

УДК 004.932

Ю. Б. Камалова, магистрант

Ижевский государственный технический университет имени М. Т. Калашникова

ВЫЧИСЛЕНИЕ СТАТИСТИЧЕСКИХ ПАРАМЕТРОВ В РАСПОЗНАВАНИИ ИЗОБРАЖЕНИЙ ЗЕРЕН ПЫЛЦЫ, ПОЛУЧЕННЫХ С ПОМОЩЬЮ РАСТРОВОГО ЭЛЕКТРОННОГО МИКРОСКОПА

Рассматриваются информативные параметры зерен пыльцы, необходимые для распознавания. Приводятся алгоритм распознавания зерен пыльцы, полученных с помощью растрового электронного микроскопа, и результаты работы компьютерной программы для автоматизированного выделения зерен пыльцы и выявления информативных параметров разных типов зерен пыльцы, результаты статистической обработки данных.

Ключевые слова: распознавание изображений зерен пыльцы, информативные параметры зерен пыльцы, алгоритм распознавания зерен пыльцы

Введение

Процедура распознавания зерен пыльцы используется в палинологии, апиологии (апидологии), апи-мониторинге и аллергодиагностике. Выделение отличительных параметров зерна высококвалифицированным оператором – дорогостоящая процедура, но до сих пор являющаяся наиболее точной и эффективной. Впервые применение автоматической системы классификации зерен пыльцы предлагал ввести Джон Фленли (John R. Flenley) в 1968 году. Однако эта идея была трудноразрешимой в то время из-за технологических ограничений. На сегодняшний день возможность создания данной системы реальна благодаря успехам в разработке компьютерного зрения.

Автоматическое распознавание пыльцы является основой развития полностью автоматизированных систем, которые сочетают отбор проб, осаждения частиц на поверхность, подходящую для оптического анализа, автоматическую подготовку, микроскопические методы визуализации, распознавание и часовой выход численной концентрации в воздухе пыльцы.

Автоматическое распознавание пыльцы развивалось, основываясь на полутоновых инвариантах, которые характеризуют гранулы пыльцы независимо от их положения и ориентации на микроскопическом образце (рис. 1) [1].

На рис. 1, а – выбор нелинейной функции ядра для объединения некоторых соседних пикселей, в этом примере сочетание двух серых значений на расстоянии в 3 пикселя; 1, б – функция ядра оценивается для всех углов, и результат суммируется, чтобы стать инвариантным к вращению объекта; 1, в – этот ряд вращающихся функций ядра оценивается во всех возможных позициях картинки, и результаты суммируются, чтобы стать инвариантными к перемещению объекта. Полученный результат не зависит от угла и положения объекта на картинке.

В работе [2] особые черты пыльцы выделяются из полутоновых изображений в прошедшем свете и флюорисцентной микроскопии: например, в флюорисцентной микроскопии все флюоресцирующие активные молекулы пыльцы сравнимы с небольшим

источником света (рис. 2), и результирующее изображение может быть представлено как измерение локальной флюоресцентной активности.

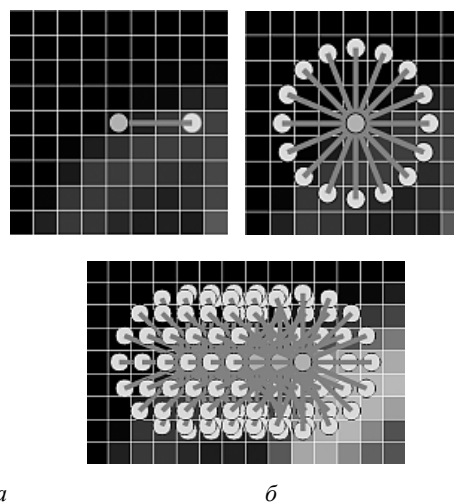


Рис. 1. Расчет двумерных полутоновых инвариантов

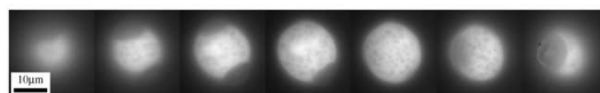


Рис. 2. Флюоресцентное микроскопическое изображение пыльцы березы *Betula*

Таким образом, задача распознавания зерен пыльцы сводится к задаче определения признаков (дескрипторов), по которым будут оцениваться объекты в выборке, т. е. к задаче кластеризации.

Цель работы – выявление и статистическая обработка дескрипторов, характерных для разных типов пыльцевых зерен.

Объект исследования – изображения зерен пыльцы, полученные на растровом электронном микроскопе по разработанной нами технологии [3].

Метод и методика проведения исследования

Выделим признаки, по которым можем разделить пыльцевые зерна вручную, затем математически формализуем и пропишем весовую функцию – какой дескриптор более важен и как соотносится с другими.

Для того чтобы выделить дескрипторы, найдем правило, как их однозначно разделить, с учетом возможного незначительного перекрытия (меньше 25 % от площади), т. е. признаки и логику разделения. Пусть также возможен исход «невозможно определить», но он должен рассматриваться как некоторое вырожденное условие.

Начинать следует с отделения от фона. Фон градиентный, освещение неровное, под углом. Слева чуть ярче, чем справа.

1-й вариант – можно было бы найти среднее фона и сравнить с этим средним каждый пиксель. В этом случае объекты получатся дырчатыми или с оборванными краями.

2-й вариант, более точный, – использование адаптивной бинаризации (на входе имеет объект, шум и фон, на выходе – списки пикселей, относящихся к объекту и фону), при которой автоматически вычисляется порог бинаризации в отношении текущего пикселя. Поиск нескольких локальных порогов осуществляли по методу Оцу (Otsu's method) [5] путем минимизации различных вариантов черных и белых пикселей. Данный метод эффективнее обычной пороговой бинаризации: на рис. 3 приведено изображение пыльцевого зерна (рис. 3, а) после применения бинаризации по методу Оцу (рис. 3, б) и после адаптивной бинаризации с использованием метода Оцу (рис. 3, в).

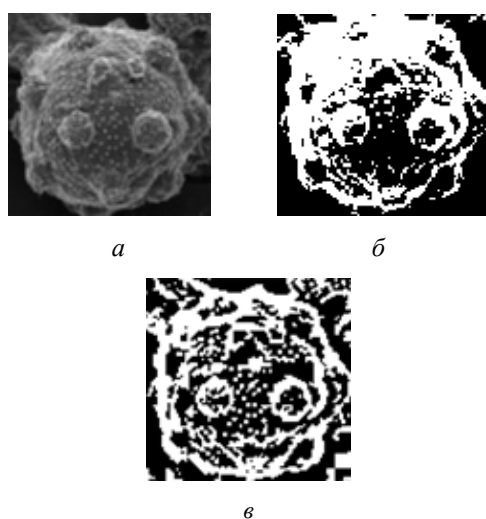


Рис. 3. Пыльцевое зерно (а), пороговая бинаризация (б), бинаризации по методу Оцу (в)

На бинарном изображении объект – это замкнутое множество точек.

Соответственно, второй шаг – определение матрицы этих точек: переход от понятия «изображение» к понятию «возможно, объект».

После определения замкнутости пространства целесообразно искать площадь при известном диапазоне значений для типов или шума на изображении. Тогда все, что выше диапазона – это перекрывающиеся объекты, а все, что ниже – ошибка выделения или часть объекта.

Теперь мы перешли к понятию «объект». Идем дальше к понятию «объект определенного типа». Для

этого мы должны проигнорировать неподходящие объекты и работать только с объектами в нашем диапазоне, т. е. отделить объекты от ошибок распознавания.

Делаем это по площади, по отношению площадь/периметр, благодаря которому получим коэффициент, описывающий форму пыльцы.

Таким образом, мы прописываем форму с помощью семейства признаков.

Допустим, у нас осталась только пыльца. С помощью формы мы разделили на семейства. Теперь нужно делить семейства на виды.

Для каждого типа зерен у нас должны быть определенные (с небольшими допусками) фильтры по площади, соотношению площадь к периметру, сигнатуре контура, сигнатуры по распределению интенсивностей. Однако мы не можем однозначно разделить все типы по дескриптору площадь/периметр, так как возникнут проблемы с теми, где периметр вычисляется по приближенной формуле, т. е. с эллипсом – предельные варианты эллипса – круг и линия. Соответственно, соотношение площадь/периметр колеблется от максимума до минимума, т. е. по данному дескриптору можем спутать эллипс со всеми фигурами.

Для того чтобы различить эллипс, можно взять отношение радиусов. Например, квадрат описан окружностью, значит, его радиус – радиус описанной окружности. Это хороший дескриптор, но мы не знаем заранее, к какой фигуре применяем дескриптор. Поэтому можно остановиться на другом дескрипторе, заодно посчитав остроту и количество углов в случайной фигуре – расстояние от центра до каждой точки периметра в полярной системе координат.

По такому дескриптору углы в случайной фигуре и их остроту можно посчитать по дисперсии выборки, рассматривая длину как зависимость от угла.

Пока мы рассматривали простые свойства. Теперь рассмотрим случай перехода от бинарного к серому изображению, но пока не рассматриваем относительное положение пикселей, только интенсивности.

Программно можно сравнить гистограммы интенсивностей путем нормирования суммы интенсивностей на площадь. Однако этот признак сильно зависит от условий съемки.

У нас есть выборка интенсивностей для объектов разных типов. Рассматривая гистограммы интенсивностей как некоторые статистические выборки, можем их сравнить, найдя дисперсию, (D) среднеквадратическое отклонение (σ) и математическое ожидание (M).

Также необходимо привести снимок к эталонному виду путем перешкалирования, чтобы можно было разделять гистограммы по сигнатурам.

Таким образом, можно выделить основные шаги алгоритма распознавания зерен пыльцы: бинаризация, фильтрация по площади, фильтрация по форме, определение характерных сигнатур по интенсивности, фильтрация по сигнатурам интенсивности.

Более детально алгоритм будет выглядеть так:

- 1) шкалирование;
- 2) адаптивная бинаризация;
- 3) выделение замкнутого множества точек;
- 4) определение матрицы найденных точек;
- 5) определение площади: все, что выше диапазона – это перекрывающиеся объекты, а все, что ниже – ошибка выделения или часть;
- 6) фильтрация по площади: если у нас много (субъективный признак) слипшихся объектов и много шума, мы возвращаемся к п. 1 – выделению от фона;
- 7) фильтрация по форме;
- 8) сравнение выборок интенсивностей для каждого объекта и эталона: нахождение D , σ , M ;
- 9) фильтрация по распределению интенсивностей: высокая медиана – объект яркий, высокая дисперсия – объект неоднороден; высокие нижняя и верхняя квантили – объект сильно неоднороден.

Результаты

Отообразим этапы нахождения информативных параметров разных типов зерен пыльцы при автоматизированном анализе изображения пыльцевых зерен полифлорного меда (рис. 4) в программной среде Matlab: шкалирование изображения (рис. 5), бинаризация (рис. 6), нахождение объектов (рис. 7), вычисление глобального порога изображения с использованием метода Оцу $\text{graythresh}(t_img) = 0,4529$, нахождение контура методом Канни (рис. 8), заполнение замкнутых контуров изображения (рис. 9), создание дискообразного структурного элемента $SE = \text{strel}('disk', 2)$ с радиусом $R = 2$, параметром $N = 4$ (когда N больше 0, дискообразный структурный элемент аппроксимируется последовательно N периодически-линейных структурных элементов [4]). Далее – выделение контуров с залитыми областями (рис. 10), выделение пыльцевого зерна – вывод изображения, ограниченного заданным прямоугольником (рис. 11), его яркостной гистограммы, а также рассчитанных параметров – площади (S), периметра (P), отношения площади к периметру (S/P), среднего значения выборки расстояний от центра до каждой точки периметра (R_{mean}), среднеквадратического отклонения выборки расстояний от центра до каждой точки периметра (R_{sco}), среднеквадратического отклонения выборки яркостной гистограммы (SKO), среднего значения выборки яркостной гистограммы (M) (рис. 12).

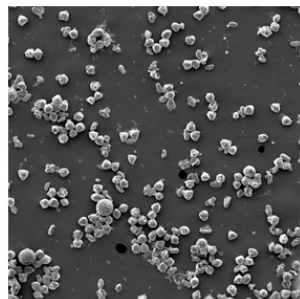
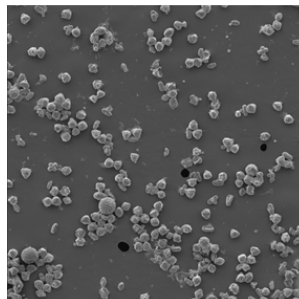


Рис. 4. Исходное изображение

Рис. 5. Шкалированное изображение

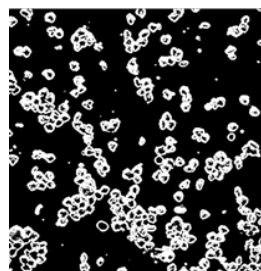


Рис. 6. Бинаризация

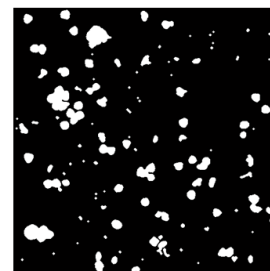


Рис. 7. Нахождение объектов

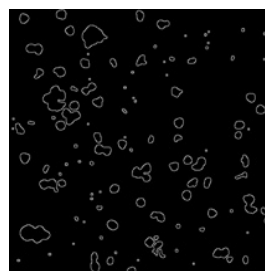


Рис. 8. Нахождение контура

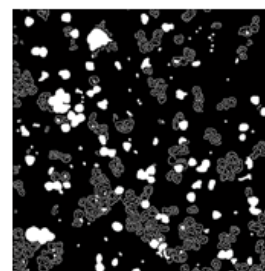


Рис. 9. Заполнение замкнутых контуров изображения

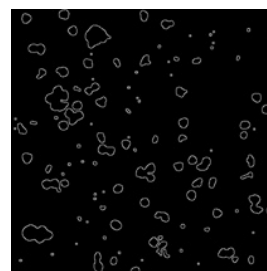


Рис. 10. Выделение контуров с залитыми областями



$S/P=13,6006$
 $R_{mean}=1575,1097$
 $R_{sco}=2,788$
 $SKO=0,34491$
 $M=0,38199$
 $S=2865$
 $P=210,6518$

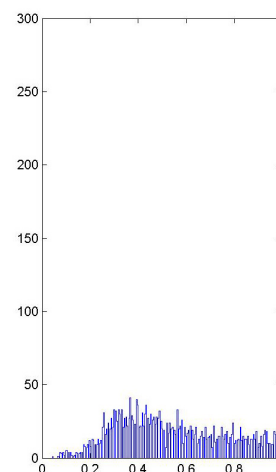


Рис. 11. Выделенное пыльцевое зерно и его информативные параметры

Обработка результатов

Для нахождения диагностического правила применяли критерий Шапиро – Уилка (Shapiro–Wilk test) проверки распределения на нормальность к выборкам порядка 20–25 объектов для каждого из трех основных выделенных типов (рис. 12).



Рис. 12. Типы пыльцевых зерен, присутствующие на рис. 4 – (а) 1-й тип, (б) 2-й тип, (в) 3-й тип

Реально достигнутый уровень значимости получился больше 10 %, это очень хорошо говорит в пользу нормальности: для всех параметров распределения нормальные с вероятностью 70–97 % (рис. 13).

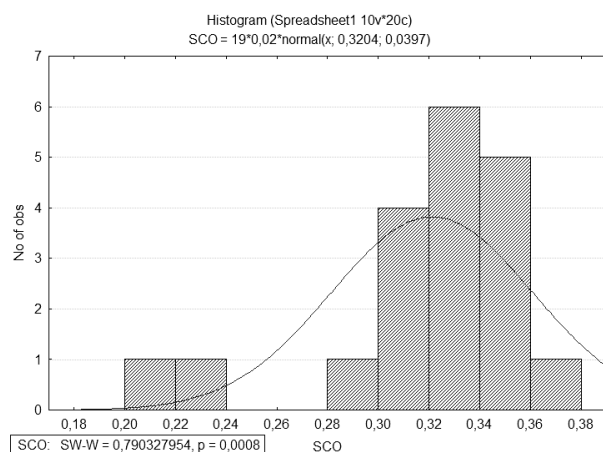


Рис. 13. Нормальное распределение для параметра СКО

Информативные параметры пыльцевых зерен

Тип/Параметр	S/P	Rmean	Rseo	SCO	M	S	P
1-й тип	Сред	14,55	1 031,80	2,31	0,32	0,39	3 238,63
	Min	11,63	-1 329,77	-2,03	0,20	0,26	2 139,91
	Max	17,48	3 393,37	6,65	0,44	0,52	4 337,35
	Сигма	0,98	787,19	1,45	0,04	0,04	366,24
	Pn, %	95	92	75	79	98	94
2-й тип	Сред	10,61	1 064,33	4,65	0,25	0,27	1 876,00
	Min	6,49	-755,73	-16,52	0,13	0,10	266,07
	Max	14,73	2 884,39	25,83	0,36	0,44	3 485,93
	Сигма	1,37	606,69	7,06	0,04	0,06	536,64
	Pn	86	95	55	85	86	71
3-й тип	Сред	4,61	952,82	5,63	0,34	0,36	273,57
	Min	3,41	-1 408,43	-19,06	0,16	0,14	121,29
	Max	5,82	3 314,07	30,31	0,52	0,57	425,85
	Сигма	0,40	787,08	8,23	0,06	0,07	50,76
	Pn	85	83	52	72	78	84

Yu. B. Kamalova, Master's degree student, Kalashnikov Izhevsk State Technical University

Calculation of statistical parameters in recognition of pollen grains, obtained by the scanning electron microscope

Informative parameters of pollen grains, which are necessary for recognition, are considered in the article. Algorithm of pollen grains recognition, obtained by the scanning electron microscope, and results of the work of computer program for automatically pollen grains and informative parameters of different types gripping are presented.

Keywords: pollen grains recognition, pollen grains informative parameters, algorithm of pollen grains recognition

Выражаю благодарность д-ру техн. наук, проф. Г. В. Ломаеву за предоставленные изображения зерен пыльцы, полученные с помощью РЭМ.

Получено: 13.01.14

Для анализа данных использовался программный пакет Statistica 6.0.

Для определения доверительного интервала (или диагностического правила) можно было применить правило «трех сигм»: практически все значения нормально распределенной случайной величины лежат в интервале $[\bar{x} - 3\sigma; \bar{x} + 3\sigma]$, где σ – среднеквадратическое отклонение с достоверностью 99,7 %. Данные были занесены в таблицу.

Выводы

В результате проделанной работы был разработан алгоритм распознавания зерен пыльцы, полученных с помощью РЭМ. Были выведены информативные параметры разных типов пыльцевых зерен, разработана компьютерная программа для автоматизированного выявления информативных параметров разных типов зерен пыльцы и проведена статистическая обработка данных, благодаря которой были найдены доверительные интервалы выявленных информативных параметров, необходимые для формирования фильтров.

Библиографические ссылки

1. Burges, Ch. J. C. A Tutorial on Support Vector Machines for Pattern Recognition. – URL: <http://research.microsoft.com/pubs/67119/svmtutorial.pdf>; <http://www.cs.columbia.edu/~cleslie/cs4761/Burges98.pdf> (дата обращения: 08.05.2014).
2. Ronneberger, O., Schultz, E., Burkhardt, H. Automated Pollen Recognition using 3D Volume Images from Fluorescence Microscopy. – URL: ftp://flores.informatik.uni-freiburg.de/papers/lmb/ro_schu_bu_aero02.pdf; http://lmb.informatik.uni-freiburg.de/Publications/2002/RB02a/2002_Ronneberger_Aerobiologia.pdf; <http://link.springer.com/article/10.1023%2FA%3A1020623724584#page-1> (дата обращения: 08.05.2014).
3. Ломаев Г. В., Камалова Ю. Б., Бекмачев Д. А. Подготовка пыльцевых зерен для анализа с использованием РЭМ // Пчеловодство. – 2012. – № 9. – С. 50–52.
4. Otsu, N. A Threshold Selection Method from Gray-Level Histograms. – URL: <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=04310076> (дата обращения: 08.05.2014).
5. Boomgaard, R. van den, Balen, R. van. Methods for Fast Morphological Image Transforms Using Bitmapped Binary Images // CVGIP: Graphical Models and Image Processing. – May 1992. – Vol. 54. – Iss. 3. – Pp. 252–258.