

# 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
2.8	实验	12
2.8.1	形状特征 +SVM Linear	12
2.8.2	形状特征 +RandomForest	13
2.8.3	灰度特征 +SVM Linear	13
2.8.4	灰度特征 +RandomForest	13
2.8.5	生物统计特征 +SVM Linear	14
2.8.6	生物统计特征 +RandomForest	14
2.8.7	不包括位置特征 +SVM Linear	14
2.8.8	不包括位置特征 +RandomForest	15
2.8.9	不包括位置特征和尺度特征 +SVM Linear	15
2.8.10	不包括位置特征和尺度特征 +RandomForest	15
2.8.11	不包括位置特征、尺度特征和灰度特征 +SVM Linear	17
2.8.12	不包括位置特征、尺度特征和灰度特征 +RandomForest	17
2.8.13	选取部分特征 +SVM Linear	17

2.8.14	选取部分特征 +RandomForest	18
2.8.15	PkID+SVM Linear	18
2.8.16	PkID+RandomForest	19
<b>3</b>	<b>计算机视觉领域形态学常用特征</b>	<b>21</b>
3.1	形状表示	21
3.1.1	多边形近似	21
3.1.2	傅里叶描述子	21
3.1.3	统计矩	21
3.1.4	骨架	21
3.1.5	形状数	21
3.1.6	拓扑描述	21
3.1.7	关系描述	21
3.1.8	主分量描述	21
3.1.9	不变矩	21
3.1.10	矩形度	21
3.1.11	体态比	21
3.1.12	圆形性	21
3.1.13	偏心率	22
3.1.14	凸率	22
3.1.15	密集度	22
3.1.16	球状性	22
3.1.17	伸长度	23
3.1.18	叶状性	23
3.2	简单描述子	23
3.2.1	边界的周长	23
3.2.2	面积	23
3.2.3	宽度和高度	23
3.2.4	边界的曲率	23
3.3	纹理表示	23
3.3.1	统计方法	23
3.3.2	结构性方法	23

3.3.3 频谱方法 . . . . .	23
----------------------	----

## 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。<sup>1</sup> 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。<sup>2</sup> 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：

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

<sup>2</sup><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。<sup>3</sup>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。<sup>4</sup>

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）。<sup>5</sup>

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<sup>6</sup>。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：

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

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

<sup>5</sup>[http://baike.baidu.com/link?url=LWmrgD\\_DVUcwOupg\\_ziOLTlJWj6quxa\\_juRrS3zUt91A-FjPM6VQwYfZ5fFZckzIyEGCaXypikXUGg2JsYMXUX-uFEkmkLqC5lfxvXvApK3WRBcWQkfbDhMlfTdgrWvh-728gSoUylWZG2UstFK](http://baike.baidu.com/link?url=LWmrgD_DVUcwOupg_ziOLTlJWj6quxa_juRrS3zUt91A-FjPM6VQwYfZ5fFZckzIyEGCaXypikXUGg2JsYMXUX-uFEkmkLqC5lfxvXvApK3WRBcWQkfbDhMlfTdgrWvh-728gSoUylWZG2UstFK)

<sup>6</sup><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** 鱼卵以及其他浮游动物的卵。

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)** 胶质的浮游动物，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，属于原生动物门，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）。该类是多种呈胶质浮游动物的集合。大部分水母都有三个主要部位：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。

Texture Contrast

Cumulation Histogram

Convex Area

Symmetry

Thickness Ratio

## 2.8 实验

### 2.8.1 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 1，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 1: 采用形状特征 SVM Linear 进行分类的结果

## 2.8.2 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 2, 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 2: 采用形状特征随机森林进行分类的结果

## 2.8.3 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 3, 其分类准确率为 37.6%。

## 2.8.4 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 4, 其分类准确率为 61.3%。

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	Appendicularia	1884	0	178	102	36	15	0	0	0	49	100	361	0	2725
Bubble	0.0566	0.6168	Bubble	0	41	0	5	94	0	0	46	0	0	1	538	0	725
Chaetognatha	0.5433	0.2582	Chaetognatha	329	0	948	5	31	5	5	0	0	48	76	298	0	1745
CladoceraPenilia	0.4415	0.4855	CladoceraPenilia	0	0	0	2042	12	0	7	0	0	2	1	2561	0	4625
Copepoda	0.6836	0.4252	Copepoda	0	6	0	28	5038	301	0	30	0	0	81	1860	26	7370
Decapoda	0.5451	0.3286	Decapoda	0	0	0	3	737	1518	0	13	0	0	66	433	15	2785
Doliolida	0.3705	0.4477	Doliolida	1	0	0	10	0	0	528	0	0	63	6	817	0	1425
Egg	0.1994	0.6163	Egg	0	3	0	67	417	35	4	343	0	14	9	776	52	1720
Fiber	0.0020	0.0000	Fiber	49	0	14	40	26	0	19	0	3	6	22	1341	0	1520
Gelatinous	0.2066	0.4842	Gelatinous	78	1	52	476	26	18	217	3	0	622	144	1373	0	3010
Multiple	0.1644	0.5639	Multiple	276	2	61	331	683	162	35	12	0	153	522	930	8	3175
Nonbio	0.7843	0.4873	Nonbio	126	49	25	860	1309	164	141	185	0	249	163	12071	48	15390
Pteropoda	0.2037	0.4027	Pteropoda	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 3: 采用灰度特征 SVM Linear 进行分类的结果

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	Appendicularia	1951	0	235	54	24	25	4	0	36	41	150	205	0	2725
Bubble	0.8166	0.1834	Bubble	0	592	0	0	3	0	0	8	0	0	5	117	0	725
Chaetognatha	0.7278	0.2853	Chaetognatha	229	0	1270	0	11	3	5	0	5	46	73	99	4	1745
CladoceraPenilia	0.8666	0.1918	CladoceraPenilia	8	0	0	4008	30	0	19	2	0	49	39	470	0	4625
Copepoda	0.7659	0.3438	Copepoda	10	0	7	13	5645	408	0	9	0	11	196	987	84	7370
Decapoda	0.7260	0.3090	Decapoda	6	0	0	1	527	2022	0	5	0	2	86	114	22	2785
Doliolida	0.7200	0.2914	Doliolida	1	0	0	23	0	0	1026	0	0	187	24	164	0	1425
Egg	0.4645	0.2455	Egg	0	41	0	12	188	45	0	799	2	23	49	491	70	1720
Fiber	0.2605	0.3333	Fiber	75	0	9	37	28	0	12	0	396	6	34	923	0	1520
Gelatinous	0.4837	0.4245	Gelatinous	56	0	78	191	16	11	200	7	2	1456	203	789	1	3010
Multiple	0.2466	0.6139	Multiple	297	6	100	90	627	172	38	34	8	220	783	777	23	3175
Nonbio	0.7637	0.3136	Nonbio	126	86	78	526	1194	187	144	163	145	486	378	11753	124	15390
Pteropoda	0.4074	0.4260	Pteropoda	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 4: 采用灰度特征随机森林进行分类的结果

### 2.8.5 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 5, 其分类准确率为 52.4%。

### 2.8.6 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 6, 其分类准确率为 69.9%。

### 2.8.7 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 7, 其分类准确率为 70.8%。

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 5: 采用生物统计特征 SVM Linear 进行分类的结果

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 6: 采用生物统计特征随机森林进行分类的结果

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 7: 不采用位置特征 SVM Linear 进行分类的结果



## 2.8.8 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 8，其分类准确率为 76.1%。

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 8: 不采用位置特征随机森林进行分类的结果

## 2.8.9 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 9，其分类准确率为 70.3%。

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 9: 不采用位置特征和尺度特征 SVM Linear 进行分类的结果

## 2.8.10 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 10，其分类准确率为 76.4%。



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 10: 不采用位置特征和尺度特征随机森林进行分类的结果

### 2.8.11 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 11, 其分类准确率为 69.2%。

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 11: 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

### 2.8.12 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 12, 其分类准确率为 77.2%。

### 2.8.13 选取部分特征 +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、

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	2222	0	110	5	0	17	0	0	41	5	166	159	0	2725
Bubble	0.8855	0.0802	0	642	0	0	3	0	0	28	0	0	0	52	0	725
Chaetognatha	0.8957	0.0849	140	0	1563	0	0	4	0	0	4	0	22	12	0	1745
CladoceraPenilia	0.8878	0.0936	0	0	0	4106	22	0	5	0	0	6	16	470	0	4625
Copepoda	0.8514	0.2060	2	0	0	6	6275	132	0	0	2	3	223	709	18	7370
Decapoda	0.7932	0.2080	8	0	0	0	307	2209	0	0	0	0	103	149	9	2785
Doliolida	0.7832	0.1240	0	0	0	0	0	0	1116	0	0	157	10	142	0	1425
Egg	0.7878	0.1032	0	51	0	12	0	0	0	1355	0	21	2	277	2	1720
Fiber	0.7461	0.2141	51	0	11	0	0	0	3	0	1134	0	59	262	0	1520
Gelatinous	0.7279	0.1945	5	0	0	54	12	1	73	21	1	2191	91	561	0	3010
Multiple	0.3288	0.5153	270	0	12	47	499	213	8	15	63	101	1044	901	2	3175
Nonbio	0.8428	0.2318	105	5	12	300	739	163	69	92	196	234	403	12970	102	15390
Pteropoda	0.6903	0.1508	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 12: 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

PerimFeret、PerimMaj、CDexc), 并采用 SVM Linear 进行训练和分类得到的结果如图 13, 其分类准确率为 60.9%。

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	2193	0	101	20	3	8	0	0	28	13	129	227	3	2725
Bubble	0.7807	0.3213	0	566	0	2	23	1	0	10	0	0	5	118	0	725
Chaetognatha	0.8315	0.1004	193	0	1451	4	2	1	0	0	8	5	38	43	0	1745
CladoceraPenilia	0.5963	0.2995	0	0	0	2758	18	0	2	0	0	56	0	1791	0	4625
Copepoda	0.7004	0.3956	0	0	0	5	5162	234	0	0	5	10	155	1779	20	7370
Decapoda	0.5059	0.2905	4	0	0	0	986	1409	0	1	0	0	78	306	1	2785
Doliolida	0.5726	0.3131	0	0	0	4	0	0	816	0	0	257	0	348	0	1425
Egg	0.5721	0.2096	0	157	0	27	42	7	0	984	0	70	1	419	13	1720
Fiber	0.6428	0.2184	53	0	26	9	26	0	0	3	977	1	39	386	0	1520
Gelatinous	0.6063	0.3074	6	0	0	133	40	8	204	95	0	1825	86	612	1	3010
Multiple	0.2485	0.4933	251	0	21	194	659	156	19	18	57	119	789	887	5	3175
Nonbio	0.7822	0.3768	114	111	14	781	1205	118	147	131	175	279	232	12038	45	15390
Pteropoda	0.2682	0.2322	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 13: 选取 16 个特征采用 SVM Linear 进行分类的结果

#### 2.8.14 选取部分特征 + 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 14, 其分类准确率为 78.4%。

#### 2.8.15 PkID+SVM Linear

PkID 系统中 SVM Linear 的参数如图 15。

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 14: 选取 16 个特征采用随机森林进行分类的结果

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 15: PkID 系统中 SVM Linear 的参数

采用以上 67 个特征，并用 SVM Linear 分类器进行训练和分类得到的混淆矩阵如图 16。在采用 SVM 进行训练之前，要先对特征进行归一化。

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

PkID 软件中用 67 个特征和 SVM Linear 分类器进行训练和分类得到的结果如图 17，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	Appendicularia	2289	0	92	0	6	11	0	1	45	6	123	151	1	2725
Bubble	0.7959	0.2640	Bubble	0	577	0	0	20	0	0	52	0	0	3	73	0	725
Chaetognatha	0.8883	0.0941	Chaetognatha	133	0	1550	0	0	10	0	0	14	1	20	17	0	1745
CladoceraPenilia	0.7894	0.1993	CladoceraPenilia	0	0	0	3651	18	0	4	0	0	29	5	918	0	4625
Copepoda	0.7984	0.2842	Copepoda	0	0	0	17	5884	187	0	0	2	1	180	1011	88	7370
Decapoda	0.6747	0.2538	Decapoda	2	0	0	0	470	1879	0	0	0	0	88	339	7	2785
Doliolida	0.7502	0.1790	Doliolida	0	0	0	8	0	0	1069	0	0	131	2	215	0	1425
Egg	0.6640	0.1766	Egg	0	129	0	25	27	2	0	1142	0	65	0	320	10	1720
Fiber	0.7342	0.2146	Fiber	44	0	28	0	0	0	0	9	1116	0	39	284	0	1520
Gelatinous	0.6844	0.2238	Gelatinous	6	0	1	119	21	10	142	36	0	2060	73	540	2	3010
Multiple	0.3002	0.4646	Multiple	264	0	21	101	589	174	15	8	50	87	953	897	16	3175
Nonbio	0.8041	0.2890	Nonbio	105	78	16	639	946	185	72	123	194	274	288	12375	95	15390
Pteropoda	0.4525	0.3085	Pteropoda	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 17: PkID 软件中 SVM Linear 分类器分类结果

2.8.16 PkID+RandomForest

采用以上 67 个特征，并用随机森林分类器进行训练和分类得到的混淆矩阵如图 18。

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

PkID 软件中用 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 19，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 19: PkID 软件中随机森林分类器分类结果

### 3. 计算机视觉领域形态学常用特征

#### 3.1 形状表示

##### 3.1.1 多边形近似

##### 3.1.2 傅里叶描述子

##### 3.1.3 统计矩

##### 3.1.4 骨架

##### 3.1.5 形状数

##### 3.1.6 拓扑描述

##### 3.1.7 关系描述

##### 3.1.8 主分量描述

##### 3.1.9 不变矩

##### 3.1.10 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。

##### 3.1.11 体态比

为目标最小外接矩形的长与宽的比值。

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

##### 3.1.12 圆形性

用目标区域的所有边界点定义的特征向量。

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

$\mu_R$  为区域重心到边界点的平均距离,  $\sigma_R$  为从区域重心到边界点的距离的平均方差。

### 3.1.13 偏心率

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。

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

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

### 3.1.14 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。

### 3.1.15 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。

### 3.1.16 球状性

内切圆的直径与外接圆的直径之比。

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

### 3.1.17 伸长度

周长与目标区域最小外接矩形面积之比。

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

### 3.1.18 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。

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

## 3.2 简单描述子

### 3.2.1 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。

### 3.2.2 面积

描述区域大小的特征。对区域内总像素点的统计。

### 3.2.3 宽度和高度

最小外接矩形的宽度和高度

### 3.2.4 边界的曲率

## 3.3 纹理表示

### 3.3.1 统计方法

### 3.3.2 结构性方法

### 3.3.3 频谱方法

- 基于灰度共生矩阵的方法
- 灰度-梯度共生矩阵分析法
- 灰度行程长度统计法



- 小波分析法
- 基于 Gabor 小波变换的纹理分析法