

浮游动物特征总结

王如晨 朱亚菲

2015 年 8 月

目录

1	13 类浮游动物的特征	4
2	PkID 中用到的特征	8
2.1	位置特征	8
2.2	尺寸特征	9
2.3	灰度值特征	9
2.4	形状特征	10
2.5	生物统计特征	11
2.6	还没有查找到的特征	11
2.7	其他特征	12
3	采用 PkID 中的特征进行实验	13
3.1	采用 PkID 中的 67 个特征进行实验	13
3.1.1	PkID 中 67 个特征 +SVM Linear	13
3.1.2	PkID 中 67 个特征 +RandomForest	14
3.2	将 67 个特征分类进行实验	15
3.2.1	形状特征 +SVM Linear	15
3.2.2	形状特征 +RandomForest	15
3.2.3	灰度特征 +SVM Linear	16
3.2.4	灰度特征 +RandomForest	16
3.2.5	生物统计特征 +SVM Linear	16
3.2.6	生物统计特征 +RandomForest	17
3.2.7	不包括位置特征 +SVM Linear	17
3.2.8	不包括位置特征 +RandomForest	17

3.2.9	不包括位置特征和尺度特征 +SVM Linear	18
3.2.10	不包括位置特征和尺度特征 +RandomForest	18
3.2.11	不包括位置特征、尺度特征和灰度特征 +SVM Linear	18
3.2.12	不包括位置特征、尺度特征和灰度特征 +RandomForest	20
3.2.13	不包括位置特征、尺度特征、灰度特征和形状特征 +SVM Linear	20
3.2.14	不包括位置特征、尺度特征、灰度特征和形状特征 +RandomForest	21
3.3	从 67 个特征中选取部分特征进行实验	21
3.3.1	选取 16 个特征 +SVM Linear	21
3.3.2	选取 16 个特征 +RandomForest	22
3.3.3	选取 18 个特征 +SVM Linear	22
3.3.4	选取 18 个特征 +RandomForest	22
3.3.5	选取 19 个特征 +SVM Linear	22
3.3.6	选取 19 个特征 +RandomForest	23
3.3.7	选取 21 个特征 +SVM Linear	24
3.3.8	选取 21 个特征 +RandomForest	24
3.3.9	选取 25 个特征 +SVM Linear	25
3.3.10	选取 25 个特征 +RandomForest	26
3.3.11	选取 27 个特征 +SVM Linear	26
3.3.12	选取 27 个特征 +RandomForest	27
3.3.13	选取 29 个特征 +RandomForest	28
3.4	实验分析	28
4	计算机视觉特征提取	31
4.1	几何参数	31
4.1.1	边界的周长	31
4.1.2	边界的曲率	31
4.1.3	面积	31
4.1.4	宽度和高度	31
4.1.5	矩形度	31
4.1.6	体态比	31
4.1.7	圆形性	31
4.1.8	偏心率	32

4.1.9	凸率	32
4.1.10	密集度	32
4.1.11	球状性	32
4.1.12	伸长度	33
4.1.13	叶状性	33
4.2	几种典型的特征描述方法	33
4.2.1	多边形近似	33
4.2.2	骨架	33
4.2.3	形状数 (边界)	33
4.2.4	统计矩 (边界)	33
4.2.5	傅里叶描述子	33
4.2.6	小波描述子	33
4.2.7	拓扑描述 (区域)	33
4.2.8	不变矩 (区域)	33
4.2.9	统计方法 (纹理)	33
4.2.10	结构性方法 (纹理)	33
4.2.11	频谱方法 (纹理)	33
4.2.12	曲率尺度空间	33
4.2.13	主分量描述	33
4.2.14	关系描述	33
5	选取特征进行实验	34
5.1	Matlab	34
5.1.1	实验一	34
5.1.2	实验二	35
5.1.3	实验三	35

1. 13 类浮游动物的特征

Appendicularia (尾海鞘纲) 属于脊索动物门，体型像蝌蚪，身体分为躯干和尾两部分。躯干为椭圆形；尾部扁平，比躯干要长。¹ 大小：小于 5mm。

观察采集的图像发现：

- 形状像蝌蚪，分为躯干和尾部。
- 躯干较大且灰度较深，并不是呈现规则的椭圆（还有部分突出的东西还不知道是什么）。
- 尾部大致呈现两种形状：一种细长弯曲；另一种较粗（粗细甚至于头部差不多），呈现柳叶状。尾部的灰度相比于躯干较浅，轮廓不太清晰。

Bubble(气泡) 非生物。

观察采集的图像发现：

- 圆形。
- 气泡四周灰度深，中间灰度很浅，呈亮白色。

Chaetognatha (毛颚动物门) 全体分为头、躯干和尾 3 部分。头部略膨大，躯干与头连接处稍微缢缩为颈部，尾部末端尖细。头部两侧各着生一系列 4 ~ 13 根镰刀状、几丁质的颚毛。² 大小：最大成年成年个体长 105mm，该类成年个体的长度一般都大于 5mm。

观察采集的图像发现：

- 身体修长，可以明显看出身体分为头、躯干和尾三部分。
- 头部小且圆滑，在头与躯干连接的地方略窄。
- 躯干较粗，轮廓清晰。
- 尾部慢慢变窄，末尾尖。

Cladocera Penilia (Penilia avirostris, 鸟喙尖头蚤) 属于节肢动物门，鳃足纲，枝角目，俗称水跳蚤。大小：大约为 1mm 左右。

观察采集的图像发现：

¹<https://zh.wikipedia.org/wiki/%e5%b0%be%e6%b5%b7%e9%9e%98%e7%ba%b2>

²<https://zh.wikipedia.org/wiki/%e6%af%9b%e9%a2%9a%e5%8a%a8%e7%89%a9%e9%97%a8>

- 身体短小，有两条长长的触角（但并不是每一幅图像中都可以看到。有时触角是向前的，可以看的很清楚；有时触角是向后的，和身体重合在了一起）。
- 该类浮游动物身体中轴线的地方灰度较深（感觉类似人体的脊柱），这个颜色较深的中轴线上还有一条纹理线连向边缘（就像人体脊柱上连着的骨骼）。由于运动，扫描得到的图像中浮游动物的中轴线并不是都在其身体中间。

Copepoda(桡脚类) 属于节肢动物门，颚足纲，桡足类属于其下的一个亚纲。体形像泪珠，有大的触角。分为前体部和后体部，前体部较为宽大，后体部较为短小。³前体部前体部由头和胸部组成，头部有两对触角，胸部有颚足、五对胸足。后体部无附肢，由 3—5 节组成。最末的腹节称尾节，末端具 1 对尾叉，尾叉的末端有 5 根不等长的刚毛，常呈羽状。⁴

观察采集的图像发现：

- 该类动物身体呈长椭圆形，尾部长在椭圆形一段（由于尾叉末端有几根不等长的刚毛，因此尾部呈一簇），触角长在椭圆的另一端（一共有两对触角，但最多只能看到一两个，有的图像甚至看不到）。
- 从该类动物正上方扫描得到的目标关于其自身的中轴对称。从该类动物侧面扫描得到的目标不对称，其身体一侧长着几对胸足。
- 该类中有八十几张图片中有多个目标，应该分在 Multiple 类中。

Decapoda(十足目) 属于节肢动物门，软甲纲。分为两类：Lucifer hansenii 和 Crab larvae。体躯延长呈虾形（腹部发达）或缩短扁圆呈蟹形（腹部化）。⁵

观察采集的图像发现：

- 该类浮游动物形状特征并不是很统一，大致分为两类：虾形和蟹形。
- 一些图像中可以看到目标有一条尾巴（像虾的尾巴）。
- 一些图像中可以看到目标有一对灰度较深的复眼。

Doliolida (海樽目) 属于脊索动物门，樽海鞘纲⁶。体型一般呈桶状，体壁最外是被囊层，其内层是外套膜。被囊层下有 8～9 条肌带环绕着体躯。

观察采集的图像发现：

³<https://zh.wikipedia.org/wiki/%e6%a9%88%e8%85%b3%e9%a1%9e>

⁴<http://baike.baidu.com/view/665478.htm>

⁵http://baike.baidu.com/link?url=LWmrgD_DVUcwOupg_ziOLTlJWj6quxa_juRrS3zUt91A-FjPM6VQwYfZ5fFZckzIyEGCaXypikXUGg2JsYMXUX-uFEkmkLqC5lfxvXvApK3WRBcWQkfbDhMlfTdgrWvh-728gSoUylWZG2UstFK

⁶<https://zh.wikipedia.org/wiki/%e6%a8%bd%e6%b5%b7%e9%9e%98%e7%ba%b2>

- 由于该类浮游动物比较透明，因此在图像中灰度较浅，并且其桶状轮廓也不完整了，但最明显的是能看到大概 7、8 条环状的肌肉带，有的图像中还能看到内部器官。

Egg 很多种类生物的卵。

观察采集的图像发现：

- 形状大致都呈圆形。
- 有的卵整体灰度都很深；有的卵中间有一块灰度较深的区域，四周灰度较浅（结构像细胞）。
- 由于是不同动物的卵，因此其灰度特征差异较大。

Fiber (纤维) 非生物。

- 弯曲的线状，有的纤维有分叉和交叉。
- 该类图像中噪声较多，纤维的边缘也不是很规则。

Gelatinous (明胶) 该类包括很多不同种类的胶状生物，它们体内都含有很高的水分。包括 Aglaura (属于刺胞动物门，这一个没有搜的中文名字，但也属于水母类)、Medusa (水母，属于刺胞动物门，水螅纲)、Siphonophora (管水母，属于刺胞动物门，水螅纲，管水母目)、Radiolaria (放射虫，属于原声动物门，辐足纲) 和 Salps (樽海鞘，属于脊索动物门 Chordata，樽海鞘纲 Thaliacea，纽鳃樽目 Salpida)。其中水母大部分都有三个主要部位：圆伞状或是钟状（寺院里面敲得那种钟）的身体，触器和口腕。

观察采集的图像发现：

- 由于该类呈胶状，因此该类物体灰度整体较浅，边缘也不是十分清晰。大部分是水母，有小部分的樽海鞘（与海樽目形态很相似），小部分的放射虫。
- 其中水母也包括很多类，形态大致呈现以下几种：
 - 一些水母身体呈现类似钟状（这里呈现钟状有长有短，有粗有细，还有的会发生一点弯曲），灰度较浅，内部有一块颜色较深的椭圆形区域。
 - 一些水母也呈钟状，但内部没有颜色较深的椭圆形区域，整个身体灰度均匀。
 - 还有的个头稍微偏小，形状有的类似圆形、像半个胶囊（应该由于拍摄原因，有的拍到顶部，有的拍到侧面），体内有颜色较深的一个大点和几个小点。（可能是灯塔水母）
 - 还有四张看不出形状的，不知道是什么。

- 放射虫：形状近似圆形（但由于整体灰度较浅，形状保存是完整），中间有一块灰度较深的区域，四周灰度较浅，可以看到淡淡的细纹从中心连接到边界。

Multiple (多个生物) 由于浮游动物的重叠，导致分割过程中多个浮游动物被分割到一张图像上。

Nonbio 非生物的集合。（不符合以上集中浮游动物的形态特征）

Pteropoda (翼足目) 属于软体动物门，腹足纲。

观察采集的图像发现：

- 该类浮游动物灰度较深，形状总体都呈现一头宽一头窄。
- 形状总体呈现三类：有的呈现象牙状，有点弯曲；有的较粗短，像一顶尖的小帽子；有的呈现细长的三角形状。

参考网站：<http://www.imas.utas.edu.au/zooplankton/home>。

2. PkID 中用到的特征

PkID 中用到的特征一共有 67 个：Area, Mean, StdDev, Mode, Min, Max, X, Y, XM, YM, Perim., BX, BY, Width, Height, Major, Minor, Angle, Circ., Feret, IntDen, Median, Skew, Kurt, %Area, XStart, YStart, Area_exc, Fractal, Skelarea, Slope, Histcum1, Histcum2, Histcum3, XMg5, YMg5, Compentropy, Compmean, Compslope, CompM1, CompM2, CompM3, Symetrieih, Symetriev, Tag, ESD, Elongation, Range, MeanPos, CentroidsD, CV, SR, PerimAreaexc, FeretAreaexc, PerimFeret, PerimMaj, Circexc, CDexc, [Nb1](#), [Nb2](#), [Nb3](#), [Symetrieihc](#), [Symetrieivc](#), [Convperim](#), [Convarea](#), [Fcons](#), [ThickR](#) (这几个特征没有找到具体的含义)

从训练集的 PID 文件文件中看到, Compentropy, Compmean, Compslope, CompM1, CompM2, CompM3 这 6 个特征在所有图像上的值都为 0, Tag 这个特征在所有图像上的值都为 1, 在训练分类器时是不起作用的, 同时这 7 个特征的具体含义也没有找到。

2.1 位置特征

BX 能够包围物体, 且平行于图像两条边的最小外界矩形的左上角顶点的 X 坐标

BY 能够包围物体, 且平行于图像两条边的最小外界矩形的左上角顶点的 Y 坐标

Height 能够包围物体, 且平行于图像两条边的最小外界矩形的高

Width 能够包围物体, 且平行于图像两条边的最小外界矩形的宽

XStart 图像最左上角像素点的 X 坐标

YStart 图像最左上角像素点的 Y 坐标

XM 物体灰度重心的 X 坐标

YM 物体灰度重心的 Y 坐标

XMg5 gamma 值为 51 时的物体灰度重心的 X 坐标 (gamma 值表示图像输出值与输入值关系的斜率)

YMg5 gamma 值为 51 时的物体灰度重心的 Y 坐标

X 物体重心点的 X 坐标

Y 物体重心点的 Y 坐标

Angle 浮游动物主轴与图片 x 轴形成的夹角，在图片切割后旋转图片测量相关参数使用

这类特征反映的是浮游动物在图像中的位置信息，浮游动物特征与位置信息无关，因此它们不适合作为特征直接用于分类（会降低分类的准确率），而是用来计算其他特征（尺寸特征、灰度特征和形状特征）。

2.2 尺寸特征

Area 物体的表面积，方形像素的个数

Perim 周长，物体最外层边缘的长度

Major 物体的最佳拟合椭圆的长轴

Minor 物体的最佳拟合椭圆的短轴

Feret Maximum feret diameter（最大费雷特征），沿物体边缘任意两个点的最长距离

Area_exc 去掉物体空洞后的表面积，空洞是指灰度值与背景相同的部分

%area 物体表面积中空洞所占的百分比，即背景所占的比例

这类特征表示了图像中目标的大小尺寸。它的根据是同类浮游动物的表面积、周长等尺寸特征应该是大致相同的。但是这些特征还存在着问题：1、同类浮游动物在不同时期（如幼年和成年）的个体大小尺寸是不同的。2、拍摄照片的方位不同（比如正面和侧面）得到的尺寸特征也是不同的。

2.3 灰度值特征

Min 物体内部所有像素点的最小灰度值 (0 = black)

Max 物体内部所有像素点的最大灰度值 (255 = white)

IntDen (Integrated density) 总密度，物体内部像素点的灰度值的总和 ($IntDen = Area * Mean$)

Slope 归一化的灰度累计直方图的斜率

Histcum1 灰度累计直方图的值为 25% 时所对应的灰度值

Histcum2 灰度累计直方图的值为 50% 时所对应的灰度值

Histcum3 灰度累计直方图的值为 75% 时所对应的灰度值

CentroidsD $\sqrt{(XM - X)^2 + (YM - Y)^2}$ 目标物体重心和灰度重心之间的距离。

根据是同类浮游动物的灰度特征（灰度的范围和整体灰度变换趋势）应该是相似的，但观察图像发现并不是所有同类浮游动物的灰度都是相似的，例如 Gelatinous 类中有的个体灰度跨度较小，整体灰度都较浅，而有的个体灰度跨度较大；同时由于拍摄时光线的原因，会造成同类浮游动物中个体灰度的深浅不一。

2.4 形状特征

Fractal 物体边界的分形维数 (Berube and Jebrak, 1999)，表明物体边界的不规则程度

Skelarea 骨架像素的表面积（在二值图像中，不断地从物体边缘处减去像素点直到仅剩一个像素的宽度，最后所得图形的像素点数）

Symetrieih 水平对称

Symetriev 垂直对称

Circ $Circularity = (4 * Pi * Area) / Perim^2$ 圆形度，表征物体接近圆的程度，值等于 1 时，说明物体为正圆形，值越接近 0，物体体形越长。

ESD $2 \times \sqrt{\frac{Area}{\pi}}$ 相应球形直径（也称为等效球直径），是指一不规则外形物体，其体积相同球体的直径。

Elongation $\frac{Major}{Minor}$ 延伸率，最佳拟合椭圆的长轴和短轴之比。

Circexc $\frac{4 \times \pi Area_exc}{Perim^2}$ 去掉目标内部空洞的圆形度。

这类特征描述的是浮游动物的灰度特征，根据的是不同种类浮游动物的形状不同。存在的问题是有不同种类的浮游动物形状相似，例如 Appendicularia 和 Chaetognatha, Bubble 和 Egg；也有同种浮游动物形状不同，例如 Decapoda、Gelatinous。

2.5 生物统计特征

Mean 物体内的平均灰度值；物体中所有像素点的灰度值的总和除以总的像素个数

Range $Max - Min$ 极差，灰度的范围。

CV $100 \times \frac{StdDev}{Mean}$ 变异系数（也称离散系数或相对偏差），是灰度标准偏差与平均值之比，用百分数表示。

SR $100 \times \frac{StdDev}{Max - Min}$ 灰度标准差比上极差。

Skew 灰度直方图的偏度，衡量灰度分布的不对称性。偏度为负就意味着在概率密度函数左侧的尾部比右侧的长，绝大多数的值位于平均值的右侧。偏度为正就意味着在概率密度函数右侧的尾部比左侧的长，绝大多数的值位于平均值的左侧。偏度为零就表示数值相对均匀地分布在平均值的两侧，但不一定意味着其为对称分布。

Kurt 峰度，描述灰度直方图的陡缓程度。

Mean_exc 物体内部去掉空洞后的平均灰度值 ($Mean_exc = IntDen / Area_exc$)

Median 物体内部像素的灰度值的中值

StdDev 物体内部像素的灰度值的标准差

Mode Modal grey value within the object（可能表示灰度的众数）

2.6 还没有查找到的特征

MeanPos $\frac{Mean - Max}{Max - Min}$

PerimAreaexc $\frac{Perim}{\sqrt{Area_exc}}$

FeretAreaexc $\frac{Feret}{\sqrt{Area_exc}}$

PerimFeret $\frac{Perim}{Feret}$

$$\text{PerimMaj} \frac{Perim}{Major}$$

$$\text{CDexc} \frac{\sqrt{(XM - X)^2 + (YM - Y)^2}}{\sqrt{Area_exc}}$$

Nb1

Nb2

Nb3

Symetriehc

Symetrievc

Convperim

Convarea

Fcons

ThickR

2.7 其他特征

这些特征并没有在 PkID 中使用，而是在作者的一个幻灯片中提到的新特征。

Neighborhood gray-tone difference Matrix (NGTDM) 邻域灰度差别矩阵

Cumulation Histogram 累积直方图

Convex Area

Symmetry

Thickness Ratio

3. 采用 PkID 中的特征进行实验

3.1 采用 PkID 中的 67 个特征进行实验

3.1.1 PkID 中 67 个特征 +SVM Linear

在 PkID 软件中，选用 67 个特征和 SVM Linear 分类器进行训练和分类得到的结果如图 1，其分类准确率为 70.6%。

Error rate			0.2593													
Values prediction			Confusion matrix													
Value	Recall	1-Precision	Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum
Appendicularia	0.8400	0.1960	2289	0	92	0	6	11	0	1	45	6	123	151	1	2725
Bubble	0.7959	0.2640	0	577	0	0	20	0	0	52	0	0	3	73	0	725
Chaetognatha	0.8883	0.0941	133	0	1550	0	0	10	0	0	14	1	20	17	0	1745
CladoceraPenilia	0.7894	0.1993	0	0	0	3651	18	0	4	0	0	29	5	918	0	4625
Copepoda	0.7984	0.2842	0	0	0	17	5884	187	0	0	2	1	180	1011	88	7370
Decapoda	0.6747	0.2538	2	0	0	0	470	1879	0	0	0	0	88	339	7	2785
Doliolida	0.7502	0.1790	0	0	0	8	0	0	1069	0	0	131	2	215	0	1425
Egg	0.6640	0.1766	0	129	0	25	27	2	0	1142	0	65	0	320	10	1720
Fiber	0.7342	0.2146	44	0	28	0	0	0	0	9	1116	0	39	284	0	1520
Gelatinous	0.6844	0.2238	6	0	1	119	21	10	142	36	0	2060	73	540	2	3010
Multiple	0.3002	0.4646	264	0	21	101	589	174	15	8	50	87	953	897	16	3175
Nonbio	0.8041	0.2890	105	78	16	639	946	185	72	123	194	274	288	12375	95	15390
Pteropoda	0.4525	0.3085	4	0	3	0	239	60	0	16	0	0	6	266	491	1085
Sum			2847	784	1711	4560	8220	2518	1302	1387	1421	2654	1780	17406	710	47300

图 1: PkID 软件中 SVM Linear 分类器分类结果

PkID 系统中 SVM Linear 的参数如图 2。根据图中的参数，自己设计的 SVM Linear 分类器得到的混淆矩阵如图 3。在采用 SVM 进行训练之前，要先对特征进行归一化。

Parameters	
Kernel type	LINEAR
Degree (poly)	1.00
Gamma in kernel function (poly/rbf/sigmoid)	0
Coef0 in kernel function(poly/sigmoid)	0
Tolerance of termination criteria (eps)	0.0001
C (Complexity Cost)	1
Compute probability estimates	0
Use shrinking heuristics	1
Data normalization	1

图 2: PkID 系统中 SVM Linear 的参数

	Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Total	Recall	1-Precision
Appendicularia	2325	0	95	3	9	15	0	0	51	5	106	116	0	2725	0.853211	0.202948
Bubble	0	591	0	1	17	0	0	60	0	0	9	47	0	725	0.815172	0.316763
Chaetognatha	144	0	1556	0	0	6	0	0	9	0	14	16	0	1745	0.891691	0.099016
CladoceraPenilia	0	0	0	3797	19	0	8	1	0	30	5	765	0	4625	0.820973	0.196402
Copepoda	0	0	0	14	6119	197	0	0	2	3	167	802	66	7370	0.830258	0.296586
Decapoda	1	0	0	0	565	1894	0	0	0	0	68	247	10	2785	0.680072	0.261022
Doliolida	0	0	0	16	0	0	1148	0	0	94	0	167	0	1425	0.805614	0.157122
Egg	0	172	0	28	90	3	1	1117	0	61	0	240	8	1720	0.649419	0.176254
Fiber	54	0	31	0	0	0	0	9	1108	0	43	275	0	1520	0.728947	0.216407
Gelatinous	8	0	2	131	22	13	127	44	2	2103	60	498	0	3010	0.698671	0.213538
Multiple	260	1	23	114	599	189	15	12	56	107	948	840	11	3175	0.298583	0.443662
Nonbio	120	101	17	621	1003	184	63	99	186	270	279	12346	101	15390	0.802209	0.256265
Pteropoda	5	0	3	0	256	62	0	14	0	1	5	241	498	1085	0.458986	0.282421
Total	2917	865	1727	4725	8699	2563	1362	1356	1414	2674	1704	16600	694	47300	0.717985	0.239877

图 3: PkID-SVM 交叉验证, folds 取 2, repetitions 取 5

3.1.2 PkID 中 67 个特征 + RandomForest

在 PkID 软件中, 选用 67 个特征和随机森林分类器进行训练和分类得到的结果如图 4, 其分类准确率为 75.4%。

Error rate			0.2155														
Values prediction			Confusion matrix														
Value	Recall	1- Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum
Appendicularia	0.8051	0.2039	Appendicularia	2194	0	111	8	1	19	0	0	57	1	157	177	0	2725
Bubble	0.8359	0.1049	Bubble	0	606	0	0	0	0	0	29	0	0	3	87	0	725
Chaetognatha	0.8934	0.0983	Chaetognatha	137	0	1559	0	5	2	0	0	7	1	18	14	2	1745
CladoceraPenilia	0.8830	0.1044	CladoceraPenilia	0	0	0	4084	22	0	7	0	0	7	19	486	0	4625
Copepoda	0.8419	0.2108	Copepoda	3	0	0	9	6205	145	0	0	0	1	234	745	28	7370
Decapoda	0.7989	0.2157	Decapoda	4	0	0	0	339	2225	0	0	0	0	87	126	4	2785
Doliolida	0.7474	0.1418	Doliolida	0	0	0	2	1	0	1065	0	0	156	7	194	0	1425
Egg	0.7424	0.1401	Egg	0	66	0	13	2	1	1	1277	0	28	3	320	9	1720
Fiber	0.7414	0.2069	Fiber	59	0	18	0	0	0	0	0	1127	0	62	254	0	1520
Gelatinous	0.6914	0.2042	Gelatinous	6	0	0	72	10	5	93	39	0	2081	79	624	1	3010
Multiple	0.3238	0.5093	Multiple	267	0	23	51	491	202	9	23	45	91	1028	941	4	3175
Nonbio	0.8406	0.2452	Nonbio	86	5	17	321	716	189	66	117	185	246	390	12937	115	15390
Pteropoda	0.6627	0.1848	Pteropoda	0	0	1	0	70	49	0	0	0	3	8	235	719	1085
			Sum	2756	677	1729	4560	7862	2837	1241	1485	1421	2615	2095	17140	882	47300

图 4: PkID 软件中随机森林分类器分类结果

自己设计随机森林分类器, 对 67 个特征进行训练和分类得到的混淆矩阵如图 5。

	Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Total	Recall	1-Precision
Appendicularia	2220	0	110	13	1	21	0	0	61	2	137	160	0	2725	0.8147	0.195069
Bubble	0	629	0	0	3	0	0	28	0	0	4	61	0	725	0.8676	0.096264
Chaetognatha	147	0	1550	0	2	6	0	0	4	1	23	12	0	1745	0.8883	0.093567
CladoceraPenilia	0	0	0	4156	18	0	7	0	0	1	11	432	0	4625	0.8986	0.079717
Copepoda	4	0	0	7	6309	161	0	0	0	0	193	679	17	7370	0.856	0.190323
Decapoda	1	0	0	0	316	2270	0	0	0	0	76	121	1	2785	0.8151	0.221269
Doliolida	0	0	0	1	0	0	1104	0	0	160	7	153	0	1425	0.7747	0.110395
Egg	0	57	0	15	3	1	0	1361	0	30	1	246	6	1720	0.7913	0.098675
Fiber	46	0	19	0	0	0	5	0	1129	0	67	254	0	1520	0.7428	0.204369
Gelatinous	6	0	0	49	12	6	67	19	0	2127	88	635	1	3010	0.7066	0.196145
Multiple	266	2	16	38	452	219	13	17	43	93	1054	955	7	3175	0.332	0.481299
Nonbio	68	8	13	237	626	187	45	85	182	227	366	13240	106	15390	0.8603	0.230635
Pteropoda	0	0	2	0	50	44	0	0	0	5	5	261	718	1085	0.6618	0.161215
Total	2758	696	1710	4516	7792	2915	1241	1510	1419	2646	2032	17209	856	47300	0.77	0.181457

图 5: PkID-RF 交叉验证, folds 取 2, repetitions 取 5

3.2 将 67 个特征分类进行实验

3.2.1 形状特征 +SVM Linear

在 PkID 软件中，从 67 个特征中选取形状特征（8 个），并采用 SVM Linear 进行训练和分类得到的结果如图 6，其分类准确率为 37%。

Error rate			0.5353														
Values prediction			Confusion matrix														
Value	Recall	1-Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum
Appendicularia	0.6095	0.4093	Appendicularia	1661	0	102	0	9	54	0	0	153	5	41	700	0	2725
Bubble	0.0000	1.0000	Bubble	0	0	0	0	1	0	0	415	0	0	0	309	0	725
Chaetognatha	0.8080	0.1334	Chaetognatha	227	0	1410	0	0	18	0	0	61	2	19	8	0	1745
CladoceraPenilia	0.0465	0.7018	CladoceraPenilia	0	0	0	215	52	0	0	1	0	39	0	4318	0	4625
Copepoda	0.0837	0.6425	Copepoda	14	0	0	17	617	63	0	0	0	16	70	6572	1	7370
Decapoda	0.2743	0.4264	Decapoda	71	0	1	8	147	764	0	0	0	25	121	1648	0	2785
Doliolida	0.3354	0.5047	Doliolida	0	0	0	1	3	0	478	0	0	495	1	447	0	1425
Egg	0.5860	0.4886	Egg	0	0	0	18	0	0	1	1008	0	72	0	621	0	1720
Fiber	0.6039	0.3035	Fiber	167	0	53	0	0	0	0	6	918	0	28	348	0	1520
Gelatinous	0.4924	0.3789	Gelatinous	7	0	0	23	73	49	201	303	0	1482	21	840	11	3010
Multiple	0.0967	0.5518	Multiple	295	0	25	41	126	191	46	17	34	70	307	2023	0	3175
Nonbio	0.8500	0.5874	Nonbio	309	0	31	397	619	120	238	204	152	166	73	13081	0	15390
Pteropoda	0.0359	0.2353	Pteropoda	61	0	5	1	79	73	1	17	0	14	4	791	39	1085
			Sum	2812	0	1627	721	1726	1332	965	1971	1318	2386	685	31706	51	47300

图 6: 采用形状特征 SVM Linear 进行分类的结果

3.2.2 形状特征 +RandomForest

在 PkID 软件中，从 67 个特征中选取形状特征（8 个），并采用随机森林进行训练和分类得到的结果如图 7，其分类准确率为 63.1%。

Error rate			0.3285														
Values prediction			Confusion matrix														
Value	Recall	1-Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum
Appendicularia	0.7006	0.3417	Appendicularia	1909	0	129	3	8	54	5	0	110	1	188	289	29	2725
Bubble	0.5959	0.3143	Bubble	0	432	0	0	3	0	0	149	0	0	0	141	0	725
Chaetognatha	0.8521	0.1053	Chaetognatha	194	0	1487	0	0	8	0	0	22	0	25	8	1	1745
CladoceraPenilia	0.7529	0.3051	CladoceraPenilia	0	0	0	3482	141	27	2	4	0	15	45	900	9	4625
Copepoda	0.7331	0.3082	Copepoda	12	0	0	197	5403	205	8	0	1	35	280	1214	15	7370
Decapoda	0.6668	0.3284	Decapoda	57	0	5	71	284	1857	0	0	0	15	174	311	11	2785
Doliolida	0.5979	0.3804	Doliolida	0	0	0	17	41	0	852	2	0	306	7	198	2	1425
Egg	0.5593	0.3109	Egg	0	175	0	81	0	0	4	962	2	71	0	421	4	1720
Fiber	0.7158	0.3008	Fiber	88	0	15	0	4	0	0	5	1088	1	59	258	2	1520
Gelatinous	0.5638	0.3279	Gelatinous	7	3	0	84	121	13	260	127	0	1697	46	593	59	3010
Multiple	0.2243	0.6229	Multiple	312	0	12	220	436	242	54	11	86	93	712	986	11	3175
Nonbio	0.7364	0.3324	Nonbio	258	20	6	845	1336	307	188	136	247	256	344	11333	114	15390
Pteropoda	0.5060	0.3189	Pteropoda	63	0	8	11	33	52	2	0	0	35	8	324	549	1085
			Sum	2900	630	1662	5011	7810	2765	1375	1396	1556	2525	1888	16976	806	47300

图 7: 采用形状特征随机森林进行分类的结果

3.2.3 灰度特征 +SVM Linear

在 PkID 软件中, 从 67 个特征中选取灰度特征 (18 个, 包括生物统计特征), 并采用 SVM Linear 进行训练和分类得到的结果如图 8, 其分类准确率为 37.6%。

Error rate			0.4549													
Values prediction			Confusion matrix													
Value	Recall	1-Precision	Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum
Appendicularia	0.6914	0.3144	1884	0	178	102	36	15	0	0	0	49	100	361	0	2725
Bubble	0.0566	0.6168	0	41	0	5	94	0	0	46	0	0	1	538	0	725
Chaetognatha	0.5433	0.2582	329	0	948	5	31	5	5	0	0	48	76	298	0	1745
CladoceraPenilia	0.4415	0.4855	0	0	0	2042	12	0	7	0	0	2	1	2561	0	4625
Copepoda	0.6836	0.4252	0	6	0	28	5038	301	0	30	0	0	81	1860	26	7370
Decapoda	0.5451	0.3286	0	0	0	3	737	1518	0	13	0	0	66	433	15	2785
Doliolida	0.3705	0.4477	1	0	0	10	0	0	528	0	0	63	6	817	0	1425
Egg	0.1994	0.6163	0	3	0	67	417	35	4	343	0	14	9	776	52	1720
Fiber	0.0020	0.0000	49	0	14	40	26	0	19	0	3	6	22	1341	0	1520
Gelatinous	0.2066	0.4842	78	1	52	476	26	18	217	3	0	622	144	1373	0	3010
Multiple	0.1644	0.5639	276	2	61	331	683	162	35	12	0	153	522	930	8	3175
Nonbio	0.7843	0.4873	126	49	25	860	1309	164	141	185	0	249	163	12071	48	15390
Pteropoda	0.2037	0.4027	5	5	0	0	356	43	0	262	0	0	6	187	221	1085
Sum			2748	107	1278	3969	8765	2261	956	894	3	1206	1197	23546	370	47300

图 8: 采用灰度特征 SVM Linear 进行分类的结果

3.2.4 灰度特征 +RandomForest

在 PkID 软件中, 从 67 个特征中选取灰度特征 (18 个, 包括生物统计特征), 并采用随机森林进行训练和分类得到的结果如图 9, 其分类准确率为 61.3%。

Error rate			0.3204													
Values prediction			Confusion matrix													
Value	Recall	1-Precision	Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum
Appendicularia	0.7160	0.2929	1951	0	235	54	24	25	4	0	36	41	150	205	0	2725
Bubble	0.8166	0.1834	0	592	0	0	3	0	0	8	0	0	5	117	0	725
Chaetognatha	0.7278	0.2853	229	0	1270	0	11	3	5	0	5	46	73	99	4	1745
CladoceraPenilia	0.8666	0.1918	8	0	0	4008	30	0	19	2	0	49	39	470	0	4625
Copepoda	0.7659	0.3438	10	0	7	13	5645	408	0	9	0	11	196	987	84	7370
Decapoda	0.7260	0.3090	6	0	0	1	527	2022	0	5	0	2	86	114	22	2785
Doliolida	0.7200	0.2914	1	0	0	23	0	0	1026	0	0	187	24	164	0	1425
Egg	0.4645	0.2455	0	41	0	12	188	45	0	799	2	23	49	491	70	1720
Fiber	0.2605	0.3333	75	0	9	37	28	0	12	0	396	6	34	923	0	1520
Gelatinous	0.4837	0.4245	56	0	78	191	16	11	200	7	2	1456	203	789	1	3010
Multiple	0.2466	0.6139	297	6	100	90	627	172	38	34	8	220	783	777	23	3175
Nonbio	0.7637	0.3136	126	86	78	526	1194	187	144	163	145	486	378	11753	124	15390
Pteropoda	0.4074	0.4260	0	0	0	4	310	53	0	32	0	3	8	233	442	1085
Sum			2759	725	1777	4959	8603	2926	1448	1059	594	2530	2028	17122	770	47300

图 9: 采用灰度特征随机森林进行分类的结果

3.2.5 生物统计特征 +SVM Linear

在 PkID 软件中, 从 67 个特征中选取生物统计特征 (16 个, 包括没有查找到的特征), 并采用 SVM Linear 进行训练和分类得到的结果如图 10, 其分类准确率为 52.4%。

Error rate			0.3764														
Values prediction			Confusion matrix														
Value	Recall	1-Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum
Appendicularia	0.7831	0.2149	Appendicularia	2134	0	118	25	11	4	0	0	31	16	191	195	0	2725
Bubble	0.2345	0.3950	Bubble	0	170	0	4	31	0	0	76	0	0	10	434	0	725
Chaetognatha	0.8350	0.1275	Chaetognatha	171	0	1457	4	8	6	0	0	32	0	31	36	0	1745
CladoceraPenilia	0.5952	0.3541	CladoceraPenilia	0	0	0	2753	9	0	7	0	0	40	8	1808	0	4625
Copepoda	0.6782	0.3999	Copepoda	0	0	0	14	4998	337	0	0	6	0	191	1715	109	7370
Decapoda	0.5110	0.3544	Decapoda	3	0	0	0	911	1423	0	3	0	0	75	351	19	2785
Doliolida	0.3586	0.3122	Doliolida	0	0	0	9	0	0	511	0	0	218	12	675	0	1425
Egg	0.3855	0.3094	Egg	0	42	0	70	53	7	4	663	0	47	3	810	21	1720
Fiber	0.6217	0.2248	Fiber	38	0	61	3	17	0	0	0	945	0	35	421	0	1520
Gelatinous	0.4286	0.3505	Gelatinous	10	2	0	384	31	8	95	9	0	1290	81	1100	0	3010
Multiple	0.2539	0.5322	Multiple	270	0	19	208	627	161	35	15	50	62	806	914	8	3175
Nonbio	0.7778	0.4178	Nonbio	89	67	15	788	1189	192	91	135	155	313	275	11971	110	15390
Pteropoda	0.3465	0.4152	Pteropoda	3	0	0	0	444	66	0	59	0	0	5	132	376	1085
			Sum	2718	281	1670	4262	8329	2204	743	960	1219	1986	1723	20562	643	47300

图 10: 采用生物统计特征 SVM Linear 进行分类的结果

3.2.6 生物统计特征 +RandomForest

在 PkID 软件中, 从 67 个特征中选取生物统计特征 (16 个, 包括没有查找到的特征), 并采用随机森林进行训练和分类得到的结果如图 11, 其分类准确率为 69.9%。

Error rate			0.2699														
Values prediction			Confusion matrix														
Value	Recall	1- Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum
Appendicularia	0.7828	0.2412	Appendicularia	2133	0	137	0	6	22	3	0	55	15	204	150	0	2725
Bubble	0.7986	0.1201	Bubble	0	579	0	0	15	1	0	29	0	6	14	81	0	725
Chaetognatha	0.8223	0.1345	Chaetognatha	210	0	1435	0	11	8	0	0	35	0	16	29	1	1745
CladoceraPenilia	0.8322	0.1980	CladoceraPenilia	0	0	0	3849	21	0	53	0	0	18	54	630	0	4625
Copepoda	0.8262	0.2703	Copepoda	2	0	0	12	6089	243	0	0	0	0	214	764	46	7370
Decapoda	0.7056	0.2660	Decapoda	2	0	1	0	513	1965	0	0	0	0	101	192	11	2785
Doliolida	0.7404	0.2596	Doliolida	0	0	0	19	0	0	1055	0	0	94	12	245	0	1425
Egg	0.7157	0.1475	Egg	0	57	0	5	8	6	5	1231	0	82	19	303	4	1720
Fiber	0.6553	0.2735	Fiber	58	0	42	1	3	0	0	0	996	1	87	326	6	1520
Gelatinous	0.5721	0.2963	Gelatinous	13	3	0	127	9	1	94	32	0	1722	109	898	2	3010
Multiple	0.2825	0.5849	Multiple	270	2	24	139	571	200	30	35	70	94	897	830	13	3175
Nonbio	0.7767	0.2817	Nonbio	123	17	19	647	917	208	185	117	215	411	427	11954	150	15390
Pteropoda	0.5797	0.2703	Pteropoda	0	0	0	0	181	23	0	0	0	4	7	241	629	1085
			Sum	2811	658	1658	4799	8344	2677	1425	1444	1371	2447	2161	16643	862	47300

图 11: 采用生物统计特征随机森林进行分类的结果

3.2.7 不包括位置特征 +SVM Linear

在 PkID 软件中, 从 67 个特征中去掉 13 个位置特征, 并采用 SVM Linear 进行训练和分类得到的结果如图 12, 其分类准确率为 70.8%。

3.2.8 不包括位置特征 +RandomForest

在 PkID 软件中, 从 67 个特征中去掉 13 个位置特征, 并采用随机森林进行训练和分类得到的结果如图 13, 其分类准确率为 76.1%。

Error rate			0.2592														
Values prediction			Confusion matrix														
Value	Recall	1-Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum
Appendicularia	0.8426	0.1901	Appendicularia	2296	0	84	1	10	11	0	1	39	4	122	157	0	2725
Bubble	0.8400	0.2784	Bubble	0	609	0	1	19	0	0	23	0	0	2	71	0	725
Chaetognatha	0.8968	0.0826	Chaetognatha	124	0	1565	0	0	8	0	0	15	0	18	15	0	1745
CladoceraPenilia	0.7948	0.2026	CladoceraPenilia	0	0	0	3676	19	0	5	0	0	31	6	888	0	4625
Copepoda	0.7986	0.2908	Copepoda	0	0	0	16	5886	185	0	0	1	0	190	1011	81	7370
Decapoda	0.6661	0.2604	Decapoda	1	0	0	0	515	1855	0	0	0	0	92	313	9	2785
Doliolida	0.7481	0.1704	Doliolida	0	0	0	7	0	0	1066	0	0	132	3	217	0	1425
Egg	0.6448	0.1515	Egg	0	154	0	23	29	1	0	1109	0	65	0	328	11	1720
Fiber	0.7342	0.2196	Fiber	54	0	21	0	0	0	0	8	1116	0	42	279	0	1520
Gelatinous	0.6973	0.2220	Gelatinous	5	0	1	114	19	10	132	31	3	2099	62	531	3	3010
Multiple	0.3002	0.4591	Multiple	241	0	19	116	592	178	12	10	58	84	953	901	11	3175
Nonbio	0.8016	0.2880	Nonbio	109	81	15	656	965	197	70	115	198	283	266	12336	99	15390
Pteropoda	0.4387	0.3101	Pteropoda	5	0	1	0	246	63	0	10	0	0	6	278	476	1085
			Sum	2835	844	1706	4610	8300	2508	1285	1307	1430	2698	1762	17325	690	47300

图 12: 不采用位置特征 SVM Linear 进行分类的结果

Error rate			0.2092														
Values prediction			Confusion matrix														
Value	Recall	1- Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum
Appendicularia	0.8048	0.2109	Appendicularia	2193	0	117	8	1	19	0	0	47	4	177	159	0	2725
Bubble	0.8262	0.1152	Bubble	0	599	0	0	1	0	0	44	0	0	0	81	0	725
Chaetognatha	0.8871	0.0979	Chaetognatha	147	0	1548	0	2	4	0	0	4	1	27	12	0	1745
CladoceraPenilia	0.8962	0.0944	CladoceraPenilia	0	0	0	4145	19	0	4	0	0	6	15	436	0	4625
Copepoda	0.8468	0.2051	Copepoda	4	0	0	5	6241	138	0	0	2	5	220	723	32	7370
Decapoda	0.7953	0.2173	Decapoda	7	0	0	0	321	2215	0	0	0	0	101	135	6	2785
Doliolida	0.7740	0.1239	Doliolida	0	0	0	2	0	0	1103	0	0	141	9	170	0	1425
Egg	0.7669	0.1052	Egg	0	68	0	12	1	2	0	1319	0	26	2	276	14	1720
Fiber	0.7586	0.2172	Fiber	37	0	12	0	0	0	0	0	1153	0	65	251	2	1520
Gelatinous	0.7213	0.2021	Gelatinous	7	0	2	58	9	4	69	27	0	2171	91	571	1	3010
Multiple	0.3203	0.5176	Multiple	291	0	19	52	486	213	5	14	57	95	1017	920	6	3175
Nonbio	0.8441	0.2344	Nonbio	93	10	12	295	711	185	78	70	209	267	375	12991	94	15390
Pteropoda	0.6562	0.1788	Pteropoda	0	0	6	0	59	50	0	0	1	5	9	243	712	1085
			Sum	2779	677	1716	4577	7851	2830	1259	1474	1473	2721	2108	16968	867	47300

图 13: 不采用位置特征随机森林进行分类的结果

3.2.9 不包括位置特征和尺度特征 +SVM Linear

在 PkID 软件中，从 67 个特征中去掉位置特征和尺度特征，并采用 SVM Linear 进行训练和分类得到的结果如图 14，其分类准确率为 70.3%。

3.2.10 不包括位置特征和尺度特征 +RandomForest

在 PkID 软件中，从 67 个特征中去掉位置特征和尺度特征，并采用随机森林进行训练和分类得到的结果如图 15，其分类准确率为 76.4%。

3.2.11 不包括位置特征、尺度特征和灰度特征 +SVM Linear

在 PkID 软件中，从 67 个特征中去掉位置特征、尺度特征和灰度特征，并采用 SVM Linear 进行训练和分类得到的结果如图 16，其分类准确率为 69.2%。

Error rate			0.2618														
Values prediction			Confusion matrix														
Value	Recall	1-Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum
Appendicularia	0.8378	0.1904	Appendicularia	2283	0	89	4	10	10	0	0	43	4	122	160	0	2725
Bubble	0.8262	0.2826	Bubble	0	599	0	1	18	0	0	32	0	0	3	72	0	725
Chaetognatha	0.8968	0.0885	Chaetognatha	125	0	1565	0	0	5	0	0	16	1	16	16	1	1745
CladoceraPenilia	0.7825	0.2098	CladoceraPenilia	0	0	0	3619	23	0	5	0	0	33	7	938	0	4625
Copepoda	0.7996	0.2926	Copepoda	0	0	0	18	5893	182	0	0	1	1	184	1010	81	7370
Decapoda	0.6614	0.2570	Decapoda	0	0	0	0	531	1842	0	0	0	0	91	311	10	2785
Doliolida	0.7396	0.1778	Doliolida	0	0	0	7	0	0	1054	0	0	140	3	221	0	1425
Egg	0.6256	0.1639	Egg	0	155	0	31	42	1	0	1076	0	69	0	337	9	1720
Fiber	0.7342	0.2201	Fiber	55	0	27	2	0	0	0	2	1116	0	39	279	0	1520
Gelatinous	0.6924	0.2259	Gelatinous	5	0	2	112	19	11	142	38	3	2084	62	530	2	3010
Multiple	0.2989	0.4555	Multiple	236	0	19	129	587	181	11	9	60	87	949	895	12	3175
Nonbio	0.8029	0.2902	Nonbio	111	81	14	657	964	190	70	120	192	273	262	12357	99	15390
Pteropoda	0.4424	0.3084	Pteropoda	5	0	1	0	244	57	0	10	0	0	5	283	480	1085
			Sum	2820	835	1717	4580	8331	2479	1282	1287	1431	2692	1743	17409	694	47300

图 14: 不采用位置特征和尺度特征 SVM Linear 进行分类的结果

Error rate			0.2085														
Values prediction			Confusion matrix														
Value	Recall	1- Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum
Appendicularia	0.8015	0.2075	Appendicularia	2184	0	106	7	0	29	0	0	54	4	182	159	0	2725
Bubble	0.8414	0.1003	Bubble	0	610	0	0	2	0	0	34	0	0	0	79	0	725
Chaetognatha	0.8951	0.0924	Chaetognatha	133	0	1562	0	1	3	0	0	13	1	19	13	0	1745
CladoceraPenilia	0.8893	0.0848	CladoceraPenilia	0	0	0	4113	21	0	5	0	0	7	17	462	0	4625
Copepoda	0.8488	0.2019	Copepoda	2	0	0	5	6256	137	0	0	2	3	228	712	25	7370
Decapoda	0.8025	0.2213	Decapoda	1	0	0	0	300	2235	0	0	0	0	96	145	8	2785
Doliolida	0.7860	0.1257	Doliolida	0	0	0	2	0	0	1120	0	0	134	10	159	0	1425
Egg	0.7669	0.1009	Egg	0	61	0	7	4	1	1	1319	0	25	7	282	13	1720
Fiber	0.7493	0.2335	Fiber	49	0	18	0	0	0	1	0	1139	0	66	247	0	1520
Gelatinous	0.7226	0.1986	Gelatinous	7	0	0	54	8	5	69	24	0	2175	84	584	0	3010
Multiple	0.3282	0.5133	Multiple	277	0	20	39	479	219	7	15	52	100	1042	921	4	3175
Nonbio	0.8427	0.2359	Nonbio	102	7	12	267	712	187	78	75	226	261	379	12969	115	15390
Pteropoda	0.6599	0.1873	Pteropoda	1	0	3	0	56	54	0	0	0	4	11	240	716	1085
			Sum	2756	678	1721	4494	7839	2870	1281	1467	1486	2714	2141	16972	881	47300

图 15: 不采用位置特征和尺度特征随机森林进行分类的结果

Error rate			0.2691														
Values prediction			Confusion matrix														
Value	Recall	1- Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum
Appendicularia	0.8374	0.1945	Appendicularia	2282	0	93	3	11	10	0	0	42	5	126	153	0	2725
Bubble	0.8055	0.2930	Bubble	0	584	0	3	19	0	0	45	0	0	1	73	0	725
Chaetognatha	0.8980	0.0884	Chaetognatha	134	0	1567	0	0	7	0	0	10	0	14	13	0	1745
CladoceraPenilia	0.7619	0.2172	CladoceraPenilia	0	0	0	3524	23	0	4	0	0	46	6	1022	0	4625
Copepoda	0.7965	0.3007	Copepoda	0	0	0	17	5870	171	0	0	1	4	189	1021	97	7370
Decapoda	0.6309	0.2667	Decapoda	2	0	0	0	592	1757	0	0	0	0	97	327	10	2785
Doliolida	0.6975	0.1990	Doliolida	0	0	0	8	0	0	994	0	0	175	2	246	0	1425
Egg	0.6105	0.1758	Egg	0	168	0	36	43	3	0	1050	0	75	0	337	8	1720
Fiber	0.7349	0.2139	Fiber	54	0	25	1	0	0	0	2	1117	0	43	278	0	1520
Gelatinous	0.6857	0.2431	Gelatinous	5	0	1	108	20	10	149	44	4	2064	60	542	3	3010
Multiple	0.2961	0.4588	Multiple	241	0	18	130	593	189	10	13	57	88	940	883	13	3175
Nonbio	0.8026	0.2955	Nonbio	111	74	14	672	978	183	84	111	190	270	254	12352	97	15390
Pteropoda	0.4323	0.3271	Pteropoda	4	0	1	0	245	66	0	9	0	0	5	286	469	1085
			Sum	2833	826	1719	4502	8394	2396	1241	1274	1421	2727	1737	17533	697	47300

图 16: 不采用位置特征、尺度特征和灰度特征 SVM Linear 进行分类的结果

3.2.12 不包括位置特征、尺度特征和灰度特征 +RandomForest

在 PkID 软件中，从 67 个特征中去掉位置特征、尺度特征和灰度特征，并采用随机森林进行训练和分类得到的结果如图 17，其分类准确率为 77.2%。

Error rate			0.2056														
Values prediction			Confusion matrix														
Value	Recall	1- Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum
Appendicularia	0.8154	0.2076	Appendicularia	2222	0	110	5	0	17	0	0	41	5	166	159	0	2725
Bubble	0.8855	0.0802	Bubble	0	642	0	0	3	0	0	28	0	0	0	52	0	725
Chaetognatha	0.8957	0.0849	Chaetognatha	140	0	1563	0	0	4	0	0	4	0	22	12	0	1745
CladoceraPenilia	0.8878	0.0936	CladoceraPenilia	0	0	0	4106	22	0	5	0	0	6	16	470	0	4625
Copepoda	0.8514	0.2060	Copepoda	2	0	0	6	6275	132	0	0	2	3	223	709	18	7370
Decapoda	0.7932	0.2080	Decapoda	8	0	0	0	307	2209	0	0	0	0	103	149	9	2785
Doliolida	0.7832	0.1240	Doliolida	0	0	0	0	0	0	1116	0	0	157	10	142	0	1425
Egg	0.7878	0.1032	Egg	0	51	0	12	0	0	0	1355	0	21	2	277	2	1720
Fiber	0.7461	0.2141	Fiber	51	0	11	0	0	0	3	0	1134	0	59	262	0	1520
Gelatinous	0.7279	0.1945	Gelatinous	5	0	0	54	12	1	73	21	1	2191	91	561	0	3010
Multiple	0.3288	0.5153	Multiple	270	0	12	47	499	213	8	15	63	101	1044	901	2	3175
Nonbio	0.8428	0.2318	Nonbio	105	5	12	300	739	163	69	92	196	234	403	12970	102	15390
Pteropoda	0.6903	0.1508	Pteropoda	1	0	0	0	46	50	0	0	2	2	15	220	749	1085
			Sum	2804	698	1708	4530	7903	2789	1274	1511	1443	2720	2154	16884	882	47300

图 17: 不采用位置特征、尺度特征和灰度特征随机森林进行分类的结果

3.2.13 不包括位置特征、尺度特征、灰度特征和形状特征 +SVM Linear

在 PkID 软件中，从 67 个特征中去掉位置特征、尺度特征、灰度特征和形状特征，并采用 SVM Linear 进行训练和分类得到的结果如图 18，其分类准确率为 63.3%。

Error rate			0.3126														
Values prediction			Confusion matrix														
Value	Recall	1-Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum
Appendicularia	0.8198	0.2053	Appendicularia	2234	0	102	15	16	10	0	0	48	1	135	164	0	2725
Bubble	0.7545	0.3058	Bubble	0	547	0	9	25	0	0	83	0	0	0	61	0	725
Chaetognatha	0.8877	0.1041	Chaetognatha	129	0	1549	0	3	3	0	0	19	2	24	16	0	1745
CladoceraPenilia	0.6904	0.3092	CladoceraPenilia	0	0	0	3193	27	0	0	0	0	6	5	1394	0	4625
Copepoda	0.7501	0.3325	Copepoda	0	0	0	32	5528	159	0	0	3	0	166	1383	99	7370
Decapoda	0.5774	0.2674	Decapoda	2	0	0	0	618	1608	0	0	0	0	100	448	9	2785
Doliolida	0.5432	0.2110	Doliolida	0	0	0	4	0	0	774	0	0	242	3	402	0	1425
Egg	0.5453	0.2514	Egg	0	152	0	98	34	0	1	938	0	39	0	447	11	1720
Fiber	0.6947	0.2201	Fiber	62	0	35	2	3	0	0	0	1056	0	40	322	0	1520
Gelatinous	0.5571	0.2583	Gelatinous	16	0	0	247	36	17	108	7	2	1677	66	834	0	3010
Multiple	0.2743	0.4772	Multiple	267	2	28	176	567	162	19	4	54	88	871	929	8	3175
Nonbio	0.7907	0.3513	Nonbio	96	86	15	846	1064	159	79	145	172	206	251	12169	102	15390
Pteropoda	0.3419	0.3817	Pteropoda	5	1	0	0	361	77	0	76	0	0	5	189	371	1085
			Sum	2811	788	1729	4622	8282	2195	981	1253	1354	2261	1666	18758	600	47300

图 18: 不采用位置特征、尺度特征、灰度特征和形状特征 SVM Linear 进行分类的结果

3.2.14 不包括位置特征、尺度特征、灰度特征和形状特征 +RandomForest

在 PkID 软件中，从 67 个特征中去掉位置特征、尺度特征、灰度特征和形状特征，并采用随机森林进行训练和分类得到的结果如图 19，其分类准确率为 75.7%。

Error rate			0.2160														
Values prediction			Confusion matrix														
Value	Recall	1-Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum
Appendicularia	0.8187	0.2066	Appendicularia	2231	0	104	3	2	16	1	0	55	9	158	146	0	2725
Bubble	0.8910	0.0785	Bubble	0	646	0	0	2	0	0	30	0	0	1	46	0	725
Chaetognatha	0.8848	0.0891	Chaetognatha	129	0	1544	0	5	4	0	0	19	1	36	7	0	1745
CladoceraPenilia	0.8778	0.1304	CladoceraPenilia	0	0	0	4060	27	0	6	0	0	7	23	502	0	4625
Copepoda	0.8472	0.2132	Copepoda	1	0	0	22	6244	183	0	0	0	4	219	667	30	7370
Decapoda	0.7842	0.2175	Decapoda	7	0	0	0	326	2184	0	0	0	1	109	155	3	2785
Doliolida	0.7825	0.1449	Doliolida	0	0	0	0	0	0	1115	0	0	127	8	175	0	1425
Egg	0.7814	0.0937	Egg	0	52	0	6	1	0	0	1344	0	22	1	285	9	1720
Fiber	0.6980	0.2106	Fiber	65	0	25	0	0	0	3	0	1061	0	75	291	0	1520
Gelatinous	0.6688	0.2034	Gelatinous	13	0	0	88	18	7	74	12	6	2013	127	652	0	3010
Multiple	0.3291	0.5276	Multiple	281	2	12	53	496	198	15	16	40	90	1045	917	10	3175
Nonbio	0.8378	0.2407	Nonbio	85	1	10	437	720	168	90	79	163	253	397	12893	94	15390
Pteropoda	0.6461	0.1724	Pteropoda	0	0	0	0	95	31	0	2	0	0	13	243	701	1085
			Sum	2812	701	1695	4669	7936	2791	1304	1483	1344	2527	2212	16979	847	47300

图 19: 不采用位置特征、尺度特征、灰度特征和形状特征随机森林进行分类的结果

3.3 从 67 个特征中选取部分特征进行实验

3.3.1 选取 16 个特征 +SVM Linear

在 PkID 软件中，从 67 个特征中根据自己的判断选取部分特征进行（选取 16 个特征：Fractal、Skelarea、Circ、ESD、Elongation、Circexc、CV、SR、Skew、Kurt、MeanPos、PerimAreaexc、FeretAreaexc、PerimFeret、PerimMaj、CDexc），并采用 SVM Linear 进行训练和分类得到的结果如图 20。

Error rate			0.3391														
Values prediction			Confusion matrix														
Value	Recall	1- Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum
Appendicularia	0.8048	0.2218	Appendicularia	2193	0	101	20	3	8	0	0	28	13	129	227	3	2725
Bubble	0.7807	0.3213	Bubble	0	566	0	2	23	1	0	10	0	0	5	118	0	725
Chaetognatha	0.8315	0.1004	Chaetognatha	193	0	1451	4	2	1	0	0	8	5	38	43	0	1745
CladoceraPenilia	0.5963	0.2995	CladoceraPenilia	0	0	0	2758	18	0	2	0	0	56	0	1791	0	4625
Copepoda	0.7004	0.3956	Copepoda	0	0	0	5	5162	234	0	0	5	10	155	1779	20	7370
Decapoda	0.5059	0.2905	Decapoda	4	0	0	0	986	1409	0	1	0	0	78	306	1	2785
Doliolida	0.5726	0.3131	Doliolida	0	0	0	4	0	0	816	0	0	257	0	348	0	1425
Egg	0.5721	0.2096	Egg	0	157	0	27	42	7	0	984	0	70	1	419	13	1720
Fiber	0.6428	0.2184	Fiber	53	0	26	9	26	0	0	3	977	1	39	386	0	1520
Gelatinous	0.6063	0.3074	Gelatinous	6	0	0	133	40	8	204	95	0	1825	86	612	1	3010
Multiple	0.2485	0.4933	Multiple	251	0	21	194	659	156	19	18	57	119	789	887	5	3175
Nonbio	0.7822	0.3768	Nonbio	114	111	14	781	1205	118	147	131	175	279	232	12038	45	15390
Pteropoda	0.2682	0.2322	Pteropoda	4	0	0	0	375	44	0	3	0	0	5	363	291	1085
			Sum	2818	834	1613	3937	8541	1986	1188	1245	1250	2635	1557	19317	379	47300

图 20: 选取 16 个特征采用 SVM Linear 进行分类的结果

3.3.2 选取 16 个特征 +RandomForest

在 PkID 软件中，从 67 个特征中根据自己的判断选取部分特征进行（选取 16 个特征：Fractal、Skelarea、Circ、ESD、Elongation、Circexc、CV、SR、Skew、Kurt、MeanPos、PerimAreaexc、FeretAreaexc、PerimFeret、PerimMaj、CDexc），并采用随机森林进行训练和分类得到的结果如图 21，其分类准确率为 72.76%。

Error rate			0.2404														
Values prediction			Confusion matrix														
Value	Recall	1- Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum
Appendicularia	0.8018	0.2295	Appendicularia	2185	0	111	14	1	12	0	0	42	4	188	167	1	2725
Bubble	0.8097	0.1186	Bubble	0	587	0	0	2	0	0	37	0	0	4	95	0	725
Chaetognatha	0.8562	0.0890	Chaetognatha	161	0	1494	0	5	3	0	0	23	0	30	29	0	1745
CladoceraPenilia	0.8603	0.1373	CladoceraPenilia	0	0	0	3979	7	0	16	0	0	20	22	581	0	4625
Copepoda	0.8303	0.2521	Copepoda	4	0	0	7	6119	186	0	1	0	5	242	760	46	7370
Decapoda	0.7382	0.2135	Decapoda	6	0	0	1	462	2056	0	1	0	0	102	149	8	2785
Doliolida	0.7495	0.1772	Doliolida	0	0	0	7	0	0	1068	0	0	169	11	170	0	1425
Egg	0.7349	0.1684	Egg	0	60	0	8	2	5	1	1264	0	45	16	308	11	1720
Fiber	0.6868	0.2457	Fiber	50	0	18	0	1	0	0	0	1044	0	75	330	2	1520
Gelatinous	0.6741	0.2329	Gelatinous	6	0	0	78	15	3	122	52	1	2029	102	601	1	3010
Multiple	0.3030	0.5603	Multiple	291	0	6	110	525	181	14	25	49	103	962	898	11	3175
Nonbio	0.8116	0.2572	Nonbio	132	19	11	408	884	145	77	140	225	267	426	12490	166	15390
Pteropoda	0.6028	0.2733	Pteropoda	1	0	0	0	159	23	0	0	0	3	8	237	654	1085
			Sum	2836	666	1640	4612	8182	2614	1298	1520	1384	2645	2188	16815	900	47300

图 21: 选取 16 个特征采用随机森林进行分类的结果

3.3.3 选取 18 个特征 +SVM Linear

在 PkID 软件中，从 67 个特征中根据自己的判断选取部分特征进行（选取 18 个特征：Fractal、Skelarea、Circ、ESD、Elongation、Circexc、CV、SR、Skew、Kurt、MeanPos、PerimAreaexc、FeretAreaexc、PerimFeret、PerimMaj、CDexc、Symetrieih、Symetriev），并采用 SVM Linear 进行训练和分类得到的结果如图 22。

3.3.4 选取 18 个特征 +RandomForest

在 PkID 软件中，从 67 个特征中根据自己的判断选取部分特征进行（选取 18 个特征：Fractal、Skelarea、Circ、ESD、Elongation、Circexc、CV、SR、Skew、Kurt、MeanPos、PerimAreaexc、FeretAreaexc、PerimFeret、PerimMaj、CDexc、Symetrieih、Symetriev），并采用随机森林进行训练和分类得到的结果如图 23，其分类准确率为 74.59%。

3.3.5 选取 19 个特征 +SVM Linear

在 PkID 软件中，从 67 个特征中根据自己的判断选取部分特征进行（选取 19 个特征：Fractal、Skelarea、Circ、ESD、Elongation、Circexc、CV、SR、Skew、Kurt、MeanPos、PerimAreaexc、FeretAreaexc、

Error rate			0.3111															
Values prediction			Confusion matrix															
Value	Recall	1-Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum	
Appendicularia	0.8213	0.2030	Appendicularia	2238	0	96	4	6	8	0	0	43	5	127	194	4	2725	
Bubble	0.8041	0.3244	Bubble	0	583	0	0	26	3	0	11	0	0	5	97	0	725	
Chaetognatha	0.8877	0.0899	Chaetognatha	139	0	1549	0	3	0	0	0	16	0	13	25	0	1745	
CladoceraPenilia	0.6986	0.2502	CladoceraPenilia	0	0	0	3231	33	0	5	0	0	65	0	1291	0	4625	
Copepoda	0.7228	0.3449	Copepoda	0	0	0	15	5327	168	0	0	0	2	175	1661	22	7370	
Decapoda	0.5436	0.2885	Decapoda	5	0	0	0	707	1514	0	0	0	0	91	466	2	2785	
Doliolida	0.5811	0.3106	Doliolida	0	0	0	8	0	0	828	0	0	271	0	318	0	1425	
Egg	0.5773	0.1894	Egg	0	168	0	27	48	1	0	993	0	70	0	403	10	1720	
Fiber	0.6967	0.2392	Fiber	54	0	29	2	5	1	0	2	1059	3	41	324	0	1520	
Gelatinous	0.6355	0.2875	Gelatinous	7	0	0	158	50	6	202	84	2	1913	60	526	2	3010	
Multiple	0.2583	0.4699	Multiple	248	0	12	129	585	185	14	18	66	90	820	1003	5	3175	
Nonbio	0.7938	0.3537	Nonbio	113	112	16	735	1048	162	152	110	206	266	205	12216	49	15390	
Pteropoda	0.2894	0.2304	Pteropoda	4	0	0	0	294	80	0	7	0	0	10	376	314	1085	
			Sum	2808	863	1702	4309	8132	2128	1201	1225	1392	2685	1547	18900	408	47300	

图 22: 选取 18 个特征采用 SVM Linear 进行分类的结果

Error rate			0.2270															
Values prediction			Confusion matrix															
Value	Recall	1-Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum	
Appendicularia	0.8143	0.2189	Appendicularia	2219	0	110	3	4	11	0	0	46	5	186	141	0	2725	
Bubble	0.8138	0.1246	Bubble	0	590	0	0	4	0	0	36	0	0	3	92	0	725	
Chaetognatha	0.8699	0.0894	Chaetognatha	172	0	1518	0	5	0	0	0	7	1	24	18	0	1745	
CladoceraPenilia	0.8720	0.1189	CladoceraPenilia	0	0	0	4033	11	0	11	0	0	24	16	530	0	4625	
Copepoda	0.8347	0.2228	Copepoda	2	0	0	9	6152	171	0	0	0	3	229	779	25	7370	
Decapoda	0.7709	0.2136	Decapoda	2	0	0	2	356	2147	0	1	0	1	107	162	7	2785	
Doliolida	0.7347	0.1704	Doliolida	0	0	0	4	0	0	1047	0	0	179	4	191	0	1425	
Egg	0.7297	0.1572	Egg	0	58	0	12	3	3	2	1255	0	43	14	317	13	1720	
Fiber	0.7342	0.2398	Fiber	50	0	10	0	5	0	1	0	1116	1	60	277	0	1520	
Gelatinous	0.6867	0.2296	Gelatinous	10	0	0	81	20	3	105	54	0	2067	74	594	2	3010	
Multiple	0.3187	0.5188	Multiple	257	1	16	62	495	194	11	20	60	82	1012	953	12	3175	
Nonbio	0.8221	0.2528	Nonbio	129	25	11	371	806	167	85	122	238	273	366	12652	145	15390	
Pteropoda	0.6949	0.2129	Pteropoda	0	0	2	0	55	34	0	1	1	4	8	226	754	1085	
			Sum	2841	674	1667	4577	7916	2730	1262	1489	1468	2683	2103	16932	958	47300	

图 23: 选取 18 个特征采用随机森林进行分类的结果

PerimFeret、PerimMaj、CDexc、Mean、stdDev、Mode)，并采用 SVM Linear 进行训练和分类得到的结果如图 24。

3.3.6 选取 19 个特征 +RandomForest

在 PkID 软件中，从 67 个特征中根据自己的判断选取部分特征进行（选取 19 个特征：Fractal、Skelarea、Circ、ESD、Elongation、Circexc、CV、SR、Skew、Kurt、MeanPos、PerimAreaexc、FeretAreaexc、PerimFeret、PerimMaj、CDexc、Mean、stdDev、Mode），并采用随机森林进行训练和分类得到的结果如图 25，其分类准确率为 73.04%。

Error rate			0.3290															
Values prediction			Confusion matrix															
Value	Recall	1-Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum	
Appendicularia	0.8070	0.2210	Appendicularia	2199	0	107	27	1	8	0	0	26	6	140	211	0	2725	
Bubble	0.7448	0.3366	Bubble	0	540	0	6	20	1	0	52	0	0	4	102	0	725	
Chaetognatha	0.8332	0.0975	Chaetognatha	193	0	1454	3	6	0	0	0	11	5	37	36	0	1745	
CladoceraPenilia	0.6067	0.2967	CladoceraPenilia	0	0	0	2806	16	0	3	0	0	57	2	1741	0	4625	
Copepoda	0.7288	0.3900	Copepoda	0	0	0	7	5371	262	0	0	3	5	177	1452	93	7370	
Decapoda	0.5436	0.2932	Decapoda	4	0	0	0	983	1514	0	0	0	0	73	205	6	2785	
Doliolida	0.5881	0.2922	Doliolida	0	0	0	4	0	0	838	0	0	253	1	329	0	1425	
Egg	0.5698	0.2379	Egg	0	172	0	31	49	3	1	980	0	65	1	409	9	1720	
Fiber	0.6520	0.2122	Fiber	56	0	23	9	18	0	0	2	991	2	40	379	0	1520	
Gelatinous	0.6116	0.3008	Gelatinous	6	0	0	121	37	12	195	103	0	1841	81	611	3	3010	
Multiple	0.2683	0.4867	Multiple	250	0	14	186	710	164	18	19	51	116	852	789	6	3175	
Nonbio	0.7760	0.3539	Nonbio	112	102	13	790	1246	140	129	125	176	283	247	11943	84	15390	
Pteropoda	0.3770	0.3295	Pteropoda	3	0	0	0	348	38	0	5	0	0	5	277	409	1085	
			Sum	2823	814	1611	3990	8805	2142	1184	1286	1258	2633	1660	18484	610	47300	

图 24: 选取 19 个特征采用 SVM Linear 进行分类的结果

Error rate			0.2384															
Values prediction			Confusion matrix															
Value	Recall	1-Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum	
Appendicularia	0.7908	0.2315	Appendicularia	2155	0	125	15	4	12	0	0	54	8	186	166	0	2725	
Bubble	0.8234	0.0982	Bubble	0	597	0	0	1	0	0	27	0	0	3	97	0	725	
Chaetognatha	0.8619	0.0967	Chaetognatha	163	0	1504	0	5	1	0	0	19	0	32	20	1	1745	
CladoceraPenilia	0.8605	0.1329	CladoceraPenilia	0	0	0	3980	16	0	22	0	0	19	23	565	0	4625	
Copepoda	0.8259	0.2459	Copepoda	1	0	0	5	6087	197	0	1	0	7	263	765	44	7370	
Decapoda	0.7594	0.2334	Decapoda	2	0	0	0	424	2115	0	1	0	0	89	146	8	2785	
Doliolida	0.7544	0.1831	Doliolida	0	0	0	4	0	0	1075	0	0	154	15	177	0	1425	
Egg	0.7320	0.1476	Egg	0	44	0	6	3	10	1	1259	0	47	17	324	9	1720	
Fiber	0.6921	0.2448	Fiber	44	0	17	1	4	0	1	0	1052	0	67	333	1	1520	
Gelatinous	0.6824	0.2313	Gelatinous	11	0	0	74	22	2	108	42	0	2054	106	591	0	3010	
Multiple	0.3017	0.5606	Multiple	295	1	9	93	521	205	18	22	48	104	958	893	8	3175	
Nonbio	0.8149	0.2559	Nonbio	133	20	9	412	823	189	91	125	219	276	413	12541	139	15390	
Pteropoda	0.5954	0.2453	Pteropoda	0	0	1	0	162	28	0	0	1	3	8	236	646	1085	
			Sum	2804	662	1665	4590	8072	2759	1316	1477	1393	2672	2180	16854	856	47300	

图 25: 选取 19 个特征采用随机森林进行分类的结果

3.3.7 选取 21 个特征 +SVM Linear

在 PkID 软件中，从 67 个特征中根据自己的判断选取部分特征进行（选取 21 个特征：Fractal、Skelarea、Circ、ESD、Elongation、Circexc、CV、SR、Skew、Kurt、MeanPos、PerimAreaexc、FeretAreaexc、PerimFeret、PerimMaj、CDexc、Symetrieih、Symetriev、Mean、stdDev、Mode），并采用 SVM Linear 进行训练和分类得到的结果如图 26。

3.3.8 选取 21 个特征 +RandomForest

在 PkID 软件中，从 67 个特征中根据自己的判断选取部分特征进行（选取 21 个特征：Fractal、Skelarea、Circ、ESD、Elongation、Circexc、CV、SR、Skew、Kurt、MeanPos、PerimAreaexc、FeretAreaexc、

Error rate			0.3008															
Values prediction			Confusion matrix															
Value	Recall	1-Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum	
Appendicularia	0.8253	0.2028	Appendicularia	2249	0	101	7	8	11	0	0	48	4	139	156	2	2725	
Bubble	0.7766	0.3249	Bubble	0	563	0	3	17	2	0	48	0	0	6	86	0	725	
Chaetognatha	0.8865	0.0916	Chaetognatha	141	0	1547	0	8	0	0	0	15	0	17	17	0	1745	
CladoceraPenilia	0.7053	0.2503	CladoceraPenilia	0	0	0	3262	21	0	7	0	0	69	1	1265	0	4625	
Copepoda	0.7419	0.3313	Copepoda	0	0	0	10	5468	180	0	0	0	2	195	1418	97	7370	
Decapoda	0.5662	0.2871	Decapoda	4	0	0	0	711	1577	0	0	0	0	100	385	8	2785	
Doliolida	0.5951	0.2886	Doliolida	0	0	0	7	0	0	848	0	0	258	1	311	0	1425	
Egg	0.5791	0.2157	Egg	0	170	0	36	47	2	0	996	0	79	1	378	11	1720	
Fiber	0.7079	0.2380	Fiber	57	0	25	1	4	0	0	2	1076	1	43	311	0	1520	
Gelatinous	0.6565	0.2809	Gelatinous	6	0	0	150	27	5	187	86	1	1976	60	511	1	3010	
Multiple	0.2781	0.4744	Multiple	247	0	13	135	596	185	16	20	62	89	883	922	7	3175	
Nonbio	0.7928	0.3316	Nonbio	114	101	16	740	996	181	134	110	210	270	223	12201	94	15390	
Pteropoda	0.3926	0.3406	Pteropoda	3	0	1	0	274	69	0	8	0	0	11	293	426	1085	
			Sum	2821	834	1703	4351	8177	2212	1192	1270	1412	2748	1680	18254	646	47300	

图 26: 选取 21 个特征采用 SVM Linear 进行分类的结果

PerimFeret、PerimMaj、CDexc、Symetrieih、Symetriev、Mean、stdDev、Mode), 并采用随机森林进行训练和分类得到的结果如图 27, 其分类准确率为 74.91%。

Error rate			0.2240															
Values prediction			Confusion matrix															
Value	Recall	1-Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum	
Appendicularia	0.7949	0.2256	Appendicularia	2166	0	148	3	0	10	0	0	50	2	191	155	0	2725	
Bubble	0.8124	0.0952	Bubble	0	589	0	0	6	0	0	31	0	0	1	98	0	725	
Chaetognatha	0.8665	0.1153	Chaetognatha	172	0	1512	0	5	1	0	0	9	0	22	24	0	1745	
CladoceraPenilia	0.8765	0.1225	CladoceraPenilia	0	0	0	4054	10	0	15	0	0	19	22	505	0	4625	
Copepoda	0.8452	0.2216	Copepoda	4	0	0	10	6229	167	0	0	0	4	194	737	25	7370	
Decapoda	0.7724	0.2089	Decapoda	5	0	0	0	349	2151	0	0	0	0	92	180	8	2785	
Doliolida	0.7600	0.1752	Doliolida	0	0	0	5	0	0	1083	0	0	153	11	173	0	1425	
Egg	0.7448	0.1262	Egg	0	46	0	8	5	1	2	1281	0	44	15	307	11	1720	
Fiber	0.7401	0.2388	Fiber	50	0	23	0	1	0	2	1	1125	0	46	272	0	1520	
Gelatinous	0.7013	0.2170	Gelatinous	11	0	0	75	20	0	104	25	0	2111	83	577	4	3010	
Multiple	0.3068	0.5233	Multiple	264	0	14	72	526	180	15	20	69	88	974	948	5	3175	
Nonbio	0.8237	0.2489	Nonbio	125	16	10	393	792	173	92	108	225	271	385	12676	124	15390	
Pteropoda	0.6940	0.1903	Pteropoda	0	0	2	0	59	36	0	0	0	4	7	224	753	1085	
			Sum	2797	651	1709	4620	8002	2719	1313	1466	1478	2696	2043	16876	930	47300	

图 27: 选取 21 个特征采用随机森林进行分类的结果

3.3.9 选取 25 个特征 +SVM Linear

在 PkID 软件中, 从 67 个特征中根据自己的判断选取部分特征进行 (选取 25 个特征: Fractal、Skelarea、Circ、ESD、Elongation、Circexc、CV、SR、Skew、Kurt、MeanPos、PerimAreaexc、FeretAreaexc、PerimFeret、PerimMaj、CDexc、Symetrieih、Symetriev、Mean、stdDev、Mode、Symetrieihc、Symetrieivc、Fcons、ThickR), 并采用 SVM Linear 进行训练和分类得到的结果如图 28。

Error rate			0.2834														
Values prediction			Confusion matrix														
Value	Recall	1-Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum
Appendicularia	0.8283	0.1957	Appendicularia	2257	0	104	3	11	10	0	0	40	5	121	174	0	2725
Bubble	0.8041	0.3125	Bubble	0	583	0	0	24	0	0	46	0	0	4	68	0	725
Chaetognatha	0.8860	0.0959	Chaetognatha	148	0	1546	0	0	4	0	0	12	0	18	17	0	1745
CladoceraPenilia	0.7332	0.2300	CladoceraPenilia	0	0	0	3391	27	0	9	0	0	61	1	1136	0	4625
Copepoda	0.7859	0.3068	Copepoda	0	0	0	17	5792	162	0	0	2	3	171	1123	100	7370
Decapoda	0.6083	0.2720	Decapoda	2	0	0	0	624	1694	0	0	0	0	94	361	10	2785
Doliolida	0.6232	0.2839	Doliolida	0	0	0	7	0	0	888	0	0	258	1	271	0	1425
Egg	0.6006	0.1749	Egg	0	183	0	33	41	1	0	1033	0	68	0	351	10	1720
Fiber	0.7329	0.2215	Fiber	50	0	34	1	0	0	0	2	1114	0	38	281	0	1520
Gelatinous	0.6591	0.2799	Gelatinous	5	0	0	142	44	6	186	46	4	1984	63	529	1	3010
Multiple	0.2693	0.4571	Multiple	239	0	11	120	583	183	12	13	65	93	855	992	9	3175
Nonbio	0.7991	0.3131	Nonbio	101	82	14	690	977	197	145	103	194	283	204	12298	102	15390
Pteropoda	0.4258	0.3343	Pteropoda	4	0	1	0	232	70	0	9	0	0	5	302	462	1085
			Sum	2806	848	1710	4404	8355	2327	1240	1252	1431	2755	1575	17903	694	47300

图 28: 选取 25 个特征采用 SVM Linear 进行分类的结果

3.3.10 选取 25 个特征 +RandomForest

在 PkID 软件中, 从 67 个特征中根据自己的判断选取部分特征进行 (选取 25 个特征: Fractal、Skelarea、Circ、ESD、Elongation、Circexc、CV、SR、Skew、Kurt、MeanPos、PerimAreaexc、FeretAreaexc、PerimFeret、PerimMaj、CDexc、Symetrieih、Symetriev、Mean、stdDev、Mode、Symetrieihc、Symetrieivc、Fcons、ThickR), 并采用随机森林进行训练和分类得到的结果如图 29, 其分类准确率为 76.5%。

Error rate			0.2133														
Values prediction			Confusion matrix														
Value	Recall	1-Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum
Appendicularia	0.8015	0.2138	Appendicularia	2184	0	126	3	1	13	0	0	57	6	189	144	2	2725
Bubble	0.9090	0.0948	Bubble	0	659	0	0	1	0	0	19	0	0	2	44	0	725
Chaetognatha	0.8716	0.0957	Chaetognatha	169	0	1521	0	0	2	0	0	17	0	19	17	0	1745
CladoceraPenilia	0.8804	0.1125	CladoceraPenilia	0	0	0	4072	20	0	19	0	0	20	23	471	0	4625
Copepoda	0.8495	0.2066	Copepoda	1	0	0	8	6261	131	0	0	2	2	212	738	15	7370
Decapoda	0.7835	0.2068	Decapoda	6	0	0	0	327	2182	0	0	0	0	96	170	4	2785
Doliolida	0.7593	0.1467	Doliolida	0	0	0	5	0	0	1082	0	0	156	13	169	0	1425
Egg	0.7878	0.1038	Egg	0	62	0	10	0	0	3	1355	0	40	1	248	1	1720
Fiber	0.7533	0.2437	Fiber	51	0	17	0	1	0	0	0	1145	0	59	247	0	1520
Gelatinous	0.7080	0.2254	Gelatinous	5	0	0	72	25	1	76	39	2	2131	99	557	3	3010
Multiple	0.3156	0.5278	Multiple	257	3	3	70	489	214	15	16	73	101	1002	923	9	3175
Nonbio	0.8361	0.2354	Nonbio	105	4	10	348	722	170	73	83	217	292	395	12868	103	15390
Pteropoda	0.6894	0.1548	Pteropoda	0	0	5	0	44	38	0	0	1	3	12	234	748	1085
			Sum	2778	728	1682	4588	7891	2751	1268	1512	1514	2751	2122	16830	885	47300

图 29: 选取 25 个特征采用随机森林进行分类的结果

3.3.11 选取 27 个特征 +SVM Linear

在 PkID 软件中, 从 67 个特征中根据自己的判断选取部分特征进行 (选取 27 个特征: Fractal、Skelarea、Circ、ESD、Elongation、Circexc、CV、SR、Skew、Kurt、MeanPos、PerimAreaexc、FeretAreaexc、

PerimFeret、PerimMaj、CDexc、Symetrieih、Symetriev、Mean、stdDev、Mode、Symetrieihc、Symetrieivc、Fcons、ThickR、Convperim、Convarea)，并采用 SVM Linear 进行训练和分类得到的结果如图 30。

Error rate			0.2819														
Values prediction			Confusion matrix														
Value	Recall	1-Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum
Appendicularia	0.8341	0.1937	Appendicularia	2273	0	94	3	11	9	0	0	37	5	122	171	0	2725
Bubble	0.8041	0.3084	Bubble	0	583	0	0	24	0	0	46	0	0	3	69	0	725
Chaetognatha	0.8888	0.0909	Chaetognatha	143	0	1551	0	0	8	0	0	15	0	16	12	0	1745
CladoceraPenilia	0.7325	0.2305	CladoceraPenilia	0	0	0	3388	29	0	10	0	0	60	1	1137	0	4625
Copepoda	0.7855	0.3078	Copepoda	0	0	0	17	5789	155	0	0	3	2	171	1136	97	7370
Decapoda	0.6144	0.2676	Decapoda	2	0	0	0	611	1711	0	0	0	0	94	357	10	2785
Doliolida	0.6344	0.2745	Doliolida	0	0	0	7	0	0	904	0	0	242	0	272	0	1425
Egg	0.6023	0.1738	Egg	0	180	0	34	41	1	0	1036	0	67	0	352	9	1720
Fiber	0.7322	0.2184	Fiber	60	0	26	1	0	0	0	2	1113	0	40	278	0	1520
Gelatinous	0.6625	0.2728	Gelatinous	5	0	1	139	37	7	183	45	4	1994	58	532	5	3010
Multiple	0.2731	0.4530	Multiple	233	0	17	122	584	189	11	13	61	90	867	978	10	3175
Nonbio	0.7992	0.3126	Nonbio	98	80	16	692	998	186	138	103	191	282	208	12299	99	15390
Pteropoda	0.4221	0.3343	Pteropoda	5	0	1	0	239	70	0	9	0	0	5	298	458	1085
			Sum	2819	843	1706	4403	8363	2336	1246	1254	1424	2742	1585	17891	688	47300

图 30: 选取 27 个特征采用 SVM Linear 进行分类的结果

3.3.12 选取 27 个特征 +RandomForest

在 PKID 软件中，从 67 个特征中根据自己的判断选取部分特征进行（选取 27 个特征：Fractal、Skelarea、Circ、ESD、Elongation、Circexc、CV、SR、Skew、Kurt、MeanPos、PerimAreaexc、FeretAreaexc、PerimFeret、PerimMaj、CDexc、Symetrieih、Symetriev、Mean、stdDev、Mode、Symetrieihc、Symetrieivc、Fcons、ThickR、Convperim、Convarea），并采用随机森林进行训练和分类得到的结果如图 31，其分类准确率为 76.3%。

Error rate			0.2123														
Values prediction			Confusion matrix														
Value	Recall	1-Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum
Appendicularia	0.8040	0.2200	Appendicularia	2191	0	122	0	0	19	0	0	67	0	178	147	1	2725
Bubble	0.8690	0.0735	Bubble	0	630	0	0	5	0	0	30	0	0	3	57	0	725
Chaetognatha	0.8831	0.1015	Chaetognatha	148	0	1541	0	2	4	0	0	10	1	27	12	0	1745
CladoceraPenilia	0.8819	0.1140	CladoceraPenilia	0	0	0	4079	12	0	22	0	0	16	20	476	0	4625
Copepoda	0.8613	0.2072	Copepoda	3	0	0	12	6348	124	0	0	0	5	211	647	20	7370
Decapoda	0.7903	0.2048	Decapoda	5	0	0	0	323	2201	0	0	0	0	99	154	3	2785
Doliolida	0.7551	0.1567	Doliolida	0	0	0	5	0	0	1076	0	0	154	9	181	0	1425
Egg	0.7942	0.1025	Egg	0	47	0	9	2	0	1	1366	0	42	2	247	4	1720
Fiber	0.7329	0.2359	Fiber	58	0	24	0	0	0	1	0	1114	0	57	266	0	1520
Gelatinous	0.7050	0.2230	Gelatinous	10	0	0	80	24	1	92	23	0	2122	88	567	3	3010
Multiple	0.3165	0.5230	Multiple	287	1	15	67	499	216	9	10	60	104	1005	896	6	3175
Nonbio	0.8340	0.2320	Nonbio	107	2	12	352	738	164	75	93	207	283	400	12836	121	15390
Pteropoda	0.6922	0.1738	Pteropoda	0	0	1	0	54	39	0	0	0	4	8	228	751	1085
			Sum	2809	680	1715	4604	8007	2768	1276	1522	1458	2731	2107	16714	909	47300

图 31: 选取 27 个特征采用随机森林进行分类的结果

3.3.13 选取 29 个特征 + RandomForest

在 PkID 软件中，从 67 个特征中根据自己的判断选取部分特征进行（选取 29 个特征：Mean, StdDev, Fractal, Skelarea, Circ, Kurt, Skew, Symetrieih, Symetriev, Symetrieihc, Symetrieivc, Convperim, Convarea, Elongation, MeanPos, CV, SR, PermAreaexc, FeretAreaexc, PerimFeret, Circexc, CDexc, PerimMaj, ESD, ThickR, Fcons, Nb1, Nb2, Nb3），并采用随机森林进行训练和分类得到的结果如图 32，其分类准确率为 76.3%。

Error rate			0.2065														
Values prediction			Confusion matrix														
Value	Recall	1-Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum
Appendicularia	0.8158	0.2027	Appendicularia	2223	0	118	9	0	16	0	0	39	2	174	144	0	2725
Bubble	0.8993	0.0856	Bubble	0	652	0	0	2	0	0	24	0	0	1	46	0	725
Chaetognatha	0.8848	0.0960	Chaetognatha	142	0	1544	0	0	5	0	0	20	0	18	16	0	1745
CladoceraPenilia	0.8902	0.0920	CladoceraPenilia	0	0	0	4117	22	0	3	0	0	11	14	458	0	4625
Copepoda	0.8516	0.2045	Copepoda	3	0	0	8	6276	133	0	0	3	3	223	704	17	7370
Decapoda	0.7849	0.2097	Decapoda	4	0	0	0	325	2186	0	0	0	0	109	156	5	2785
Doliolida	0.7832	0.1129	Doliolida	0	0	0	0	0	0	1116	0	0	159	5	145	0	1425
Egg	0.7791	0.1037	Egg	0	56	0	9	0	0	0	1340	0	25	1	286	3	1720
Fiber	0.7441	0.2290	Fiber	59	0	22	0	0	0	0	0	1131	1	44	263	0	1520
Gelatinous	0.7256	0.2093	Gelatinous	8	0	2	49	18	1	71	26	0	2184	103	548	0	3010
Multiple	0.3169	0.5182	Multiple	255	4	6	54	484	219	9	17	64	104	1006	948	5	3175
Nonbio	0.8464	0.2327	Nonbio	92	1	11	288	723	155	59	88	210	270	375	13026	92	15390
Pteropoda	0.6756	0.1427	Pteropoda	2	0	5	0	39	51	0	0	0	3	15	237	733	1085
			Sum	2788	713	1708	4534	7889	2766	1258	1495	1467	2762	2088	16977	855	47300

图 32: 选取 29 个特征采用随机森林进行分类的结果

3.4 实验分析

- 位置特征、尺度特征和灰度特征对分类结果没有太大影响。
- 使用 PkID 中的特征能达到的最高分类识别率大概在 76% 左右。
- 我们自己的 SVM 和随机森林分类结果要比 PkID 中的分类结果高 1% 左右。
- 随机森林的分类准确率要比 SVM 要高。

根据以上实验，判断可以从 PkID 的 67 个特征中选取一下特征：Mean, StdDev, Fractal, Skelarea, Circ, Kurt, Skew, Symetrieih, Symetriev, Elongation, MeanPos, CV, SR, PermAreaexc, FeretAreaexc, PerimFeret, Circexc, CDexc, [Symetrieihc](#), [Symetrieivc](#), [Convperim](#), [Convarea](#), (PerimMaj, ESD, [ThickR](#), [Fcons](#), Nb1, Nb2, Nb3)。

前 22 个特征，在 PkID 软件采用 SVM Linear 进行训练和分类得到的结果如图 33，其分类准确率为 66.07%。

Error rate			0.2947														
Values prediction			Confusion matrix														
Value	Recall	1-Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum
Appendicularia	0.8316	0.1973	Appendicularia	2266	0	99	0	5	13	0	0	47	7	103	181	4	2725
Bubble	0.8483	0.2832	Bubble	0	615	0	2	25	0	0	7	0	0	5	71	0	725
Chaetognatha	0.8928	0.0968	Chaetognatha	133	0	1558	0	1	5	0	0	18	0	9	21	0	1745
CladoceraPenilia	0.6737	0.2847	CladoceraPenilia	0	0	0	3116	36	0	10	0	0	72	0	1391	0	4625
Copepoda	0.7902	0.3163	Copepoda	0	0	0	16	5824	139	0	0	1	2	172	1200	16	7370
Decapoda	0.5921	0.2671	Decapoda	3	0	0	0	656	1649	0	0	0	0	84	391	2	2785
Doliolida	0.6140	0.2792	Doliolida	0	0	0	13	0	0	875	0	0	255	1	281	0	1425
Egg	0.6110	0.1665	Egg	0	168	0	31	38	2	0	1051	0	69	0	351	10	1720
Fiber	0.7039	0.2229	Fiber	61	0	31	1	0	1	0	3	1070	1	33	319	0	1520
Gelatinous	0.6385	0.2871	Gelatinous	5	0	0	201	40	5	180	41	5	1922	56	554	1	3010
Multiple	0.2699	0.4351	Multiple	241	0	21	139	601	187	12	9	52	90	857	961	5	3175
Nonbio	0.7926	0.3319	Nonbio	109	75	15	837	1009	179	137	126	184	278	192	12198	51	15390
Pteropoda	0.3309	0.1987	Pteropoda	5	0	1	0	283	70	0	24	0	0	5	338	359	1085
			Sum	2823	858	1725	4356	8518	2250	1214	1261	1377	2696	1517	18257	448	47300

图 33: 选取 22 个特征采用 SVM Linear 进行分类的结果

前 22 个特征，用自己设计的 SVM Linear 进行训练和分类得到的结果如图 34。

	Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Total	Recall	1-Precision
Appendicularia	2309	0	88	4	6	14	0	0	49	6	89	157	3	2725	0.847339	0.199653
Bubble	0	633	0	3	24	1	0	14	0	0	6	44	0	725	0.873103	0.330867
Chaetognatha	134	0	1565	0	0	7	0	0	10	0	10	19	0	1745	0.896848	0.096942
CladoceraPenilia	0	0	0	3528	37	0	10	1	0	67	0	982	0	4625	0.762811	0.271074
Copepoda	0	0	0	22	6062	131	0	0	3	0	153	985	14	7370	0.822524	0.328236
Decapoda	3	0	0	0	706	1677	0	0	0	0	63	336	0	2785	0.602154	0.268005
Doliolida	0	0	0	28	0	0	966	0	0	209	0	222	0	1425	0.677895	0.253478
Egg	0	203	0	40	134	0	1	1013	0	58	0	260	11	1720	0.588953	0.171031
Fiber	53	0	38	2	0	1	0	5	1078	0	40	303	0	1520	0.709211	0.218274
Gelatinous	5	0	0	209	39	11	154	45	5	1951	56	535	0	3010	0.648173	0.269288
Multiple	273	0	25	164	603	184	15	14	49	100	826	917	5	3175	0.260157	0.42439
Nonbio	103	110	17	840	1044	174	148	108	185	279	187	12143	52	15390	0.789019	0.291334
Pteropoda	5	0	0	0	369	91	0	22	0	0	5	232	361	1085	0.332719	0.190583
Total	2885	946	1733	4840	9024	2291	1294	1222	1379	2670	1435	17135	446	47300	0.677762	0.254858

图 34: 选取 22 个特征采用 SVM Linear 进行分类的结果

前 22 个特征，在 PkID 软件采用随机森林进行训练和分类得到的结果如图 35，其分类准确率为 76.66%。

前 22 个特征，用自己设计的随机森林进行训练和分类得到的结果如图 35。

Error rate			0.2116														
Values prediction			Confusion matrix														
Value	Recall	1-Precision		Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Sum
Appendicularia	0.8136	0.2042	Appendicularia	2217	0	119	2	0	16	0	0	46	2	179	143	1	2725
Bubble	0.8952	0.0846	Bubble	0	649	0	0	3	0	0	29	0	0	1	43	0	725
Chaetognatha	0.8831	0.0957	Chaetognatha	137	0	1541	0	0	5	0	0	24	0	19	19	0	1745
CladoceraPenilia	0.8709	0.1188	CladoceraPenilia	0	0	0	4028	17	0	19	0	0	22	23	516	0	4625
Copepoda	0.8498	0.2076	Copepoda	2	0	0	10	6263	130	0	0	3	5	234	712	11	7370
Decapoda	0.7752	0.2057	Decapoda	8	0	0	0	332	2159	0	0	0	0	110	172	4	2785
Doliolida	0.7656	0.1490	Doliolida	0	0	0	7	0	0	1091	0	0	148	8	171	0	1425
Egg	0.7965	0.1087	Egg	0	54	0	10	0	0	4	1370	0	32	6	242	2	1720
Fiber	0.7500	0.2375	Fiber	63	0	14	0	1	0	0	0	1140	0	54	248	0	1520
Gelatinous	0.6977	0.2111	Gelatinous	8	0	0	84	25	4	91	30	2	2100	101	562	3	3010
Multiple	0.3294	0.5263	Multiple	253	1	12	79	504	210	6	21	72	99	1046	865	7	3175
Nonbio	0.8400	0.2321	Nonbio	96	5	12	351	712	153	71	87	207	253	413	12927	103	15390
Pteropoda	0.6986	0.1474	Pteropoda	2	0	6	0	47	41	0	0	1	1	14	215	758	1085
			Sum	2786	709	1704	4571	7904	2718	1282	1537	1495	2662	2208	16835	889	47300

图 35: 选取 22 个特征采用随机森林进行分类的结果

	Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Total	Recall	1-Precision
Appendicularia	2276	0	97	4	0	15	0	0	53	2	134	144	0	2725	0.835229	0.198592
Bubble	0	632	0	0	3	0	0	30	0	0	1	59	0	725	0.871724	0.063704
Chaetognatha	143	0	1549	0	1	5	0	0	12	0	22	12	1	1745	0.887679	0.080712
CladoceraPenilia	0	0	0	4051	15	0	26	0	0	12	17	504	0	4625	0.875892	0.106922
Copepoda	0	0	0	10	6334	137	0	0	0	4	208	660	17	7370	0.85943	0.189404
Decapoda	2	0	0	0	309	2216	0	0	0	0	92	163	3	2785	0.795691	0.208571
Doliolida	0	0	0	7	0	0	1092	0	0	149	3	174	0	1425	0.766316	0.137441
Egg	0	41	0	8	1	0	0	1414	0	33	1	218	4	1720	0.822093	0.097639
Fiber	53	0	12	0	0	0	1	0	1130	0	57	267	0	1520	0.743421	0.228669
Gelatinous	5	0	0	81	28	3	77	22	3	2064	86	638	3	3010	0.685714	0.208589
Multiple	274	1	12	72	458	222	12	11	61	94	1022	930	6	3175	0.32189	0.499265
Nonbio	87	1	11	303	629	163	58	90	206	249	387	13093	113	15390	0.850747	0.233745
Pteropoda	0	0	4	0	36	39	0	0	0	1	11	225	769	1085	0.708756	0.16048
Total	2840	675	1685	4536	7814	2800	1266	1567	1465	2608	2041	17087	916	47300	0.771122	0.185672

图 36: 选取 22 个特征采用随机森林进行分类的结果

4. 计算机视觉特征提取

4.1 几何参数

4.1.1 边界的周长

轮廓边界的周长。对轮廓边缘上的像素点的统计。

4.1.2 边界的曲率

4.1.3 面积

描述区域大小的特征。对区域内总像素点的统计。

4.1.4 宽度和高度

最小外接矩形的宽度和高度

4.1.5 矩形度

反映被检测目标的最小外接矩形的充满程度，当目标的形状越接近矩形时，矩形度的值越接近 1。

$$R = \frac{A}{WH}$$

A 为目标的面积，W、H 分别为最小外接矩形的宽度和高度。

4.1.6 体态比

为目标最小外接矩形的长与宽的比值。

$$C = \frac{W}{H}$$

4.1.7 圆形性

用目标区域的所有边界点定义的特征向量。

$$C_I = \frac{\mu_R}{\sigma_R}$$

μ_R 为区域重心到边界点的平均距离， σ_R 为从区域重心到边界点的距离的平均方差。

4.1.8 偏心率

在一定程度上反映了区域的紧凑程度。定义为目标区域长短主轴的平方根的比值。

$$E = \frac{p}{q}$$

设目标区域在 XY 平面上，区域像素点绕 X 轴的转动惯量为 A，绕 Y 轴的转动惯量为 B，惯性积为 C。目标区域的长度分别是 p 和 q。

$$p = \sqrt{\frac{2}{(A+B) + \sqrt{(A-B)^2 + 4C^2}}}$$

$$q = \sqrt{\frac{2}{(A+B) - \sqrt{(A-B)^2 + 4C^2}}}$$

4.1.9 凸率

为目标区域面积与目标区域凸包面积之比，该特征包含着描述边界不规则特性的信息。

$$C_R = \frac{A}{\sum_{x=1}^M \sum_{y=1}^N k(x, y)}$$

分母为凸包区域的面积。

4.1.10 密集度

描述目标密集度的量化特征，提供了目标形状的重要信息。在周长确定后，密集度越高，所围成的面积越大。

$$C_2 = \frac{L^2}{4\pi A}$$

L 为周长。

4.1.11 球状性

内切圆的直径与外接圆的直径之比。

$$S = \frac{r_i}{r_c}$$

4.1.12 伸长度

周长与目标区域最小外接矩形面积之比。

$$P = \frac{L}{WH}$$

4.1.13 叶状性

叶状反映了边界的幅度特征，为区域重心到边界的最短距离与目标区域的最大宽度之比。

$$B = \frac{R_1}{W_{max}}$$

4.2 几种典型的特征描述方法

4.2.1 多边形近似

4.2.2 骨架

4.2.3 形状数（边界）

4.2.4 统计矩（边界）

4.2.5 傅里叶描述子

4.2.6 小波描述子

4.2.7 拓扑描述（区域）

4.2.8 不变矩（区域）

4.2.9 统计方法（纹理）

4.2.10 结构性方法（纹理）

4.2.11 频谱方法（纹理）

4.2.12 曲率尺度空间

4.2.13 主分量描述

4.2.14 关系描述

5. 选取特征进行实验

5.1 Matlab

该部分进行的实验：用 Matlab 实现对浮游动物特征的提取（特征包括 PkID 中部分特征以及计算视觉中的一些特征提取方法），并进行分类。

5.1.1 实验一

去噪方法： 去掉连通区域小于 20 的噪声。

选用特征： Mean、StdDev、CV、SR、MeanPos、Elongation、Circ、Feret、PerimAreaexc、CDexc、Skelarea、FeretAreaexc、PerimFeret、矩形度、体态比、凸率、伸长度、灰度共生矩阵（对比度）、对称性（左右），前 13 个特征为 PkID 系统中使用的部分特征。

分类器： 采用随机森林进行训练和分类得到的结果如图 39，其分类准确率为 71.9%。采用 SVM Linear 进行训练和分类得到的结果如图 40，其分类准确率为 58.4%

	Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Total	Recall	1-Precision
Appendicularia	2105	0	173	39	12	14	0	0	60	0	151	171	0	2725	0.772477	0.255131
Bubble	0	620	0	0	5	0	0	12	0	0	0	88	0	725	0.855172	0.026688
Chaetognatha	203	0	1463	0	4	10	0	0	5	0	46	14	0	1745	0.838395	0.136873
CladoceraPenilia	7	0	0	3840	22	0	12	0	0	43	22	679	0	4625	0.83027	0.17366
Copepoda	3	0	0	8	6034	233	0	0	0	1	212	856	23	7370	0.818725	0.2389
Decapoda	6	0	0	0	464	2003	0	3	0	3	95	206	5	2785	0.71921	0.254836
Doliolida	0	0	0	4	1	0	1020	0	0	233	0	167	0	1425	0.715789	0.237668
Egg	0	11	0	12	9	9	0	1334	0	22	7	311	5	1720	0.775581	0.122368
Fiber	80	0	12	3	2	6	0	0	1091	0	59	266	1	1520	0.717763	0.224591
Gelatinous	5	0	0	153	8	8	198	32	0	1785	103	718	0	3010	0.593023	0.318962
Multiple	262	0	36	119	441	219	16	17	58	154	860	988	5	3175	0.270866	0.555326
Nonbio	155	6	11	469	783	163	92	122	193	376	366	12588	66	15390	0.817934	0.271233
Pteropoda	0	0	0	0	143	23	0	0	0	4	13	221	681	1085	0.62765	0.133588
Total	2826	637	1695	4647	7928	2688	1338	1520	1407	2621	1934	17273	786	47300	0.719451	0.22691

图 37: Matlab-19 个特征采用随机森林进行分类的结果

	Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Total	Recall	1-Precision
Appendicularia	2139	0	147	53	19	6	0	0	62	3	132	164	0	2725	0.784954	0.346271
Bubble	0	640	0	20	16	2	0	2	0	0	4	41	0	725	0.882759	0.126876
Chaetognatha	330	0	1360	0	0	11	0	0	11	0	24	9	0	1745	0.77937	0.168704
CladoceraPenilia	1	0	0	3296	28	0	5	0	0	34	14	1247	0	4625	0.712649	0.368098
Copepoda	3	0	1	32	5765	113	0	0	0	0	165	1268	23	7370	0.782225	0.416262
Decapoda	9	0	0	0	1209	1230	0	0	0	0	82	255	0	2785	0.441652	0.261261
Doliolida	0	0	0	5	0	0	612	0	0	558	2	248	0	1425	0.429474	0.431755
Egg	0	51	0	97	160	0	10	1084	0	5	4	299	10	1720	0.630233	0.202355
Fiber	250	0	55	6	39	11	0	0	799	0	36	324	0	1520	0.525658	0.24906
Gelatinous	9	0	1	473	74	3	285	2	0	1249	83	831	0	3010	0.41495	0.460475
Multiple	333	0	46	259	701	161	18	9	32	140	684	786	6	3175	0.215433	0.524008
Nonbio	198	42	26	975	1387	102	147	262	160	326	196	11541	28	15390	0.749903	0.333121
Pteropoda	0	0	0	0	478	26	0	0	0	0	11	293	277	1085	0.2553	0.194767
Total	3272	733	1636	5216	9876	1665	1077	1359	1064	2315	1437	17306	344	47300	0.584966	0.314078

图 38: Matlab-19 个特征采用 SVM Linear 进行分类的结果

5.1.2 实验二

去噪方法： 去掉连通区域小于 50 的噪声。

选用特征： 特征与实验一相同。

分类器： 采用随机森林进行训练和分类得到的结果如图 37，其分类准确率为 72.9%。采用 SVM Linear 进行训练和分类得到的结果如图 38，其分类准确率为 58.9%

	Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Total	Recall	1-Precision
Appendicularia	2106	0	166	41	10	16	0	0	57	0	152	177	0	2725	0.772844	0.241628
Bubble	0	624	0	0	5	0	0	14	0	0	0	82	0	725	0.86069	0.023474
Chaetognatha	173	0	1491	0	2	11	0	0	7	1	41	19	0	1745	0.854441	0.126026
CladoceraPenilia	4	0	0	3856	22	1	13	0	0	42	37	650	0	4625	0.83373	0.175716
Copepoda	2	0	1	15	6076	208	0	0	3	0	193	851	21	7370	0.824423	0.234182
Decapoda	6	0	0	0	453	2027	0	3	0	4	86	203	3	2785	0.727828	0.236822
Doliolida	0	0	0	1	0	0	1061	0	0	215	0	148	0	1425	0.744561	0.226112
Egg	0	10	0	13	2	7	0	1322	0	23	12	328	3	1720	0.768605	0.127393
Fiber	75	0	11	4	2	2	0	0	1081	1	55	286	3	1520	0.711184	0.229508
Gelatinous	5	0	0	148	10	6	194	34	0	1810	89	714	0	3010	0.601329	0.29736
Multiple	253	1	27	125	463	209	14	27	62	129	914	949	2	3175	0.287874	0.534149
Nonbio	153	4	10	475	770	144	89	115	193	348	367	12646	76	15390	0.821702	0.267026
Pteropoda	0	0	0	0	119	25	0	0	0	3	16	200	722	1085	0.665438	0.13012
Total	2777	639	1706	4678	7934	2656	1371	1515	1403	2576	1962	17253	830	47300	0.728819	0.219194

图 39: Matlab-19 个特征采用随机森林进行分类的结果

	Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Total	Recall	1-Precision
Appendicularia	2154	0	150	46	18	7	0	0	50	0	147	153	0	2725	0.790459	0.330432
Bubble	0	636	0	15	16	7	0	4	0	0	0	47	0	725	0.877241	0.105485
Chaetognatha	314	0	1371	0	1	11	0	0	6	1	33	8	0	1745	0.785673	0.164534
CladoceraPenilia	1	0	0	3321	34	0	8	0	0	24	13	1224	0	4625	0.718054	0.362082
Copepoda	1	0	0	32	5762	108	0	0	2	0	164	1283	18	7370	0.781818	0.417568
Decapoda	7	0	0	0	1209	1251	0	0	0	0	84	234	0	2785	0.449192	0.256684
Doliolida	3	0	0	5	0	0	606	0	0	547	2	262	0	1425	0.425263	0.437848
Egg	0	44	0	98	171	0	10	1053	0	6	5	328	5	1720	0.612209	0.2171
Fiber	224	0	52	7	30	10	1	0	823	1	39	333	0	1520	0.541447	0.22723
Gelatinous	7	0	0	445	76	2	260	14	0	1326	80	800	0	3010	0.440532	0.434059
Multiple	320	0	45	278	699	161	30	12	28	149	699	749	5	3175	0.220157	0.524813
Nonbio	186	31	23	959	1392	96	163	262	156	289	193	11614	26	15390	0.754646	0.32902
Pteropoda	0	0	0	0	485	30	0	0	0	0	12	274	284	1085	0.261751	0.159763
Total	3217	711	1641	5206	9893	1683	1078	1345	1065	2343	1471	17309	338	47300	0.589111	0.305125

图 40: Matlab-19 个特征采用 SVM Linear 进行分类的结果

根据实验一、二的分析：去掉小连通区域的噪声可以提高分类准确率，但是去掉的小连通区域面积阈值设置相差不大时不会对实验结果产生很大影响。

5.1.3 实验三

去噪方法： 去掉连通区域小于 50 的噪声。

选用特征： 在实验一、二所用特征的基础上增加的了不变矩特征。

分类器： 采用随机森林进行训练和分类得到的结果如图 41，其分类准确率为 73.7%。采用 SVM Linear 进行训练和分类得到的结果如图 42，其分类准确率为 61.0%

	Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Total	Recall	1-Precision
Appendicularia	2209	0	120	34	11	10	0	0	36	0	163	142	0	2725	0.810642	0.207391
Bubble	0	627	0	0	5	0	0	14	0	0	1	78	0	725	0.864828	0.021841
Chaetognatha	131	0	1537	0	2	5	0	0	15	0	39	16	0	1745	0.880802	0.108469
CladoceraPenilia	2	0	0	3902	23	3	6	0	0	19	18	652	0	4625	0.843676	0.161762
Copepoda	1	0	0	15	6112	195	0	0	0	0	191	833	23	7370	0.829308	0.227405
Decapoda	5	0	0	0	440	2033	0	4	0	4	78	215	6	2785	0.729982	0.227877
Doliolida	0	0	0	3	0	0	1073	0	0	205	0	144	0	1425	0.752982	0.207533
Egg	0	13	0	9	6	9	0	1358	0	20	5	300	0	1720	0.789535	0.100066
Fiber	64	0	13	2	1	1	0	0	1121	0	59	259	0	1520	0.7375	0.222068
Gelatinous	5	0	0	129	19	7	182	23	0	1800	88	757	0	3010	0.598007	0.291896
Multiple	247	0	33	112	461	197	13	18	62	147	875	1007	3	3175	0.275591	0.526771
Nonbio	123	1	21	449	719	150	80	92	207	346	317	12823	62	15390	0.833203	0.266083
Pteropoda	0	0	0	0	112	23	0	0	0	1	15	246	688	1085	0.634101	0.120205
Total	2787	641	1724	4655	7911	2633	1354	1509	1441	2542	1849	17472	782	47300	0.736935	0.206874

图 41: Matlab-20 个特征采用随机森林进行分类的结果

	Appendicularia	Bubble	Chaetognatha	CladoceraPenilia	Copepoda	Decapoda	Doliolida	Egg	Fiber	Gelatinous	Multiple	Nonbio	Pteropoda	Total	Recall	1-Precision
Appendicularia	2170	0	154	48	12	9	0	0	40	1	140	151	0	2725	0.79633	0.332513
Bubble	0	630	0	15	20	2	0	3	0	0	2	53	0	725	0.868966	0.111425
Chaetognatha	308	0	1379	0	0	11	0	0	8	1	28	10	0	1745	0.790258	0.163228
CladoceraPenilia	3	0	0	3356	18	0	8	0	0	16	15	1209	0	4625	0.725622	0.349108
Copepoda	2	0	0	34	5881	101	0	0	3	1	175	1142	31	7370	0.797965	0.372225
Decapoda	11	0	0	0	877	1488	0	5	0	0	83	321	0	2785	0.534291	0.237314
Doliolida	1	0	0	1	0	0	554	0	0	592	1	276	0	1425	0.388772	0.438134
Egg	0	43	0	75	52	9	10	1196	0	2	4	329	0	1720	0.695349	0.10412
Fiber	237	0	51	7	41	11	2	0	816	0	39	316	0	1520	0.536842	0.214629
Gelatinous	6	0	0	408	72	4	243	3	0	1373	70	831	0	3010	0.456146	0.438905
Multiple	318	0	42	280	674	163	19	3	25	152	718	774	7	3175	0.226142	0.51552
Nonbio	195	36	22	932	1274	125	150	125	147	309	193	11837	45	15390	0.769136	0.322322
Pteropoda	0	0	0	0	447	28	0	0	0	0	14	218	378	1085	0.348387	0.180043
Total	3251	709	1648	5156	9368	1951	986	1335	1039	2447	1482	17467	461	47300	0.610323	0.29073

图 42: Matlab-20 个特征采用 SVM Linear 进行分类的结果