

数学形态特征应用于昆虫自动鉴别的研究

赵汗青¹ 沈佐锐^{1*} 于新文²

(1 中国农业大学植物保护学院, 北京 100094)

(2 西南林学院动物保护与利用系, 昆明 650224)

摘要 用虫体面积、周长等 11 项数学形态特征对 40 种昆虫实现自动鉴别, 得出了各项数学特征的权重。在昆虫的自动鉴别中, 11 项特征所起的作用大小为: 面积> 偏心率> 形状参数> 周长> 纵轴长> 孔洞数> 横轴长、似圆度> 叶状性> 圆形性> 球状性。以面积、周长等数学形态特征为分类依据, 对隶属 8 目 25 科的 40 种昆虫进行了二义式分类, 并以此为鉴别机理, 实现了对昆虫自动识别软件 BugV isux 的升级, 使其能够自动鉴别的昆虫种类由 3 种增加到 40 种, 准确率达到 97. 5%。

关键词 数学形态特征; 自动鉴别; 分类机理

中图分类号 Q 964

On Computer-aided Insect Identification Through Math-Morphology Features

Zhao Hanqing¹ Shen Zuorui^{1*} Yu Xinwen²

(1 College of Plant Protection, China Agricultural University, Beijing 100094, China)

(2 Department of Animal Protection and Utilization, Southwest Forestry College, Kunming 650224, China)

Abstract The paper presents the result of computer-aided identification of 40 species of insects by using their 11 mathematical morphological features (MMF). The weight and importance of each MMF is evaluated on the basis of their reference frequency in the process of identification. The result demonstrated that the roles of 11 MMF in computer-aided insect identification (CAII) is, from high to low: area> eccentricity> form factor> perimeter> length> Holumber> (x-length, roundness)> lobation> circularity> sphericity. According to 11 MMF, 40 species of insect which belonging to 8 Orders, 25 Families were identified through dichotomous method. And it also was the identification mechanism of CAII. Finally, the authors put forward an updated version of BugV isux and CAII software that can identify 3 species of insects, and thus enable it to identify as much as 40 species of insects with an accuracy rate as high as 97. 5%.

Key words MMF (mathematical morphological feature); CAII (computer-aided insect identification); identification mechanism

昆虫在地球上的种类和数量都是最大的一个群体。全世界存在的昆虫总数可能超过一千万种, 现已鉴定的近一百万种^[1]。因为昆虫种类繁多, 大多数个体较小, 所以分类比较困难, 现从事昆虫鉴定的人员还仅限于数量极有限的昆虫分类学专家。这对昆虫知识的普及, 害虫的防治以及对昆虫学领域乃至自然界生物规律的探讨都是不利的。

收稿日期: 2001-08-30

高等学校博士点专项科研基金和国家高技术研究发展计划课题资助项目(863-306-ZD05-02-03)。

* 沈佐锐, 教授, 博士生导师, 研究方向为植保信息技术与分子昆虫学技术。联系作者。北京圆明园西路 2 号



计算机视觉技术 (computer vision technology) 的出现和快速发展使得计算机处理和分析图像的能力大大增强, 不少生物学工作者和医学工作者开始尝试利用这种基于图像数字技术的先进手段来帮助完成本学科的研究, 并且取得了较大的进展^[2~8]。由于利用计算机视觉技术可以把昆虫图像的一些原来人们所难以提取的特征提取出来, 如面积、周长、似圆度等。因此能否利用这种技术, 把从昆虫图像中提取出的昆虫数学形态特征作为分类依据, 实现计算机鉴别昆虫, 即实现昆虫的自动鉴别成了一个很有意义的问题。于新文、沈佐锐对昆虫图像的分割及边缘检测算法进行了研究, 利用昆虫的数学形态特征, 实现了对棉铃虫 (*Helicoverpa armigera* (Hübner)) 等 3 种昆虫的自动鉴别, 准确率达到 90% 以上^[9~11], 但是这些研究主要集中在昆虫图像处理技术方面, 没有将昆虫数学形态学特征在分类上的应用进行系统的研究。本文根据对隶属于 8 个目 25 科的 40 种昆虫 (表 1) 的自动鉴别研究, 探讨了根据昆虫的数学形态特征对昆虫实现自动鉴别的可行性, 并对面积、周长等 11 项数学特征的作用作了简要评估。

表 1 40 种昆虫的名称及自动识别结果

昆 虫 种 名	特征 提取 样本 数	识别 检验 样本 数	准确 率/%	昆 虫 种 名	特征 提取 样本 数	识别 检验 样本 数	准确 率/%
长额负蝗 <i>tractanorpha lata</i> (Motschulsky)	24	24	100	尖钩粉蝶 <i>Gonepteryx mahaguru aspasia</i> Ménétriér	21	21	100
黄衣 <i>Pantala flavescens</i> Fabricius	36	30	100	黄环蛱蝶 <i>Nepitisthenis</i> Leech	28	28	100
蚱蝉 <i>Cryptotympana pustulata</i> (Fabricius)	15	15	100	独角仙 <i>Illomyrina dichotoma</i> (L.)	17	17	100
斑衣蜡蝉 <i>Lycomadelicatulawhite</i>	13	13	100	棕色鳃金龟 <i>Holotrichia titanis</i> Reitter	30	23	82.6
碧蜡 <i>Palanena angulosa</i> Motschulsky	35	30	100	华北大黑鳃金龟 <i>Holotrichia oblita</i> (Faldemann)	30	23	100
麻皮蜡 <i>Erthesina fullo</i> (Thunberg)	32	21	100	中华弧绿丽金龟 <i>Popillia quadriguttata</i> Fabricius	30	27	92.6
褐奇缘蜡 <i>Dereperyx fuliginosa</i> (Uhler)	30	30	100	铜绿丽金龟 <i>Anomala corpulenta</i> Motschulsky	30	30	96.7
波原缘蜡 <i>Coreus potanini</i> (Jakovlev)	30	30	100	白星花金龟 <i>Potosia brevitarsis</i> Lewis	30	23	100
黑咬猎蜡 <i>Ectanocoris atrox</i> (Stål)	55	55	98.2	黄斑星天牛 <i>Anoplophora nobilis</i> (Ganglbauer)	30	23	100
槐尺蠖 <i>Samiothisa cinerearia</i> Brønner et Grey	32	20	95	松幽天牛 <i>Senum amurense</i> Eschscholtz	30	23	95.7
小地老虎 <i>Agrotis ypsilon</i> (Rottenberg)	24	12	91.7	榆绿天牛 <i>Chelidonium provosti</i> (Faimaire)	26	26	100
棉铃虫 <i>Helicoverpa armigera</i> (Hübner)	50	24	100	绿翅梨天牛 <i>Saperda viridipennis</i> Gressitt	31	31	100
白点雍夜蛾 <i>Oederonia esox</i> Draudt	38	30	100	漆伪叶甲 <i>Podontia lutea</i> (Olivier)	26	12	100
玉米螟 <i>Ostrinia furnacalis</i> Guenée	40	24	100	多型虎甲红翅亚种 <i>Cicindela hybrida nitida</i> Lichtenstein	40	20	100
雀纹天蛾 <i>Theretra japonica</i> (Dary)	21	21	100	红胸萤 <i>Luciola lateralis</i> Motschulsky	28	28	100
大蚕蛾 <i>Rhodiniinjinkovskiihattoriae</i> Inoue	28	14	92.9	血红沟胸叩甲 <i>Agyrpnus davidi</i> (Faimaire)	34	34	94.1
丁目大蚕蛾 <i>Agliatitula</i>	35	35	97.2	黑带食蚜蝇 <i>Bacchamacula</i> Walker	30	23	100
黄目大蚕蛾 <i>Caligula anna</i> Moore	33	12	100	黄胸木蜂 <i>Xylocopa appendiculata</i> Smith	30	23	100
猫目大蚕蛾 <i>Salassa thepisthenis</i> Leech	30	30	100				
菜粉蝶 <i>Pieris rapae</i> L.	26	26	84.6				
黄粉蝶 <i>Colias hyale</i> L.	30	20	90				
山楂粉蝶 <i>Aporia crataegi dilata</i> Verity	28	28	89.3				
				总 计	1 206	979	97.5

1 材料与方法

1.1 供试材料

隶属于直翅目等8目25科的40种昆虫(表1),每种昆虫50头个体。虫种的主要选择依据为:覆盖一定的目、科,使研究具有一定的代表性;同时在一些形态较典型的科下要有一定数量的种类;昆虫的大小体态分布在较大的范围。

1.2 研究方法

1.2.1 昆虫图像的获取与处理 用Olympus数码相机将固定好的昆虫标本照相,输入计算机。照相时的焦距统一为25cm。对于因标本破损而有残缺的图像,根据昆虫左右对称的原理利用图像处理软件Photoshop及Paintshop进行修补。

1.2.2 图像中昆虫数学形态特征的提取 利用中国农业大学IPMIST实验室用C++ Builder 5.0开发的昆虫图像处理与自动识别系统BugVisux,分别对每种昆虫的各幅重复图像进行处理和昆虫特征提取。提取的特征包括面积、周长、横轴长、纵轴长、形状参数、叶状性、球状性、圆形性、偏心率、似圆度、孔洞数共11项,对这些数学特征的定义和计算方法见沈佐锐等的报道^[2-4]。以重复图像各项提取值的均值作为这种昆虫的特征值。

1.2.3 自动鉴别机理 所谓自动鉴别就是完全依靠计算机根据昆虫图像对昆虫进行种类鉴别。计算机鉴别机理如图1。其中,图中虚线部分是指新的昆虫种类添加步骤。

用计算机根据昆虫的图像进行自动识别,首先要对昆虫的图像进行特征提取,建立不同种昆虫的特征值数据库,然后根据不同昆虫的特征值差别,设立分类阈值,从而将昆虫区分。在对某种昆虫图像进行鉴别时,系统在提取图像的特征值后,根据分类阈值进行判定,看其特征值与数据库中的昆虫特征值阈值是否相符,从而对昆虫进行识别。

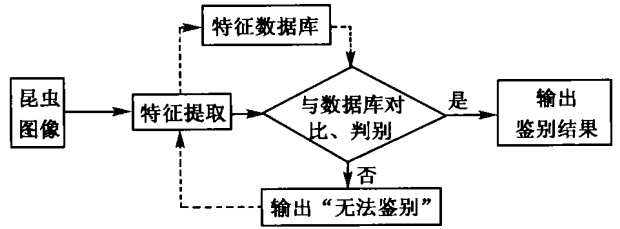


图1 BugVisux 昆虫自动鉴别机理

2 结果与分析

2.1 40种昆虫的11项数学形态特征值(表2)

2.2 自动鉴别的二叉分类图

根据不同昆虫的各项数学特征值,在BugVisux中构建二叉分类图,以实现昆虫的自动鉴别。图2以中华弧绿丽金龟为例示范了程序中的二叉分类。

在对中华弧绿丽金龟的二叉分类中,应用了面积(2次)、偏心率(2次)以及周长、孔洞数各1次。经检验,BugVisux对40种昆虫总的鉴别准确率达到了97.5%(表1)。

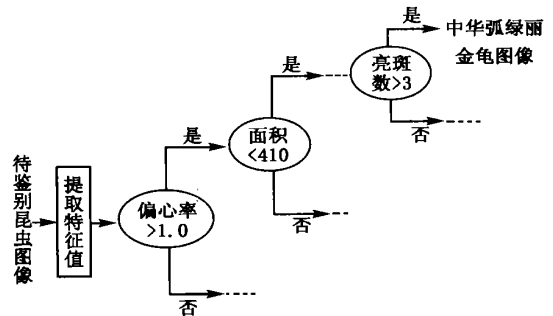


图2 二叉式分类示意图

表 2 40 种昆虫成虫 11 项数学形态特征提取值

昆虫种名	面积	周长	横轴 长	纵轴 长	形状 参数	叶状 性	球状 性	圆形 性	似圆 度	偏心 率	孔洞 数
长额负蝗	5 541.5	579.4	45.3	187.5	5.338	0.296	0.105	3.894	3.347	4.235	2.57
黄衣	5 747.9	614.7	44.6	246.8	5.252	0.128	0.041	5.427	3.709	5.554	2.11
蚱蝉	17 567.0	992.1	107.3	209.5	4.900	0.266	0.189	5.392	1.956	1.958	6.87
斑衣蜡蝉	6 088.8	368.3	67.9	115.5	1.791	0.341	0.316	5.365	1.700	1.716	6.15
碧螬	2 718.0	215.9	48.3	74.6	1.369	0.408	0.481	4.416	1.484	1.547	1.00
麻皮螬	4 725.7	302.7	60.4	104.5	1.459	0.356	0.329	5.352	1.645	1.731	1.00
褐奇缘螬	5 216.5	378.3	62.2	117.6	2.201	0.276	0.221	4.070	1.735	1.905	1.17
波原缘螬	2 085.4	231.0	41.3	75.1	2.262	0.337	0.296	4.017	1.562	1.823	1.07
黑哎猎螬	2 846.3	515.6	72.3	91.1	7.647	0.146	0.132	2.482	1.036	1.640	2.04
槐尺蛾	17 506.0	1 038.0	223.0	72.9	4.411	0.103	0.175	5.813	0.454	0.351	22.81
小地老虎	20 365.0	932.9	273.5	117.8	3.315	0.089	0.156	5.900	0.335	0.431	10.12
棉铃虫	10 394.0	768.8	188.2	86.0	4.686	0.085	0.151	4.909	0.372	0.441	26.17
白点雍夜蛾	13 841.0	717.8	228.4	96.0	2.980	0.076	0.139	5.873	0.338	0.409	1.84
玉米螟	5 196.7	521.3	147.1	38.0	4.345	0.063	0.111	4.072	0.348	0.311	13.08
雀纹天蛾	43 916.0	1 448.0	489.1	228.9	3.912	0.057	0.108	7.355	0.233	0.464	16.20
大蚕蛾	72 155.0	1 685.0	485.1	124.7	3.142	0.048	0.084	7.750	0.442	0.254	5.20
丁目大蚕蛾	82 879.0	1 821.0	503.4	147.8	3.201	0.069	0.087	9.366	0.426	0.279	3.40
黄目大蚕蛾	103 368.0	2 143.0	561.0	156.4	3.545	0.041	0.072	9.243	0.420	0.270	53.17
猫目大蚕蛾	149 687.0	2 486.0	640.9	178.4	3.297	0.050	0.086	9.758	0.464	0.261	10.63
菜粉蝶	37 948.0	1 206.0	298.0	92.7	2.901	0.086	0.144	6.923	0.551	0.316	7.73
黄粉蝶	38 479.0	1 105.0	305.9	105.1	2.538	0.110	0.189	6.886	0.526	0.344	1.22
山楂粉蝶	54 812.0	1 494.0	375.5	117.5	3.278	0.064	0.109	8.275	0.495	0.264	1.39
尖钩粉蝶	60 053.0	1 344.0	377.2	125.4	2.404	0.114	0.198	8.520	0.542	0.321	1.76
黄环蛱蝶	42 783.0	1 317.0	404.3	102.2	3.348	0.070	0.133	6.818	0.334	0.228	43.55
独角仙	31 773.0	1 231.0	169.4	250.4	3.930	0.314	0.324	5.710	1.432	1.493	5.35
棕色鳃金龟	5 331.4	303.0	60.7	113.6	1.373	0.420	0.424	5.554	1.844	1.875	1.63
华北大黑鳃金龟	5 252.0	299.6	60.5	113.0	1.362	0.422	0.421	5.564	1.828	1.870	1.57
中华弧绿丽金龟	4 350.3	266.8	59.3	98.0	1.306	0.436	0.492	5.163	1.571	1.652	6.74
铜绿丽金龟	4 606.4	275.8	61.3	101.0	1.319	0.445	0.514	5.342	1.558	1.648	2.77
白星花金龟	6 091.6	320.1	69.3	114.5	1.341	0.453	0.516	5.517	1.614	1.653	12.04
黄斑星天牛	8 011.5	488.4	60.8	167.3	2.389	0.349	0.222	5.011	2.743	2.750	16.20
松幽天牛	2 402.9	234.4	32.0	91.2	1.859	0.363	0.225	4.139	2.991	2.853	1.37
榆绿天牛	1 871.9	220.0	27.3	88.7	2.067	0.409	0.223	3.682	3.214	3.265	1.00
绿翅契天牛	2 700.8	262.0	32.9	98.9	2.032	0.411	0.249	3.919	3.176	3.013	1.35
漆伪叶甲	2 190.3	206.2	36.6	79.7	1.549	0.369	0.307	4.098	2.120	2.185	1.00
多型虎甲红翅 亚种	2 526.3	258.5	37.9	92.5	2.109	0.299	0.211	3.465	2.263	2.448	3.60
红胸萤	3 664.2	287.3	52.5	79.8	1.813	0.403	0.401	3.939	1.687	1.530	1.71
血红沟胸叩甲	2 179.0	216.0	31.8	86.6	1.699	0.386	0.261	4.084	2.771	2.727	1.65
黑带食蚜蝇	2 933.4	294.0	63.8	81.8	2.376	0.282	0.366	3.711	0.966	1.313	2.23
黄胸木蜂	9 411.3	607.5	172.0	106.2	3.183	0.180	0.282	4.835	0.447	0.642	1.20

图2的二叉式分类图对40种昆虫的分类完全不同于传统分类学的方法,其分支与昆虫所属的分类阶元没有直接关系,而传统分类则是把昆虫按分类阶元由高到低逐步分开。这是数学形态特征和传统分类学的一个显著的区别。当然也可以先对同阶元昆虫的数学形态特征进行归纳分析后按传统分类学的方式进行区分检索,即把40种昆虫先按目、科等阶元归类。但那将比直接按数学形态学本身的差别进行区分繁复得多。

2.3 各项数学形态特征的权重

根据各项特征在40种昆虫的分类中所应用的次数,可以计算这些特征的权重。有了权重,既可以作为判断该项特征在分类中所起作用的大小,还可以作为将来增加鉴别种类时所采用特征的依据。在分类过程中,一个特征每被使用一次,就加1(各特征的初始值均为0),最后得到累加值。各项特征的权重等于其累加值所占所有使用特征累加值之和的百分比。表3是各项特征的权重计算结果。

各项特征权重的大小依次为:面积> 偏心率> 形状参数> 周长> 纵轴长> 孔洞数> 横轴长、似圆度> 叶状性> 圆形性> 球状性。对涉及的40种昆虫来说,面积、偏心率等数学形态特征的权重反映出它们在分类中的相对作用或者说重要性较大;而球状性、圆形性作用较小。

表3 11项数学形态特征的权重

%

特征名称	面积	周长	横轴长	纵轴长	形状参数	叶状性	球状性	圆形性	似圆度	偏心率	孔洞数
权重	21.92	10.27	6.16	8.90	10.96	4.79	1.85	2.74	6.16	15.75	8.22

3 结论与讨论

上述结果表明,不同的数学形态特征在昆虫分类中的作用不同。从11项数学形态特征对40种昆虫自动鉴别的结果来看,昆虫的数学形态特征在昆虫分类上具有很好的应用前景。本研究的昆虫种类只有40种,不能全面代表各项数学形态特征在众多昆虫种类的分类中所起的准确作用。但是可以大体看出这些数学形态特征在昆虫分类中的作用,对于进一步探讨数学形态特征在昆虫分类学中的作用和前景具有重要意义。今后随着研究昆虫种类的增加,各项数学形态特征的权重也必然会有变化,然后逐步稳定。需要关注的问题是各项特征在分类中权重大小的机理和本质,借此从更深层次上探讨昆虫形态进化等问题。

参 考 文 献

- 1 袁锋 昆虫分类学 北京:中国农业出版社,1996
- 2 李月景 图像识别技术及其应用 北京:机械工业出版社,1985
- 3 沈佐锐,于新文 温室白粉虱自动计数研究初报 生态学报,2001,21(1):94~99
- 4 应义斌,景寒松,马俊福 用计算机视觉进行黄花梨果梗识别的新方法 农业工程学报,1998,14(2):221~225
- 5 Mui J K, Fu K S Automated classification of nucleated blood cells using binary tree classifier IEEE Trans On Pattern Analysis and Machine Intelligence, 1980, PAM I-2(5):429~443
- 6 McDonald T, Chen Y R. Application of morphological image processing in agriculture Trans A S A E 1990, 33(4):1345~1351
- 7 Parker J R. Algorithms for image processing and computer vision. New York: Wiley Computer Publishing, 1996
- 8 沈佐锐,于新文 昆虫数学形态学研究及其应用展望 昆虫学报,1998,41(增刊):140~148
- 9 于新文 昆虫图像数字技术的研究开发:[学位论文] 北京:中国农业大学 1999
- 10 于新文,沈佐锐 几种图像分割算法在棉铃虫图像处理中的应用 中国农业大学学报,2001,6(5):69~75