian difference Feature.....	错误！未定义书签。
2.3.4 Basic Shape Feature.....	错误！未定义书签。
2.4 Experimental Images of Red Tide Algae	错误！未定义书签。
2.5 Summary of Chapter 2	错误！未定义书签。
Chapter3 Feature Selection of Red Tide Algae Images	错误！未定义书

签。

3.1 Feature Selection Problem	错误！未定义书签。
3.2 Theory of Feature Selection	错误！未定义书签。
3.2.1 Introduction to Feature Dimension Reduction.....	错误！未定义书签。
3.2.2 The ways of Feature Selection.....	错误！未定义书签。
3.3 Methods of Feature Selection	错误！未定义书签。
3.3.1 ReliefF Algorithm.....	错误！未定义书签。
3.3.2 Sequential Backward Search (SBS) strategy.....	错误！未定义书签。
3.3.3 ReliefF-SBS Algorithm.....	错误！未定义书签。
3.3.4 BW Algorithm (ratio of variables between categories to within categories sum of squares)	错误！未定义书签。
3.4 Results and Analysis of Feature Selection of Red Tide Algae Images .	错误！未定义书签。
3.4.1 Experimental Data of Feature Selection	错误！未定义书签。
3.4.2 Results and Analysis of ReliefF-SBS Algorithm.	错误！未定义书签。
3.4.3 Results and Analysis of BW Algorithm.....	错误！未定义书签。
3.4.4 Feature Selection on the impact of SVM Classifier	错误！未定义书签。
3.5 Summary of Chapter 3	错误！未定义书签。
Chapter4 Classification of Red Tide Algae Images	错误！未定义书签。
4.1 Introduction to Pattern Recognition	错误！未定义书签。
4.2 Support Vector Machine(SVM)	错误！未定义书签。
4.2.1 Theory of Support Vector Machine.....	错误！未定义书签。
4.2.2 Multi-class Classification with Support Vector Machines	错误！未定义书签。
4.2.3 Experimental Data and SVM software	错误！未定义书签。
4.2.4 Selection of Multi-class Classification with Support Vector Machine	错误！未定义书签。
4.3 Theory of K nearest neighbors	错误！未定义书签。
4.3.1 Principle of nearest neighbor and KNN Algorithm	错误！未定义书签。

4.3.2 Results and Analysis of KNN Classification	错误！未定义书签。
4.4 SVM-KNN Classifier.....	错误！未定义书签。
4.4.1 Theory of SVM-KNN Classifier.....	错误！未定义书签。
4.4.2 SVM-KNN Algorithm	错误！未定义书签。
4.4.3 Results and Analysis of SVM-KNN Classification	错误！未定义书签。
4.5 Summary of Chapter 4.....	错误！未定义书签。
Chapter5 Summary and Outlook	错误！未定义书签。
Reference	错误！未定义书签。
Achievements	错误！未定义书签。
Thanks	错误！未定义书签。

厦门大学博硕士学位论文

第一章 绪论

1.1 课题的背景和意义

赤潮是一种海洋生态灾害，指的是一些在海洋中浮游生活的赤潮生物（主要是浮游植物或原生动物）在短时间内暴发性繁殖或高度聚集、且引起水色异常的现象。以浮游植物来说，在世界各地已报道的 4000 多种微藻中，能引起赤潮的种类约有 200 多种，而其中大约有 70 多种微藻能产生毒素。赤潮不仅严重破坏海洋生态平衡，恶化海洋环境，危害海洋水产资源，危及海洋生物，甚至威胁着人类的健康和生命安全，已引起许多国家的严重关注^{[1][2]}。

1990 年联合国将赤潮列为世界三大近海污染问题之一。国际海洋考察理事会于 1992 年发表了“有害赤潮对海水养殖业和海洋渔业影响”的报告。为加强全球范围赤潮的研究和监测，联合国教科文组织的政府间海洋学委员会、国际科联海洋学研究会、联合国粮农组织等均成立了赤潮研究专家组或工作组，制定赤潮研究或监测计划。美国总统克林顿于 1998 年 11 月还签署了有毒微藻水华和缺氧研究和控制的行动纲领。我国于 1985 年在广州成立了“南海赤潮研究中心”，1990 年成立了 IC0 / SCOR 有害赤潮专家组中国委员会。除了对赤潮进行监测，对赤潮生物的分类、生态、生理及赤潮毒素进行大量调查研究外，近几年还利用遥感等先进技术开展赤潮的预测、预报。近年来，我国也开始着手制定有害赤潮生态学与海洋学研究战略计划^[3]。

随着我国近岸海域富营养化程度的加剧，赤潮爆发频率不断增加，规模不断扩大，新赤潮藻种不断出现，近岸海域赤潮灾害日趋严重。据不完全统计，从 70 年代以来，我国赤潮发生频率以每 10 年增加三倍的速度上升，1998 年至今，每年都发生了面积超过 1000 平方公里、世界罕见的特大赤潮，造成了重大的经济损失，对海洋生态环境、资源和公众健康构成了严重威胁。1998 年—2001 年，在渤海、东海都发生了面积达到几千平方公里的特大赤潮，这在国际上都是十分罕见的^[4]。2002 年—2005 年，我国海域每年的赤潮累计面积从 10000 平方公里迅速上升到 27,070 平方公里，直接经济损失由 2300 万元扩大到 6900 万元。赤潮

灾害已经引起了国家的高度重视，国家海洋局采取了一系列有效措施，先后在我国近海建立了 19 个赤潮监控区，制定了赤潮应急监测方案，同时加强了赤潮监测的技术手段和能力建设，使得我国海域的赤潮得到了一些控制，2007 年赤潮累计面积约 11,610 平方公里，直接经济损失为 600 万元。赤潮频发预示着我国赤潮高发区的海洋生态环境已经受到了严重的干扰，生态系统的正常结构和功能可能已经或正在被改变，而且生态环境一旦失衡恶化，将很难在短期内恢复。目前有害赤潮已成为制约我国沿海经济可持续发展的一个重要因素。

因此，如何准确地进行赤潮监测，以便及时采取有效防治和减灾措施，减少赤潮造成的危害和损失，已成为海洋环境保护工作的当务之急。针对日益严重的赤潮灾害，美国、日本、加拿大等国家已分别启动了国家赤潮监测研究计划，通过对赤潮的产生、发展、维持和消亡过程进行监视、监测，获取赤潮的动态信息，及时发出信息，采取相关的措施，减少灾害造成的损失。我国也在国家科技攻关计划和国家 973 计划中，增设赤潮专项研究课题，进行有关赤潮爆发机理研究、检测技术研究和预测及治理工作^{[5][6]}。

1.2 赤潮生物图像分类识别技术的国内外研究现状

赤潮生物隶属于海洋浮游生物，是海洋浮游生物的一个组成部分。利用数字图像处理技术进行海洋浮游生物的分析研究开始于 20 世纪 70 年代，初期的研究应用主要是将计算机视觉系统与常规显微镜相结合，产生更适合于人观察和识别的图像，并采用简单的图像处理方法对显微镜影像中的浮游生物的外型轮廓尺寸大小（直径和面积等）进行简单测量、分类、记数^[7]，因此这种技术方法依赖专业人员进行浮游生物的种、属识别和标注。

Rodenacker 等利用决策树对浮游生物形态、灰度和不变矩等特征进行识别分类研究^[8]。但是该文只是基于用于分类识别的全局图像特征和目标特征对分类的贡献进行了讨论，并没有给出具体的分类识别率，难以对其分类效果进行比较。

欧美国家的一些硅藻学家在 20 世纪末期就开始硅藻自动鉴别系统的研究，利用标准的硅藻图版的形体、壳面文饰等等建立一套数据库，通过对显微镜目镜下的目标形态在电脑里的扫描结果，自动寻找与之最接近的属种，从而进行定名。ADIAC(Automatic Diatom Identification and Classification)系统是一项非

Degree papers are in the "[Xiamen University Electronic Theses and Dissertations Database](#)". Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.

厦门大学博硕士论文摘要库