

АВТОМАТИЗАЦИЯ ПРОЦЕССА РАСПОЗНАВАНИЯ ЧЕРНОМОРСКИХ ДИАТОМОВЫХ ВОДОРΟΣЛЕЙ ПО ИХ ЦИФРОВЫМ ИЗОБРАЖЕНИЯМ

Е. В. Маньковская*, А. М. Лях**

*Морской гидрофизический институт
НАНУ, г. Севастополь, ул. Капитанская, 2
E-mail: ErmakVV@mail.ru

**Институт биологии южных морей НАНУ,
г. Севастополь, пр. Нахимова, 2
E-mail: antonlyakh@yahoo.com

Рассматривается вариант расчета морфологических характеристик створок диатомовых водорослей при помощи специально разработанного программного обеспечения. Основой для расчетов служит контур клетки, по которому определяются его признаки. Полученные вектора признаков контура диатомовых водорослей будут использоваться для решения задачи автоматического распознавания микроводорослей. Также предлагается возможность построения трехмерных геометрических моделей диатомовых, что позволяет наиболее точно вычислять объем и площадь поверхности клеток водорослей.

Введение. Диатомовые – наиболее широко распространенная группа водорослей, их насчитывается около 200 тысяч видов. Диатомовые водоросли – одноклеточные, живущие одиночно или образующие колонии характерной формы (нити, цепочки, звезды, ленты и т. п.). Размеры клеток варьируются от 2 мкм до 2000 мкм.

Наиболее характерная особенность клеток диатомовых – наличие наружного скелета, состоящего в основном из кремния и называемого панцирем. Панцирь состоит из двух заходящих одна на другую половинок (створок) с четко обособленной областью, лежащей между створками и называемой пояском (рис. 1). Верхняя створка (эпитека) и нижняя створка (гипотека) соединяются вместе напоподобие двух половинок чашки

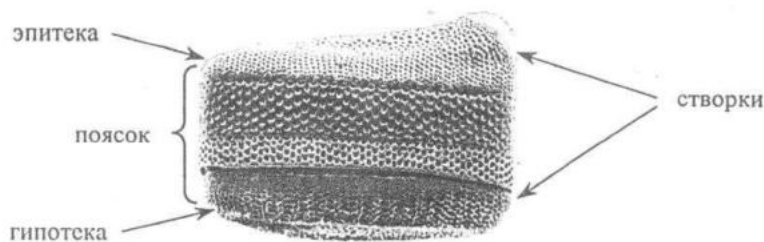


Рис. 1 – Строение клетки диатомовой водоросли

Петри. Каждая створка состоит из плоской или выпуклой пластины, форма которой может быть круглой, эллиптической, треугольной, четырехугольной, многоугольной и т.п. [2].

Для классификации диатомовых, т.е. систематического разделения на роды и виды, используются форма, размер и структура створок. Традиционно распознавание микроводорослей, т.е. определение принадлежности к известным видам, осуществляется в лабораторных условиях экспертом-таксономистом и занимает много времени (от нескольких часов до нескольких дней для одной пробы). В связи с этим актуальным является создание автоматизированной системы распознавания микроводорослей по их цифровым изображениям, которая значительно сократит время обработки пробы.

В последние годы данное направление активно разрабатывается учеными во многих странах. При этом большое внимание уделяется диатомовым. Это обусловлено несколькими причинами. Во-первых, каждый вид диатомовых приспособлен для жизни в довольно узком диапазоне условий среды обитания, что используется для экологического мониторинга. Во-вторых, кремниевый панцирь диатомовых может сохраняться в осадочных породах тысячи и даже миллионы лет, что применяется при реконструкции условий окружающей среды прошлого. В-третьих, диатомовые используются в качестве прикладного аналитического инструмента во многих науках: экологии, палеонтологии, геологии, климатологии, археологии, судебной медицине.

В 1998-2002 гг. в рамках проекта ADIAC (*Automatic Diatom Identification and Classification*) разработана система автоматического распознавания 37 видов диатомовых водорослей по их форме, размерам и внутренним признакам клеток [3]. При тестировании системы на 800 изображениях

правильно было определено 97% видов, что сравнимо с точностью распознавания экспертом. В проекте ADIAC были исследованы и применены различные алгоритмы распознавания диатомовых: на основе дерева решений, по кривизне сегментов контура, по очертаниям контура и полиномам Лежандра, по признакам Габора, по морфологическим признакам, а также смешанные подходы [4].

Однако число видов, распознаваемых данной системой, довольно мало и, следовательно, система является узкоспециализированной, предназначенной для определенных видов водорослей, большинство из которых не обитает в Черном море. Таким образом, на сегодняшний день не существует универсальной системы для автоматического определения диатомовых. Поэтому представляется целесообразным разработать автоматизированную систему для распознавания диатомовых водорослей Черного моря. При этом первоочередной задачей является создание программного продукта, способного определять признаки клеток, необходимые для последующего автоматического распознавания.

Таковыми признаками могут служить признаки контура створки клетки, т. к. диатомовые водоросли под микроскопом чаще всего видны именно со стороны створки. При этом для описания контура можно использовать различные наборы признаков: морфометрические, признаки Фурье, полиномы Лежандра и др. В данной работе различие форм створок проводится по морфологическим признакам, предложенным Shahbazkia H., Candéias T. et al. [6].

Материалы исследования. Процесс автоматического распознавания диатомовых по их цифровым изображениям состоит из трех основных этапов:

1. Выделение контура створки на изображении.
2. Вычисление морфологических признаков контура.
3. Распознавание на основе полученного вектора признаков.

В данной работе рассматривается специально разработанный программный продукт, который обеспечивает реализацию первых двух этапов, а также вычисляет объем и площадь поверхности клеток диатомовых водорослей.

Выделение контура створки на изображении. Выделения объекта на изображении – *сегментация* – необходимый и наиболее сложный вычислительный процесс. Разработано множество различных методов сегментации [5]. Они разделяются на несколько основных групп: *пороговые, краевые, наращивания областей и гибридные*. Задача сегментации диатомовых микроводорослей затрудняется наложением клеток друг на друга, низкой контрастностью контура клетки на изображении, наличием посторонних объектов (частичек детрита). В связи с этим в настоящей работе первоначально рассматривались единичные изображения клеток диатомовых, а контур клетки обозначался множеством точек, которые автоматически объединялись в замкнутую ломаную линию – многоугольник. Полученный оцифрованный контур использовался для последующего вычисления морфологических признаков.

Вычисление морфологических признаков контура. Контур диатомовых дает представление о форме клетки. При цифровой обработке изображений для представления и описания форм объектов используются различные морфологические признаки: *симметричность, выпуклость, округлость, эллиптичность, прямоугольность, вытянутость, размерное соотношение и компактность*. Наборы (вектора) морфологических признаков можно использовать для классификации объектов по их форме.

Выпуклость характеризуется двумя параметрами CA и CP .

CA – это отношение площади выпуклой оболочки ($A_{\text{вып.об.}}$) к площади контура ($A_{\text{конт.}}$):

$$CA = \frac{A_{\text{вып.об.}}}{A_{\text{конт.}}}, CA \in [1; +\infty).$$

CP – это отношение периметра выпуклой оболочки ($P_{\text{вып.об.}}$) к периметру контура ($P_{\text{конт.}}$):

$$CP = \frac{P_{\text{вып.об.}}}{P_{\text{конт.}}}, CP \in (0; 1].$$

Выпуклая оболочка - наименьший выпуклый многоугольник, содержащий данные точки (рис. 2).

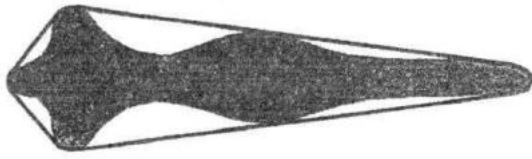


Рис. 2 – Выпуклая оболочка

Округлость (Ec). Мерой округлости служит эксцентриситет (Ec). Фактически эксцентриситет – это отношение меньшей длины оси к большей длине оси наименьшего описанного эллипса (рис. 3). Для его расчета удобно использовать следующую формулу:

$$Ec = \frac{(\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2}{(\mu_{20} + \mu_{02})^2},$$

где μ_{ij} – ij -й центральный момент контура.

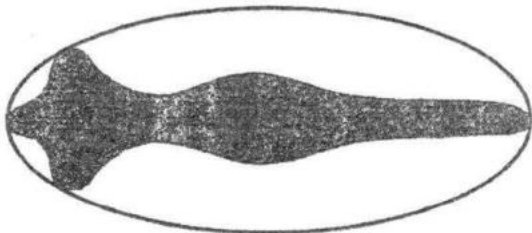


Рис. 3 – Минимальный описанный эллипс

Центральный момент вычисляется следующим образом:

$$\mu_{ij} = \sum_x \sum_y (x - \bar{x})^i (y - \bar{y})^j f(x, y),$$

$$\bar{x} = \frac{m_{10}}{m_{00}}, \quad \bar{y} = \frac{m_{01}}{m_{00}},$$

$$m_{ij} = \sum_x \sum_y x^i y^j f(x, y),$$

где x, y – координаты пикселя контура; \bar{x}, \bar{y} – это средние значение координат пикселей контура; $f(x, y)$ – интенсивность цвета пикселя; m_{ij} – момент изображения.

Эллиптичность (Ell). Данный признак часто используется астрономами для описания формы галактик и определяется следующим выражением:

$$Ell = \frac{(\mu_{20} + \mu_{02}) + \sqrt{(\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2}}{(\mu_{20} + \mu_{02}) - \sqrt{(\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2}}.$$

Прямоугольность (Rec) определяется отношением площади контура ($A_{\text{конт.}}$) к произведению длины и ширины минимального описанного прямоугольника (рис. 4):

$$Rec = \frac{A_{\text{конт.}}}{\text{длина} \cdot \text{ширина}}.$$

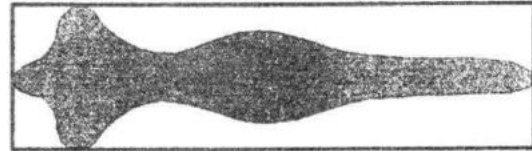


Рис. 4 – Минимальный описанный прямоугольник

Вытянутость (El) рассчитывается как отношение площади контура ($A_{\text{конт.}}$) к его максимальной толщине (Th_{max}):

$$El = \frac{A_{\text{конт.}}}{Th_{\text{max}}}.$$

Th_{max} может быть определена как количество шагов оператора эрозии, которые необходимо применить к изображению до его полного исчезновения.

Размерное соотношение (Rs) – это отношение ширины минимального описанного прямоугольника (рис. 4) к его длине:

$$Rs = \frac{\text{ширина}}{\text{длина}}.$$

Компактность (Com). Значение компактности можно вычислить по площади контура ($A_{\text{конт.}}$) и его периметру ($P_{\text{конт.}}$):

$$Com = \frac{P_{\text{конт.}}^2}{A_{\text{конт.}}}.$$

Самой компактной фигурой является круг, поэтому нормированное значение компактности вычисляется следующим образом:

$$Com = \frac{4\pi \cdot A_{\text{конт.}}}{P_{\text{конт.}}^2}, Com \in [1; +\infty).$$

Вытянутость, размерное соотношение, прямоугольность и компактность инвариантны к перемещению, повороту и масштабированию.

Вычисление объема и площади поверхности диатомовых. При оценках биомассы, продукции, потоков вещества и энергии через поверхность микроводорослей используют объемы и площади поверхности клеток. Традиционный гидробиологический метод расчета этих параметров – расчет при помощи геометрических фигур (метод истинных объемов). Данный метод не всегда позволяет получить достаточно точные результаты, особенно когда клетки имеют сложное морфологическое строение.

Более точный расчет указанных параметров можно произвести, используя *цилиндрические* и *каркасные модели* клеток [1]. Первые представляют собой цилиндры, основания которых совпадают с формой створок клеток, а высоты – с высотами клеток. Их следует использовать для моделирования клеток имеющих прямой панцирь. В случае, когда панцирь диатомовых изогнут или каким-то образом закручен, необходимо использовать *каркасные модели*. Они строятся путем перемещения замкнутого контура вдоль некоторой оси. В качестве оси берутся вертикальные оси клеток, а в качестве замкнутого контура – оцифрованный контур створок диатомовых.

Разработанная программа, используя цилиндрические и каркасные модели, вычисляет объем и площадь поверхности диатомовой водоросли, имеющей заданную форму створки и определенную высоту панциря. Высота панциря клетки задается пользователем.

Заключение. Для ускорения обработки планктонных проб необходимо разработать автоматизированную систему распознавания диатомовых микроводорослей по их цифровым изображениям. В качестве классифицирующих признаков в данной работе предлагается использовать морфологические признаки контура клетки: выпуклость, округлость, эллиптичность, прямоугольность, вытянутость, размерное соотноше-

ние и компактность.

Процедуре распознавания предшествуют два важных этапа обработки изображения – выделение контура клетки и вычисление его морфологических параметров. Для их реализации разработано специальное программное обеспечение, позволяющее на цифровом изображении выделить контур диатомовой водоросли и рассчитать его морфопараметры. Также программа способна выполнять построение геометрических моделей клеток и вычислять их объемы и площади поверхности.

При помощи программы можно получить наборы векторов морфопризнаков черноморских диатомовых, что позволит осуществить их последующее распознавание.

ЛИТЕРАТУРА

1. Лях А. М. Сравнение методов оценки объемов и площадей поверхности диатомовых водорослей // Наукові записки Тернопільського національного педагогічного університету. Сер. Біологія. Спец. випуск "Гідроекологія". – 2005. – № 4 (27). – С. 144-146.
2. Раймонт Дж. Планктон и продуктивность океана: Т. 1. Фитопланктон. Пер. с англ. – М.: Легкая и пищевая промышленность, 1983. – 586 с.
3. du Buf H. et al. Diatom identification: a Double Challenge Called ADIAC // Proc. 10th Int. Conf. on Image Analysis and Processing (Sept. 27-29, Venice). – Venice, 1999. – P. 734-739.
4. du Buf H., Bayer M. M. Automatic Diatom Identification // Series in Machine Perception and Artificial Intelligence. – Singapore: World Scientific Publishing Co., 2002. – Vol. 51. – 316 p.
5. Jalba A. C., Wilkinson M. H., Roerdink J. B. T. M. Automatic segmentation of diatom images for classification // Microscopy Research and Technique, 2004. – Vol. 65. – P. 72-85.
6. Shahbazkia H., Candeias T. et al. Weighted morphometric shape analysis of diatoms // Proc. of the Portuguese Conf. on Pattern Recognition (June 27-28, Aveiro, Portugal). – Aveiro, 2002.