

文章编号: 1674-8085(2017)01-0060-05

基于 PCA/ICA 的海洋浮游生物识别

*肖晓红^{1,2}, 刘冬生^{1,2}, 欧阳春娟^{1,2}

(1. 井冈山大学电子与信息工程学院, 江西, 吉安 343009; 2. 流域生态与地理环境监测国家测绘地理信息局重点实验室, 江西, 吉安 343009)

摘 要: 通过海洋浮游生物图像对浮游生物进行识别和分类, 有助于合理地开发、管理、利用海洋资源。针对直接使用 PCA 或 ICA 形成的特征空间, 不能达到各类别元素未能较好分离的问题, 提出了一种浮游生物识别算法。采用 PCA 和 ICA 相结合的多层混合分类模型, 提取 WHOI 数据集浮游生物样本特征, 放入到分类模型中进行测试, 并对不同距离度量的识别效果进行了比较。结果表明基于耦合度量的混合分类模型具有良好的分类性能。

关键词: 模式识别; 浮游生物; 主成分分析 (PCA); 独立成分分析 (ICA); 有监督学习

中图分类号: TP391.4

文献标识码: A

DOI:10.3969/j.issn.1674-8085.2017.01.012

PCA/ICA BASED MARINE PLANKTON RECOGNITION

*XIAO Xiao-hong^{1,2}, LIU Dong-sheng^{1,2}, OUYANG-Chunjuan^{1,2}

(1. School of Electronics and Information Engineering, Jinggangshan University, Ji'an, Jiangxi 343009, China;

2. Key laboratory of watershed ecology and geographical environment monitoring, NASG, Ji'an, Jiangxi 343009, China)

Abstract: The recognition and classification of marine plankton via its image can facilitate the development, management and utilization of marine resources. According to a direct implementation of PCA or ICA fail to separate the points in different classes in the eigenspace, we propose an algorithm for plankton recognition combining PCA and ICA to form a multi-layer classification mechanism. An experiment on plankton samples collected by WHOI is conducted to test the performance of our model, in which different features as well as measures of distance are explored. The result shows that the hybrid model using coupled measure outperforms others, and we further depict the shape of several sea planktons.

Key words: pattern recognition; marine plankton; principal component analysis; independent component analysis; supervised learning

0 引言

海洋浮游生物识别是一个新颖的问题, 有助于提高人们对海洋浮游生物种类的认识, 近来受到了科学家广泛的关注。海洋浮游生物识别与图像识别的算法和步骤类同, 主要分为预处理, 特征提取以

及模式分类等阶段。在分类器方面, 传统的 K-最近邻 (KNN) 算法忽略了类别内元素在特征空间中的整体特性。目前, 海洋浮游生物识别的方法可以分为以下几类: 特征提取方式有主成分分析 (PCA); 独立成分分析 (ICA); 线性判别式 (LDA)。分类方法: 有 K-最近邻 (KNN); 决策树; 人工神经网络; 朴素贝叶斯 (Naïve Bayesian); 支持向量机

收稿日期: 2016-07-12; 修改日期: 2016-12-14

基金项目: 国家自然科学基金项目(61462046); 江西省教育厅科学技术研究项目基金(GJJ14559; GJJ13553); 江西省科技厅自然科学基金项目基金(20151BAB207026); 井冈山大学博士科研启动项目(JZB1311)资助。

作者简介: *肖晓红(1971-), 女, 江西吉安人, 副教授, 硕士, 主要从事机器学习和计算机应用等研究(E-Mail:jgsxxh5659@126.com);

刘冬生(1969-), 男, 江西安福人, 讲师, 主要从事网络和算法研究(E-Mail:644973997@qq.com);

欧阳春娟(1974-), 女, 江西吉安人, 副教授, 博士, 主要从事图像处理, 信息隐藏等方面的研究(E-mail:oycj001@163.com).

(SVM)等。特征提取方面, Kirby 和 Sirovich 最先使用 PCA 进行数据压缩和人脸特征化, Turk 和 Pentland 将 PCA 推广到了人脸识别^[1-2]。他们使用 PCA 计算出训练人脸集的一组子空间的基向量(特征脸),然后将人脸集里的图像投影到该子空间。Yamvor 等^[3]详细讨论了 PCA 算法中特征向量的选取方式和对不同距离度量进行了评估。PCA 寻找使输入信号各分量方差最大的方向,然而对于输入非高斯分布的信号,其特征信息可能蕴含在更高阶矩中,因此使用 PCA 易导致信息的丢失。考虑到 PCA 需要输入服从高斯分布,若不是,则特征值和特征向量不能代表数据的特征,这样 PCA 失去了意义。于是 Comon^[4]提出了独立成分分析(ICA), Bartlett 等^[5]将 ICA 应用到人脸识别中。ICA 将信号分解为若干个统计独立分量的线性组合,从而保留了信号的高阶统计特性。Belhumeur 等^[6]在文献中提出了使用 LDA 的方法进行特征提取。LDA 的思想最小化类内距离,最大化类间距离。这种方法对光照,面部表情不敏感,且经测试准确率优于 PCA 的方法。分类方面,传统方法是 KNN 算法,实现容易,但效果一般。C. Cortes 和 Vapnik 等^[7]首先提出了支持向量机(SVM),通过使用非线性映射算法将低维输入空间的输入样本映射到高维特征空间使其线性可分。然而 SVM 主要应用于二元分类,对多元分类表现不佳。近年来随着深度学习(Deep Learning)的兴起,深度卷积神经网络(CNN)被广泛使用于模式识别中^[8]。神经网络具有强大的非线性计算能力,但是其理论还有待进一步完善。在已有带标签的源数据条件下进行分类是一种有监督学习,即能通过训练资料中学到或建立一个模式(learning model),并依此模式推测新的实例。

针对已有方法存在的问题,我们提出了一种优化的多层混合分类机制,利用多层分类器的结构,将 PCA 和 ICA 相结合,找到了解决 PCA 和 ICA 单独使用识别率不高问题的有效途径。用 Woods Hole Oceanographic Institution (WHOI) 的 Imaging FlowCytobot (IFCB) 所采集的浮游生物的图像数据^[9]对新方法进行了测试,结果表明,使用该方法进行浮游生物检测的准确率相比普通 KNN 算法有所提高。

1 模型

测试使用的数据为 Woods Hole Oceanographic Institution (WHOI) 的 Imaging FlowCytobot (IFCB), 所采集的浮游生物图像数据。在 WHOI 数据集中已将浮游生物的图片分成了 103 类, 移除其中分类信息不够明确的少数几类数据, 预处理之后的数据作为测试集和训练集。

将输入图像经预处理得到尺寸为 $D = M \times N$ 的图片, 故输入数据可以看作实数集 R 空间 D 维随机变量 X , $X \in R^D$ 的观测集 $\{x_n\}$, $n = 1, \dots, S$, 每次观测结果出现概率均为 $1/S$, 可以先假设 X 服从 D 维高斯分布。每个 x_n 带有一个标签。该问题化为, 根据随机变量 X 的观测集 $\{x_n\}$ 对与观测集近似同分布的输入观测值 x 进行分类。

该分类过程可以被划分为两个阶段:

- 1) 对带标签的已知观测集进行预处理和特征提取。
- 2) 根据已有训练集中的标签和特征, 对输入的未知类型进行分类。

因此, 下文对数据的特征提取和分类算法进行介绍。

2 基本理论和方法

2.1 特征提取

2.1.1 主成分分析(PCA)

D 维随机变量 X , $E(X) = 0$ 。寻找使 X 在其方向投影的方差:

$$E(|\alpha^T|^2) = \alpha^T \Sigma \alpha \quad (1)$$

最大的方向 α , Σ 为 X 的协方差阵。容易知道 α 为 Σ 的特征向量 Φ_i , 即

$$E\phi_i = \lambda_i \phi_i \quad (2)$$

方差为特征向量 Φ_i 对应的特征值 λ_i 。将 X 投影到它的特征空间, 考虑到 α 的正交性, 有

$$\Omega = \Phi^T X \quad (3)$$

其中 $\phi = [\phi_1, \dots, \phi_D]$

通过保留最大的前 M 个特征值及其对应的特征向量, 将 X 投影到一个 R^D 的一个 M 维子空间中, 可实现数据压缩。可以证明, 对于 M 一定时, 以 PCA 这种方式进行降维在均方意义下是最优

的。

2.1.2 独立成分分析 (ICA)

假设 D 维随机变量 X 是 D 维各分量互相独立的随机变量 S 的线性组合, 即

$$X=AS \quad (4)$$

其中, A 是 D 维方阵。期望找到一个 D 维方阵 W 使得

$$U=WX \quad (5)$$

其中 U 是 S 的估计值。矩阵 W 的估计算法有 InfoMax、FastICA、JADE、Kernel-ICA 等。

结合 PCA 进行降维, X 的估计值为 $\tilde{X} = \Phi\Omega$, 可以有两种方式进行 ICA。

方法一: 将输入图像看作独立的图像的线性组合, 对 PCA 的特征向量进行 ICA $U = \Phi^T$

方法二: 将输入图像看作独立的像素的线性组合, 对 PCA 的特征空间坐标进行 ICA, $U=W\Omega$

2.2 分类方法

2.2.1 K 邻近 (K-Nearest Neighbor)

这是最基本的分类方法。如果一个样本在特征空间中的 k 个最相似 (即特征空间中最邻近) 的

样本中的大多数属于某一个类别, 则该样本也属于这个类别。

2.2.2 支持向量机 (SVM)

SVM 针对线性不可分情况, 通过使用非线性映射算法将低维输入空间的输入样本映射到高维特征空间使其线性可分。SVM 学习问题可表示成凸优化问题, 因此可以取得全局最优化。

3 研究方法

本文首先基于传统架构从以下两个角度对浮游生物分类进行了研究。

1) 游生物特征提取方法

2) 浮游生物分类方法

下面是本文提出的一种优化的多层混合分类机制, 如图 1 所示。考虑到浮游生物种类间的内在的联系和区别对传统模式进行了改进, 取得了可观的效果。

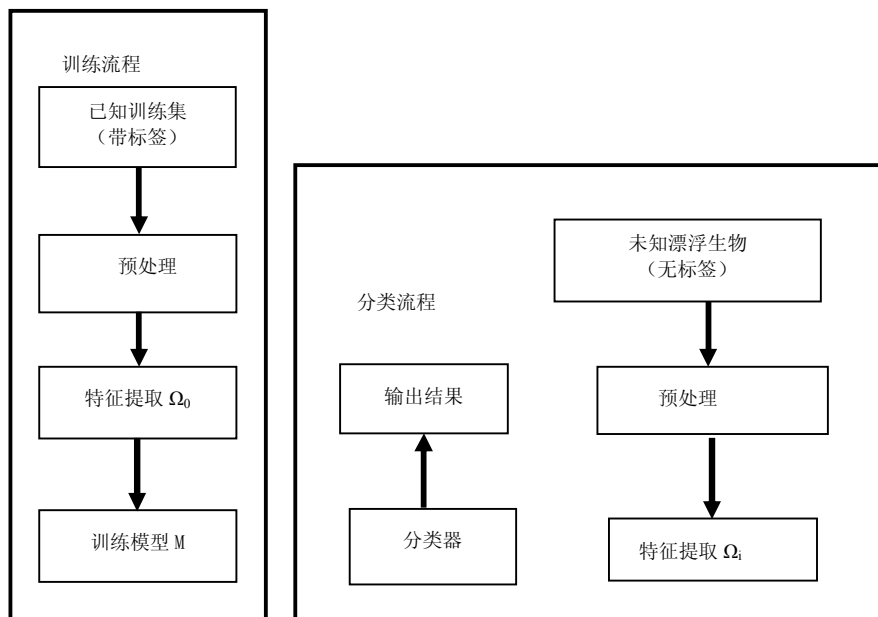


图 1 传统的浮游生物分类流程

Fig.1 Traditional procedure for plankton classification

3.1 浮游生物特征提取方法

3.1.1 基于 PCA 的特征提取

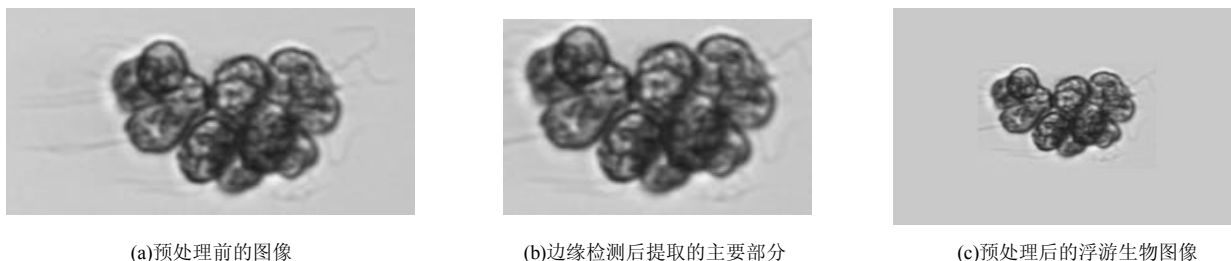
1) 预处理

由于 IFCB 所采集的数据尺寸各不相同, 首先应对训练集和测试集的图像数据进行标准化预处理。对输入的 8bit 灰度图像进行边缘检测, 提取

出主要部分, 并将尺寸统一为 300×200 大小。在进行尺寸变换时, 应保证图像未发生变形, 即处理后的图像中, 浮游生物的比例与边缘检测中提取的主要成分相同。将尺寸取成 300×200 的原因是, 基于浮游生物形态既有长条形也有圆球形这一事实, 为使浮游生物的占图像面积尽量大, 对长宽比

进行的折衷。之后进行零均值化和方差归一化,这样做是为了让不同的输入在尺度上一致,便于对输入数据进行比较。考虑到采集的浮游生物图像中,浮游生物的朝向各不相同,可以考虑寻找其主方向

并将其旋转到同一水平方向。这里我们采用一种简单的方法进行处理:将训练集中的图像分别进行 n 次旋转,每次旋转 $360^\circ/n$,这样就将图像数据进行了扩充。



(a)预处理前的图像

(b)边缘检测后提取的主要部分

(c)预处理后的浮游生物图像

图2 经过预处理的浮游生物图像

Fig.2 Plankton image after preprocessing

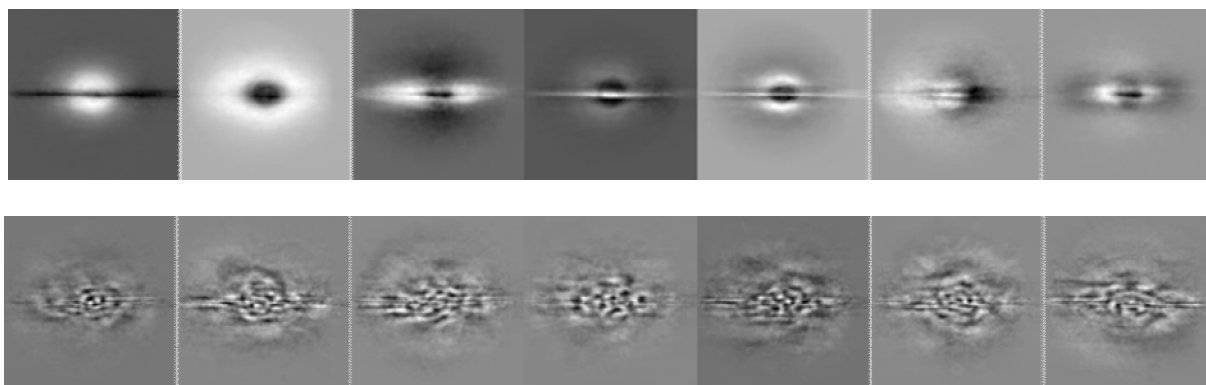
2) 训练及特征生成

经过以上的预处理后,我们从带标签的训练集中,每一类别的浮游生物选出部分图像进行训练,提取其特征。设 X 为 $D \times S$ 阶的矩阵,其每一列代表一个训练集中的图像向量,共有 S 个训练图像。首先进行中心化,将 X 各列减去所有列的均值,等效于将坐标原点取到平均点。应该求 X 的协方差阵 Σ , 并对其进行特征分解,即

$$\Sigma = \frac{1}{S}(X - \bar{X})(X - \bar{X})^T \quad (6)$$

在不引起混淆的情况下,这里将 X 默认为 $X - \bar{X}$ 。但此时的 Σ 为 $D \times D$ 维矩阵,而 D 高达

60,000, 对于一般计算机无法进行如此大规模的计算。因此,考虑到 $X^T X$ 与 XX^T 有相同的特征值,且两者特征向量之间仅仅差一个 X , 因此我们转而计算 $X^T X$ 的特征分解,进而得到 XX^T 的特征分解。我们可以得到相互正交的单位特征向量 $\varphi = [\varphi_1, \dots, \varphi_D]$ 和特征值 $\lambda = [\lambda_1, \dots, \lambda_D]$ 。由 PCA 的原理,我们可以仅保留前 M' 个大特征值对应的特征向量分量,并达到最小均方误差。即将特征空间的维数降低到 M' 维。特征向量剩下 $\varphi = [\varphi_1, \dots, \varphi_{M'}]$, 将 X 投影到 Φ 上可以得到 $\Omega = \Phi^T X$ 简而言之,训练过程输入带标签的训练集,输出了子空间的一组正交基 Φ , 以及训练集中元素在子空间的坐标 Ω 。



上排是前 7 个特征向量的图案,下排是后 7 个特征向量的图案

图3 特征向量的图案

Fig.3 Pattern of eigenvectors

在这里,特征向量的选取很有讲究。应该注意到,由于总共有 S 个输入向量, Σ 特征向量个数不

会超过 $S-1$, 因此 M' 不能超过 $S-1$ 。在人脸识别中,由于前几个特征向量主要反映光照等信息,往往可

以将其舍弃。而在本系统中,由于采集方式等原因,输入图像的亮度基本一致,此时将前面的特征舍弃可能不是一个好的处理。而对于特征值很小的特征向量往往是噪声分量,应该进行舍弃,报告后面部分会对特征向量的选择方式的结果进行评估。

3.1.2 基于 ICA 的特征提取

考虑到 PCA 仅保留了原信号的一二阶统计量,而自然界中信号的信息往往不仅体现在前两阶统计量上,在高阶统计量中也包含部分信息。除了对于高斯分布的信号,前两阶统计量能完全表现信号的性质,其他非高斯的信号经 PCA 降维后都会损失部分信息。对于浮游生物而言,其形态图像信号未必服从良好的高斯分布,因此可以利用 ICA 对其进行特征提取。



图4 处理后的图像

Fig.4 The processed image

3.2 浮游生物分类方法

无论是使用 PCA 还是 ICA 进行特征提取,在进行分类时都将待分类信号进行相同的预处理,再计算其在特征空间中的系数 Ω_1 ,然后送入分类器进行分类。对于分类器而言,最常使用的是 KNN 算法,还有卷积神经网络(CNN),SVM 等方法。以下主要讨论 KNN 算法在浮游生物分类的具体实现中的问题。

3.2.1 改进的 K-最近邻算法

K-最近邻算法最重要的就是距离的度量。关于距离的度量常用以下度量:L1 距离(City Block)、欧氏距离(Euclidean)、余弦距离(Cosine)、马氏距离(Mahalanobis) 计算出输入待分类浮游生物在特征空间中与测量集中每个元素的距离 d 。求出距离最近的 N 个元素的种类,计算众数 M ,即将该输入归为 M 类。

在 K-最近邻算法中,我们仅仅利用了训练集中同一类生物中的个体信息。注意到同类生物的形态的相似性,我们可以猜测其在特征空间中应聚集在某个中心周围。简单处理,我们将这个中心视为该

I 预处理

该部分应该与 PCA 基本相同。

II 训练及特征生成

由于时间原因,这里仅对 ICA 的实现方法一进行了尝试。首先对输入 X 的协方差阵进行特征分解(考虑 $X^T X$),得到 M' 个特征向量 $\varphi = [\varphi_1, \dots, \varphi_{M'}]$,利用 FastICA 作用在特征向量上, $U = W\Phi^T$,得到混合矩阵 W ,以及位于 U 的每一行的 M' 个统计独立的基图像。由 X 的估计值 $\tilde{X} = \Phi\Omega$,以及假设 W 可逆,我们可以得到

$$\tilde{X} = \Phi\Omega = U^T(\Phi W^{-1})^T X \quad (7)$$

所以, M' 个统计独立的基图像的线性组合系数为 $(\Phi W^{-1})^T X$,亦即将 PCA 系数乘上 $(W^{-1})^T$ 。

类别的均值。这样,我们可以在距离度量中引入该中心的影响,即在距离中附加上到中心的距离,以某种耦合机制合成一个新的距离。则待测向量 Y 到训练集中向量 X 的距离可以表示为

$$d_{\text{syn}}(X, Y) = \Phi(d(X, Y), d(X, Y)) \quad (8)$$

通过引入到类别中心的距离,我们有希望将识别率提高。这里 Φ 可简单选取为算术平均或几何平均。

3.2.2 多层混合分类(Multi-layer)

使用 KNN 算法在进行分类时,我们需要将输入图像分类到多达 100 种类型中的一类。可以想象,如果只需要将输入分类到 10 类,甚至 5 类中,应该比前者更容易。同时,结合海洋浮游生物结构相似的特点,即某几种浮游生物在形态上具有相似,我们可以将这个分类问题拆分成多个阶段。首先,根据各类别中心互相的距离,可以分出距离相近的类别对 (Type1, Type2),根据这些关系对可以构成几个等价类,如 $\{T_{A1}, \dots, T_{AK}\}$, $\{T_{B1}, \dots, T_{BK}\}$, ...。一层分类时,将等价类中类型看做同一类;第二层分类时,将等价类内的类型进一

步分为几个等价类，继续提取特征进行分类，以此类推。分类器的层数可以自己设计，而且将层次分

开可以对各层使用不同的特征提取方式，从而将各种特征提取方式结合了起来，如图 5 所示。

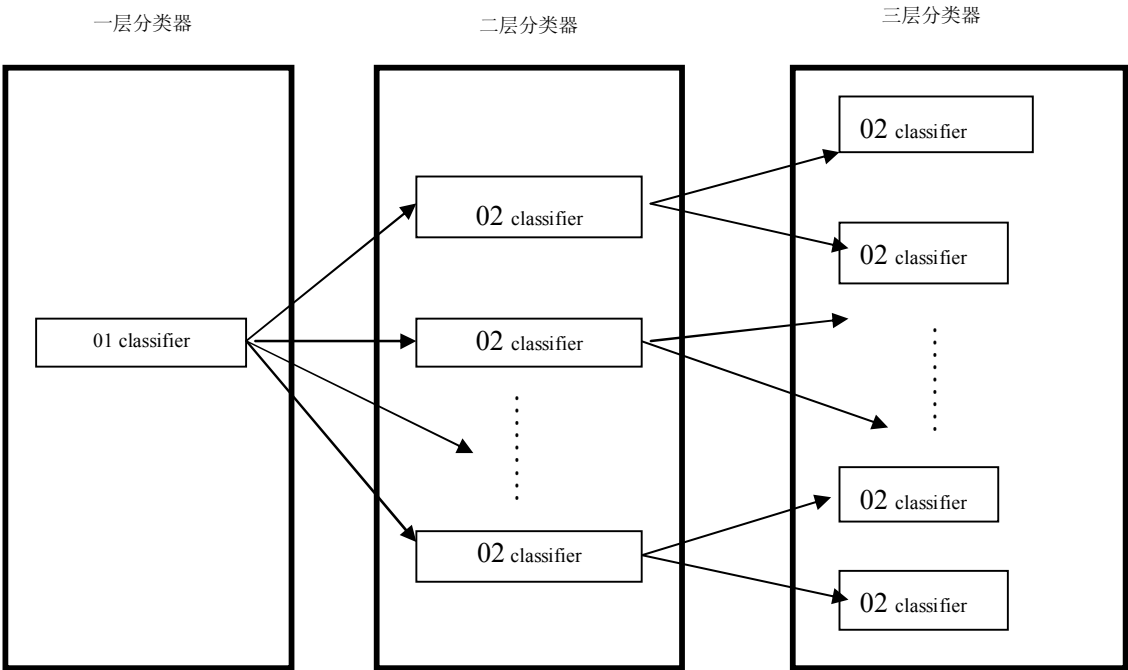


图 5 多层混合分类器的结构
Fig.5 The structure of multi-layer hybrid classifier

4 结果与分析

实验在 Windows 8.1 系统 MABLAB 2013b Intel i5-4200U 处理器，6GB 内存的计算机进行测试。测试内容有：①基于 PCA 的特征提取，②基于 ICA 的特征提取，③采用 KNN 分类，④多层混合分类的分类算法的组合进行了仿真。

测试数据使用了 WHOI 在 IFCB 所采集的浮游生物的图片数据。在 WHOI 数据集中已将浮游生物的图片分成了 103 类，移除其中分类信息不够明确的少数几类数据如（如 mix，bad，other_interaction）。为模拟真实的浮游生物分类需求，使用了 2013 年的采集数据作为训练，2014 年的数据作为待测集，这符合利用前期数据对未来数据进行预测的思路。测试集经过精简后剩余 89 类浮游生物，我们对这 89 个浮游生物进行了分类测试。从 2013 年的采集数据中的每一浮游生物类别中随机抽取 20 个样本构成训练集，从 2014 年

的采集数据中每类随机抽取 400 个样本（若不足则全部取出），共对 12270 个样本进行了分类测试。

考虑到时间原因以及测试条件的限制，仅对 2 层混合分类器和 K=5 的情况进行了仿真。实验结果显示，对于可以使用多种距离度量的基于 PCA 的浮游生物分类算法，在特征数较少的情况下，欧氏距离度量的表现最佳，而随着特征数增加，马氏距离获得的准确率更高。这是由于马氏距离将样本总体的特征值考虑进去，体现了不同特征分量的重要程度的差别。总体来说结果符合之前 3.1.1 部分对特征向量选取的分析，即特征值偏小的特征向量大体是一些噪声分量，将其舍弃能提高识别准确率。值得注意的是，若使用欧氏距离，在仅保留 25 个特征时即可达到最佳识别率，相比马氏距离的 120 个特征减少了很多。然而，其效果随特征数增加衰减得较快，相比马氏距离随特征数增加衰减得慢一些。若增大训练集的规模，可以想象，各种方法的识别率应该都会增加，而马氏距离是可以期待有更佳的表现。

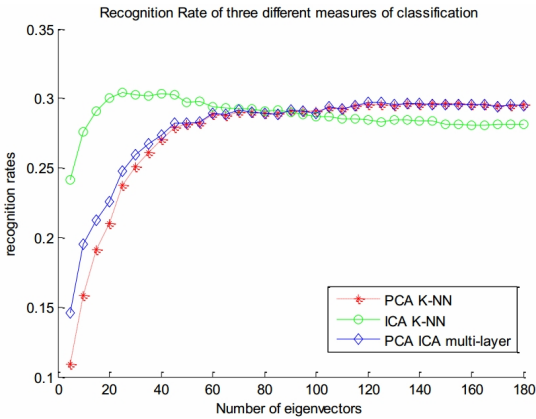


图6 多种距离度量对基于 PCA 特征提取的分类算法识别率的影响
Fig.6 Influence of different distance metrics on the recognition of PCA based algorithms

对于不同的分类算法,结果显示基于 ICA 和 KNN 的分类算法有最佳识别率,而混合分层的分类算法相比独立的 PCA 性能有提升。进一步,表 2 中结果显示经过优化使用合成距离度量的 KNN 识别算法比传统方法有更高的识别率,这说明引入到类别中心的距离可以充分有效的利用带标签训练集中浮游生物类别信息。

表 1 不同分类方法的最高准确率

Table 1 The highest recognition rates of different methods

Classification Metrics	Max Recognition Rate
PCA + Euclidean	0.3039
PCA + Mahalanobis	0.2966
PCA + Cosine	0.2959
PCA + L1	0.2997
ICA	0.3039
PCA ICA multi-layer	0.2971

表 2 使用合成距离度量对三种分类方式识别率的影响
Table 2 Influence of the application of synthesized distance metric to the three methods on their recognition rate

Classifier		Recognition Rate	
Type	#feature	Single Distance	Synthesized Distance
PCA(Euclidean)	25	0.2723	0.2888
ICA	95	0.2967	0.3039
PCA+ICA	120	0.2849	0.2971

另外,从图 8 中我们可以看到,针对不同的类别的识别率是不同的。表 3 中列举了三种具有特点的浮游生物,分别是识别率较高,识别率较低,

表 3 一些具有特点的浮游生物类型

Table 3 Some typical plankton

	High Performance	Low Performance	Easiest be Predicted
Plankton Name	Cylindrotheca	Thalassiosira	pennate
Image			
Recognition Rate	66%	0.75%	37.7%

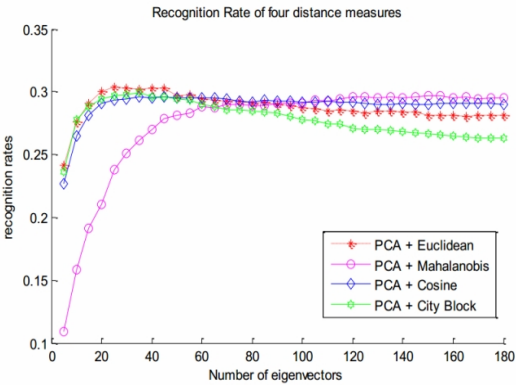


图7 三种分类算法的总体表现
Fig.7 overall performance of the three classification algorithms

以及最易被误判成的浮游生物类别。可以发现,识别效果差的类别内部元素形态较为不确定,而针对某些内部形态较为统一的浮游生物类型,分类算法有较好的表现。可以将这种区别归于该类型在特征空间中的分布并未很好的分离,而是混叠在一起。可以考虑使用 LDA 的方法,最小化类内距离,最大化类间距离,从而实现差异化。

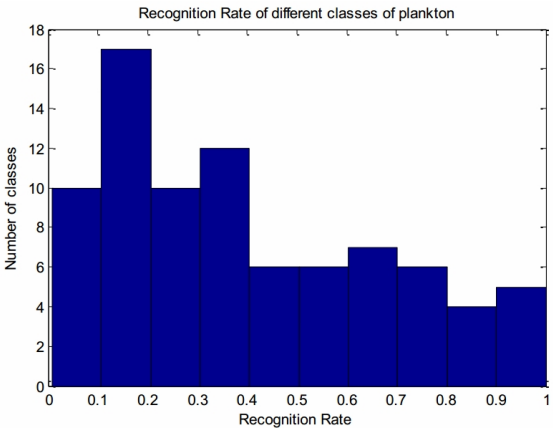


图8 不同类别浮游生物识别准确率统计,采用基于 PCA 的特征提取改进的和 KNN 分类法

Fig.8 Recognition rate of difference plankton classes, using PCA based features and the improved KNN classifier.

5 总结

本文考察了传统算法中的 PCA 方法和 ICA 方法结合 K-最近邻分类器在海洋浮游生物识别中的运用,详细分析了传统方法在该问题上应用时应该注意的问题,如距离度量的选择,特征向量的选择等。并通过仿真实验,对各传统方法的性能进行了评估。实验结果显示,基于 PCA 特征提取的分类算法中,欧氏距离在特征数较少的情形下表现最佳,最高可达到 30.4%的总体识别率;而马氏距离在特征数较大的情形效果最佳,最高可达到 29.6%的总体识别率。而基于 ICA 特征提取的分类算法仅仅取得与 PCA 类似的结果,原因可能是海洋浮游生物的图像分布为较好的高斯分布,从而使 ICA 退化为 PCA;也可能是因为 ICA 算法的选择有问题,没有正确的实现解混。另外,还有一种较常用的特征提取方法 LDA,由于时间的关系未能对其进行测试,我认为其最小化类内距离,最大化类间距离的特点在本问题上可以得到较好的结果。

基于同种浮游生物在特征空间中呈聚集的这一特点,将传统的 KNN 算法进行了改进,将特征空间中两点间距离与到类型的某一分布中心的距离相结合,形成一种耦合的距离度量。随后的仿真实验中的结果显示,采用这种度量的各分类算法的准确率相比传统方法都有所提升。同时,考虑到进行多元分类可以分解成多个规模较小的分类问题,由于直接使用 PCA 或 ICA 形成的特征空间中各类别元素的差异化程度不够,我们提出了一种多层的混合分类模型,通过多层分类器的结构,将复杂的分类分解成多层的简单分类,提供了将 PCA 和 ICA 以及不同分类器结合的途径。我们利用 WHOI 的浮游生物样本图像对构造的分类模型进行的测试结果显示,基于耦合度量的混合分类模型具有良好的性能,最高可达到 29.7%的总体识别率。

对于各个类别的浮游生物,分类算法的识别率不尽相同。对于测试样本数较大的类别(>400),最高能达到 66%以上,最低仅有 0.75%。同时,有些类别识别率为 37.7%,却被分类算法错判了 1200 次。这也反映了浮游生物各类别之间的复杂联系,我认为从各类别的错判率中能提取出类别之间的一些内在关系,今后有可能运用到对其的分类当中。

浮游生物不比人脸,其形态较为多变,并且有些不同类别之间的差别极小甚至有较大重叠,故识别起来较为困难。由于源数据的每类类别中的图像数差异较大,有的类别中仅有数张图像,同时由于时间及测试条件有限,仅取较少的一部分图像作为训练集,因此,训练集对浮游生物의 刻画的代表性不够强,导致最终识别效果不佳。若加大训练量将有可能提高识别率。

参考文献:

- [1] Turk M, Pentland A. Eigenfaces for recognition[J]. Journal of cognitive neuroscience, 1991, 3(1): 71-86.
- [2] Turk M, Pentland A P. Face recognition using eigenfaces[C]. Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91, IEEE Computer Society Conference on. IEEE, 1991: 586-591.
- [3] Yambor W, Draper B, Beveridge R. Analyzing pca-based face recognition algorithms: Eigenvector selection and distance measures[C]. 2nd Workshop on Empirical Evaluation in Computer Vision, Dublin, Ireland. 2000.
- [4] Comon P. Independent component analysis, a new concept?[J]. Signal Processing, 1994, 36(3):287-314.
- [5] Bartlett M S, Movellan J R, Sejnowski T J. Face recognition by independent component analysis[J]. IEEE Transactions on neural networks, 2002, 13(6): 1450-1464.
- [6] Belhumeur P N, Hespanha J P, Kriegman D J. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection[J]. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 1997, 19(7): 711-720.
- [7] Cortes C, Vapnik V. Support-Vestor Networks[J]. Machine Learning, 1995, 20(3):273-297.
- [8] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]// International Conference on Neural Information Processing Systems. Curran Associates Inc. 2012:1097-1105.
- [9] Orenstein E C, Beijbom O, Peacock E E, et al. WHOI-Plankton-A Large Scale Fine Grained Visual Recognition Benchmark Dataset for Plankton Classification[J]. arXiv preprint arXiv:1510.00745, 2015.