

ОЦЕНИВАНИЕ ЧИСЛА КЛЕТОК НА ИЗОБРАЖЕНИЯХ ЦИТОЛОГИЧЕСКИХ РАСТИТЕЛЬНЫХ ПРЕПАРАТОВ

М.С. Тарков, М.И. Осипов*

Институт физики полупроводников СО РАН им. А.В. Ржанова, г. Новосибирск

E-mail: tarkov@isp.nsc.ru

*Новосибирский государственный университет

E-mail: inq1201@mail.ru

Предложен алгоритм для подсчета вакуолей (растительных клеток) на изображении. Алгоритм позволяет выделить контуры клеток в массивах слипшихся клеток путем разбиения диапазона яркости изображения на слои с последующим применением в каждом слое полутоновой эрозии и дилатации, преобразования «верх шляпы» и пороговой сегментации. При подсчете клеток используется аппроксимация их контуров эллипсами. Алгоритм может быть использован в различных приложениях, например при определении скорости распада клеточных структур по серии изображений.

Ключевые слова:

Оценивание числа клеток, цитологические растительные препараты, анализ изображений.

Key words:

Estimating the number of cells, cytological herbal preparations, image analysis.

Введение

В настоящее время анализ изображений цитологических препаратов осуществляется преимущественно вручную, что сильно замедляет процесс, а плохое качество изображения или наличие на изображении большого числа клеток увеличивает вероятность ошибки при выяснении тех или иных характеристик препарата. В частности, при большом объеме (10^2 – 10^3) выборки подсчет клеток вручную может занять много времени, а также возможен некорректный подсчет.

Использование автоматических методов анализа изображений цитологических препаратов может значительно улучшить качество анализа и ускорить его, однако универсальный алгоритм разработать трудно в силу разнородности изображений. Клеточные структуры могут выглядеть совершенно по-разному, иметь прозрачную, полупрозрачную или непрозрачную структуру, обладать ярко выраженными границами или не обладать таковыми. Поэтому целесообразно разрабатывать алгоритмы под конкретные классы изображений, что повысит их эффективность, но, очевидно, применение их будет возможно только в пределах соответствующих классов.

Разработка алгоритмов для подсчета числа клеток ведется, и для некоторых классов изображений они уже есть [1, 2], но такие алгоритмы обычно основаны на сегментации, то есть целью является сопоставление клетки с конкретным сегментом изображения. Такой подход эффективен в том случае, когда клетки непрозрачны, то есть на изображении цвет такой клетки значительно отличается от цвета фона, что позволяет выяснить, является ли сегмент клеткой. К тому же при наличии шума на изображении, а при съемке цитологических препаратов это не редкость, возможно появление множества мелких сегментов, сходных по цветовым характеристикам с клетками, вследствие чего получится некачественный подсчет. Существуют классы изображений, для которых использование алгоритмов, основанных на сегмента-

ции, неприемлемо. Один из таких классов образуют изображения вакуолей (растительных клеток) [3].

Целью данной работы является разработка алгоритма для подсчета числа вакуолей на изображении. При разработке необходимо учитывать специфику изображений данного класса, поэтому необходимо решить ряд проблем, а именно (рис. 1):

1. Клетки могут слипаться или перекрываться, вследствие чего несколько клеток могут восприниматься как одна, поэтому необходимо решить проблему разделения слипшихся клеток.
2. Клетки полупрозрачны, поэтому цветовые характеристики клеток и фона практически совпадают, что препятствует использованию сегментации изображения классическими способами, так как затруднительно по цветам пикселей некоторого сегмента определить, является ли он клеткой или нет.
3. Необходимо решить проблему отделения клеток от фона, так как в некоторых случаях клетки имеют нечеткую границу.

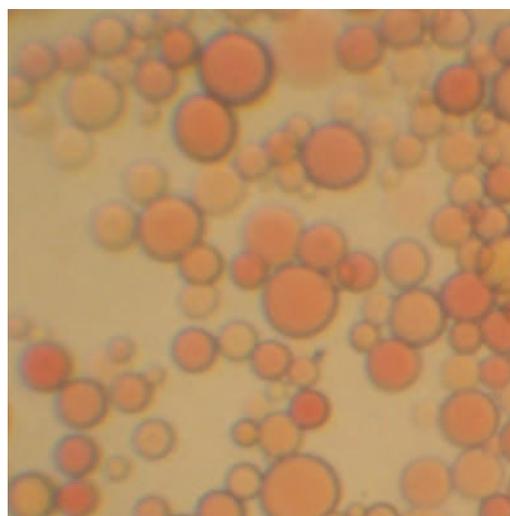


Рис. 1. Фрагмент снимка растительных клеток

Используемый подход

Для решения задачи используется подход, предложенный в [1]. После некоторых модификаций, связанных со спецификой рассматриваемого нами класса изображений, этот подход применим к изображениям растительных клеток (вакуолей). Особенность подхода заключается в том, что после выделения контуров клеток исходное изображение больше не используется. Для дальнейшего анализа используется бинарное изображение контуров. Анализируются геометрические свойства полученных контуров, что позволяет выделять слипшиеся клетки даже в случае, когда граница между ними плохо различима.

В алгоритме можно выделить два этапа:

1. *Выделение и предобработка контуров клеток.* Выделение контуров осуществляется с помощью комбинированного использования морфологических преобразований эрозии и дилатации [4], а также порогового преобразования карты евклидовых расстояний. Контур аппроксимируются многоугольниками: из каждого контура выделяются точки, являющиеся вершинами аппроксимирующего многоугольника. На контурах отыскиваются точки вогнутости. Этими точками контур сегментируется на части.
2. *Выделение клеток с помощью эллипсов.* Строятся эллипсы, аппроксимирующие контуры клеток. Выявляются эллипсы, которые не могут аппроксимировать клетку (слишком узкие или слишком длинные). Объединяются сегменты. Общаются эллипсы, не отобранные на этапе селекции. Они объединяются с другими наиболее подходящими эллипсами. Производится коррекция выделения и подсчет эллипсов.

Алгоритм выделения контуров

Перед тем как приступить непосредственно к нахождению контуров, входное цветное изображе-

ние преобразуется в монохромное. Далее осуществляется следующая последовательность действий:

1. Применение морфологических операций эрозии и дилатации с ядром 7×7 .
2. Выравнивание гистограммы.
3. Сглаживание изображения путем свертки его с гауссовым ядром 3×3 .
4. Вычитание фона.
5. Выделение десяти слоев (интервалов яркости) размером 20 ($[0,19]$, $[20,39]$, ..., $[180, 199]$).
6. Каждый слой подвергается последовательному применению морфологических операций дилатации и эрозии для удаления шумовых выбросов. Также производится пороговое преобразование евклидовой карты расстояний для выделения контуров. Выделение контуров для каждого слоя производится независимо.

Вычитание фона является необходимой процедурой. Её задача – не допустить, чтобы в рассматриваемые слои попали фрагменты фона (рис. 2). Рис. 2 показывает, что после вычитания фона значительная его часть не попадает в рассматриваемые далее слои яркости. Чтобы выделить фон на изображениях данного класса необходимо [3]:

1. Преобразовать исходное изображение в монохромное.
2. Осуществить медианную фильтрацию.
3. Выровнять гистограмму.
4. Взять дополнение.
5. Построить размыкание изображения. В качестве структурообразующего элемента используется круг с радиусом 25 пикселей.
6. Вычесть размыкание из изображения, полученного в ходе шагов 1–4 (преобразование «верх шляпы» [4]).
7. Осуществить пороговое преобразование (порог подобран эмпирически и равен 25) (рис. 3).

Пороговое преобразование карты евклидовых расстояний [5] в данном случае применяется к бинарным изображениям (каждый пиксель является либо черным, либо белым). Карта евклидовых рас-

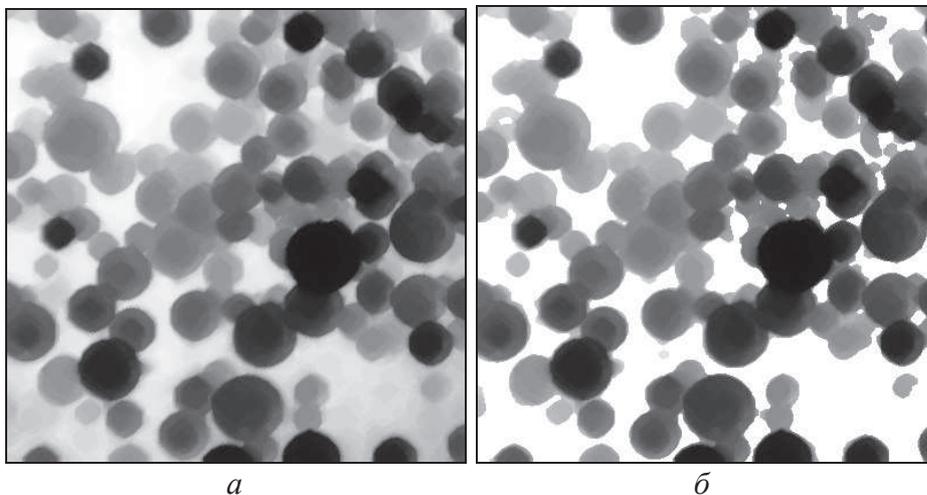


Рис. 2. Изображение рис. 1: а) после действий 1–3; б) после действий 1–4



Рис. 3. Слои изображения рис. 2, б: а) слой [0, 19]; б) слой [20, 39]; в) слой [180, 199]

стояний изображения составляется таким образом: для каждого пикселя вычисляется расстояние до ближайшего черного пикселя (если пиксель черный, оно равно нулю). Пороговое преобразование состоит в изменении цвета пикселя с белого на черный, если его расстояние до ближайшего черного пикселя не попадает в заданный промежуток значений.

Аппроксимация контуров многоугольниками

Аппроксимация контуров многоугольниками основана на двух соображениях. Первое – сгладить контуры, чтобы при поиске точек вогнутости не возникало много ложных точек. Второе – уменьшить число точек контуров. После аппроксимации, как правило, количество точек контура уменьшается. Как следствие, уменьшается количество информации о геометрии контура и сокращается объем последующих вычислений.

Для аппроксимации используется алгоритм Дугласа–Пекера [6]. Алгоритм рекурсивно делит линию. Входом алгоритма служат координаты всех точек контура кроме первой и последней, которые сохраняются неизменными. Далее алгоритм находит точку, наиболее удаленную от прямой, проведенной через первую и последнюю точки контура. Если точка находится на расстоянии меньше, чем ε , то все точки, которые ещё не были отмечены к сохранению, могут быть выброшены из набора, и получившаяся прямая сглаживает кривую с точностью не ниже ε . Если же расстояние больше ε , то алгоритм рекурсивно вызывает себя на наборах точек от начальной до данной и от данной до конечной. Это означает, что данная точка будет отмечена к сохранению. По окончании всех рекурсивных вызовов результирующая ломаная строится только из тех точек, что были отмечены к сохранению.

Поиск точек вогнутости и сегментация контуров

Точки вогнутости – это точки, разделяющие части контура, которые принадлежат различным клеткам. Обозначим $p_c(x_c, y_c)$ текущую точку, а $p_{pre}(x_{pre}, y_{pre})$ и $p_{next}(x_{next}, y_{next})$ предыдущую и следующую точки соответственно. Далее определим величину $concavity(p_c)$ в точке p_c [1]:

$$a(p_{pre}, p_c) = \tan^{-1}((y_{pre} - y_c) / (x_{pre} - x_c));$$

$$a(p_{next}, p_c) = \tan^{-1}((y_{next} - y_c) / (x_{next} - x_c));$$

$$concavity(p_c) = \begin{cases} |a(p_{pre}, p_c) - a(p_{next}, p_c)|, \\ \text{если } |a(p_{pre}, p_c) - a(p_{next}, p_c)| < \pi, \\ \pi - |a(p_{pre}, p_c) - a(p_{next}, p_c)|, \text{ иначе.} \end{cases}$$

Условия, при выполнении которых точка p_c считается точкой вогнутости, выглядят так [1]:

1. $a_1 < concavity(p_c) < a_2$;
2. Отрезок $\overline{p_{pre}p_{next}}$ не должен проходить внутри контура.

Значения a_1 и a_2 задаются пользователем.

Сегментация контуров предусматривает разделение контуров на части найденными точками вогнутости (рис. 4). Отдельные части контуров предположительно являются частями контуров отдельных клеток. По этим фрагментам осуществляется построение эллипсов [1]. Однако сегменты контура в ходе дальнейшей работы алгоритма могут объединяться, поэтому их первоначальное количество в общем случае не равно числу клеток, обнаруженных в области, контур которой рассматривается.

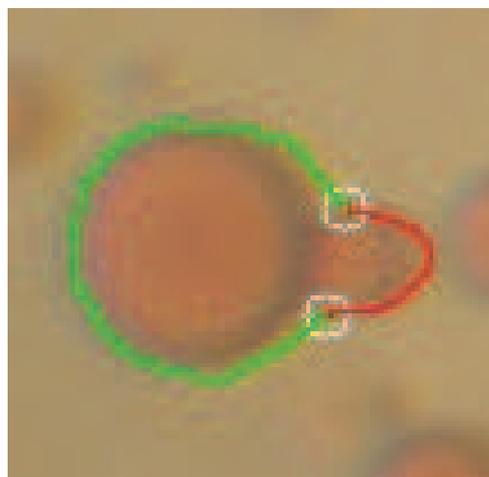


Рис. 4. Пример контура из двух сегментов, разделенных точками вогнутости (точки вогнутости обведены белыми окружностями)

Выделение клеток

Основной целью аппроксимации эллипсами является разделение выделенных областей на отдельные клетки. На данном этапе мы имеем набор контуров, каждый из которых разделен на сегменты точками вогнутости. Изначально для каждого сегмента строится свой эллипс. Далее выполняется анализ геометрических свойств эллипсов и их расположения относительно друг друга. Целью анализа является уточнение результатов, полученных при первоначальном построении.

Построение эллипса производится для каждого фрагмента каждого контура. Сам фрагмент представляет собой набор точек, которые по предположению принадлежат контуру клетки. Однако эти точки образуют только часть реального контура, а построенный эллипс должен дополнить картину (рис. 5).

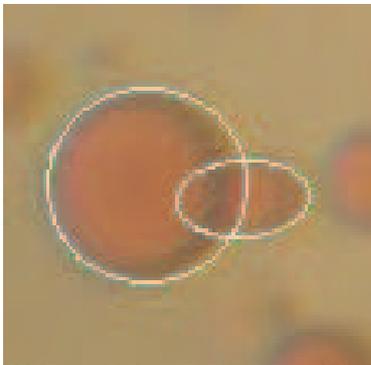


Рис. 5. Пример эллипсов, построенных по точкам сегментов

Рассмотрим уравнение кривой второго порядка:

$$(A, X) = 0,$$

где $A=[a,b,c,d,e,f]$, $X=[x^2,xy,y^2,x,y,1]$. Если для точки (x_i, y_i) сформировать вектор $X_i=[x_i^2, x_i y_i, y_i^2, x_i, y_i, 1]$, то можно определить алгебраическое расстояние (A, X_i) от точки X_i до кривой $(A, X)=0$.

Коэффициенты для искомого эллипса находятся методом наименьших квадратов, то есть решается задача условной оптимизации [7]. Минимизируется сумма квадратов алгебраических расстояний при условии, что получившаяся в результате кривая второго порядка – эллипс:

$$\begin{cases} E(A) = \sum_{i=1}^N (A, X_i)^2 \rightarrow \min_A, \\ (a+c) \cdot \begin{vmatrix} a & b/2 & d/2 \\ b/2 & c & e/2 \\ d/2 & e/2 & f \end{vmatrix} < 0, \\ b^2 - 4ac < 0, \end{cases}$$

где N – количество точек фрагмента контура.

Минимизация проводится по вектору коэффициентов квадратичной формы $A=[a,b,c,d,e,f]$, а ограничения-неравенства гарантируют, что в результате минимизации получится эллипс.

Селекция

Процедура селекции заключается в проверке двух условий для каждого эллипса. Если эти условия выполнены, эллипс участвует в следующем шаге – объединении сегментов. Остальные эллипсы обрабатываются отдельно.

Выбираются эллипсы, которые достаточно точно аппроксимируют контур, и те, которые имеют приемлемую форму для аппроксимации контура клетки, то есть не являются слишком узкими или слишком вытянутыми. Чтобы эллипс участвовал в следующем шаге, необходимо выполнение следующих двух условий [1]:

1. Обозначим сегмент контура $L_i=\{p_1, \dots, p_n\}$, $p_{ij}=(x_{ij}, y_{ij})$, а соответствующий этому сегменту эллипс $F(A_i, X)$. Определим алгебраическое расстояние $dis(A_i, L_i)$ от точек сегмента L_i до соответствующего ему эллипса:

$$dis(A_i, L_i) = \frac{1}{s} \sum_{j=1}^s |(A_i, X_{ij})|.$$

Первое условие: величина $dis(A_i, L_i)$ должна быть меньше заранее выбранного порога $disTh$.

2. Обозначим $MaxA$ и $MinA$ длины большей и малой полуосей эллипса соответственно. Второе условие: отношение $eRatio=MinA/MaxA$ должно быть не меньше заданного порога eTh , что означает близость эллипса к окружности. Порог eTh необходимо выбирать, исходя из формы рассматриваемых клеток.

Объединение сегментов

На данном шаге объединяются сегменты, эллипсы которых построены по фрагментам контуров одной клетки (рис. 6). Берется два фрагмента, строится новый эллипс по точкам из этих фрагментов, проверяется выполнение условий селекции, и в зависимости от ситуации принимается решение: оставить два изначально построенных эллипса или заменить их новым эллипсом, объединив рассматриваемые сегменты.

Возможны следующие ситуации [1]:

1. Явно соприкасающиеся клетки должны быть разделены, т. е. сегменты объединяться не должны. Для идентификации данной ситуации необходимо пользоваться следующими признаками:

- а) Расстояние между центром нового эллипса C_{new} и центрами старых эллипсов C_1 и C_2 должно быть больше, чем величина $dMinTh$, то есть

$$\|C_i - C_{new}\| > dMinTh, i = 1, 2.$$

При соблюдении этого условия эллипсы не объединяются. В качестве $dMinTh$ используется величина, близкая к длине меньшей полуоси наименьшей клетки.

- б) Расстояние между центрами старых эллипсов слишком велико для того, чтобы они были частями одной клетки, а именно

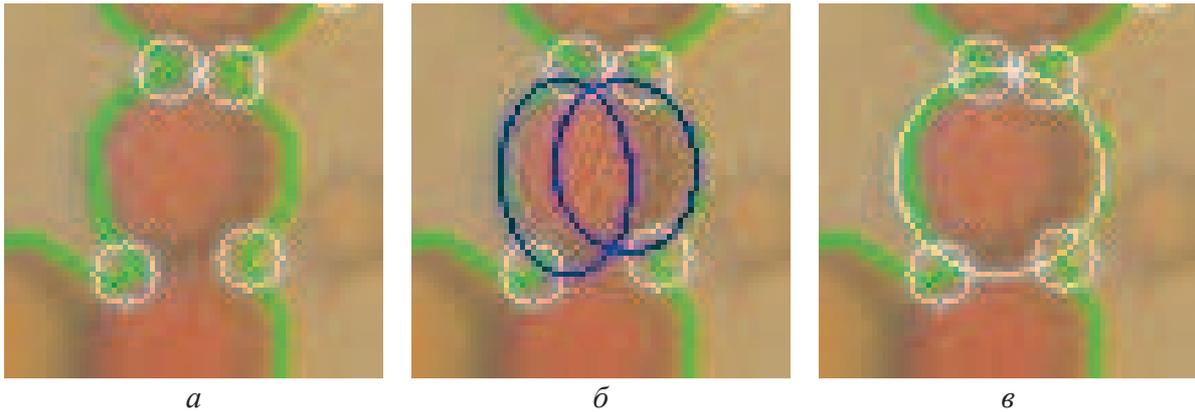


Рис. 6. Пример работы механизма объединения сегментов: а) сегментированный контур; б) эллипсы, построенные первоначально; в) объединение сегментов ввиду того, что они представляют одну клетку

$$\|C_1 - C_2\| > 3 \cdot dMinTh.$$

При выполнении этого условия эллипсы также не объединяются.

2. Два несмежных сегмента должны быть объединены, то есть нужно заменить два старых эллипса новым, построенным по точкам обоих сегментов. Этот случай идентифицируется следующими признаками:

- а) Длины меньших полуосей старых эллипсов меньше величины $dMinTh$, то есть эллипсы слишком малы для аппроксимации контура какой-либо клетки, поэтому их нужно объединить.
- б) Эллипсы, построенные по сегментам контура схожи, то есть длины их больших и малых полуосей соответственно близки:

$$|MaxA_1 - MaxA_2| < dMinTh,$$

$$|MinA_1 - MinA_2| < dMinTh / 2.$$

Такие эллипсы объединяются в один, так как, вероятно, они аппроксимируют одну и ту же клетку.

3. В остальных случаях необходимо пользоваться следующим правилом: если $dis(C_{new}, L_{new}) < dis(C_1, L_1)$ или $dis(C_{new}, L_{new}) < dis(C_2, L_2)$, где $L_{new} = L_1 \cup L_2$, то старые эллипсы заменяются новым.

Обработка эллипсов, не отобранных в ходе селекции

На данном шаге рассматриваются эллипсы, для которых условия селекции не выполнены (необработанные сегменты). Поскольку они являются частями контуров, то соответствующие им сегменты объединяются с наиболее подходящими сегментами эллипсов, прошедших процедуру селекции (т. е. с обработанными сегментами).

Пусть для текущего контура мы имеем m необработанных сегментов L_1, \dots, L_m и n обработанных E_1, \dots, E_n . Берется сегмент $L_i, i=1..m$, строится последовательность P_{ij} точек L_i и $E_j, j=1..n$. По последовательности точек P_{ij} строится эллипс T_{ij} и считается величина $dis(T_{ij}, P_{ij})$. Находится сегмент $E_k, k \in \{1, \dots, n\}$, для которого эта величина наименьшая. Объединение L_i с E_k происходит только в том

случае, если эллипс, построенный по точкам двух сегментов, удовлетворяет условиям селекции, иначе сегменты не объединяются, а сегмент L_i исключается из дальнейшего рассмотрения.

Коррекция выделения клеток и подсчет

Растительные клетки (вакуоли) имеют форму круга. Поэтому для лучшего визуального восприятия результирующее выделение нужно сделать при помощи окружностей, а не эллипсов. По полученному набору эллипсов достаточно просто строится набор окружностей следующим образом: по каждому эллипсу строится окружность с центром, совпадающим с центром эллипса, а радиус окружности равен полусумме полуосей эллипса.

Контуры клеток выделяются из различных слоев яркости. При этом возможны попадания разных частей одной и той же клетки в разные слои. В этом случае произойдет наложение окружностей друг на друга. Данный шаг нужен для отслеживания чрезмерного наложения. Рассматриваются все окружности. Если площадь пересечения некоторого круга с другими кругами составляет более 75 % собственной площади, то данный круг удаляется и в дальнейшем не рассматривается. Результирующее число клеток равно числу оставшихся кругов.

Эксперименты

При разработке алгоритма большое внимание уделялось подбору различных параметров, от которых существенно зависит качество работы алгоритма. Подбор осуществлялся, исходя из геометрии клеточных структур и специфики рассматриваемых изображений. Алгоритм программно реализован в среде Visual Studio C++ с использованием процедур обработки изображений библиотеки OpenCV [8]. В табл. 1 приведены значения параметров алгоритма, при которых получены удовлетворительные результаты. Здесь ε – параметр аппроксимации в алгоритме Дугласа–Пекера, a_1 и a_2 – параметры алгоритма поиска точек вогнутости (в градусах), $disTh$ и $dMinTh$ – пороги расстояний, eTh – порог отношения полуосей эллипса.

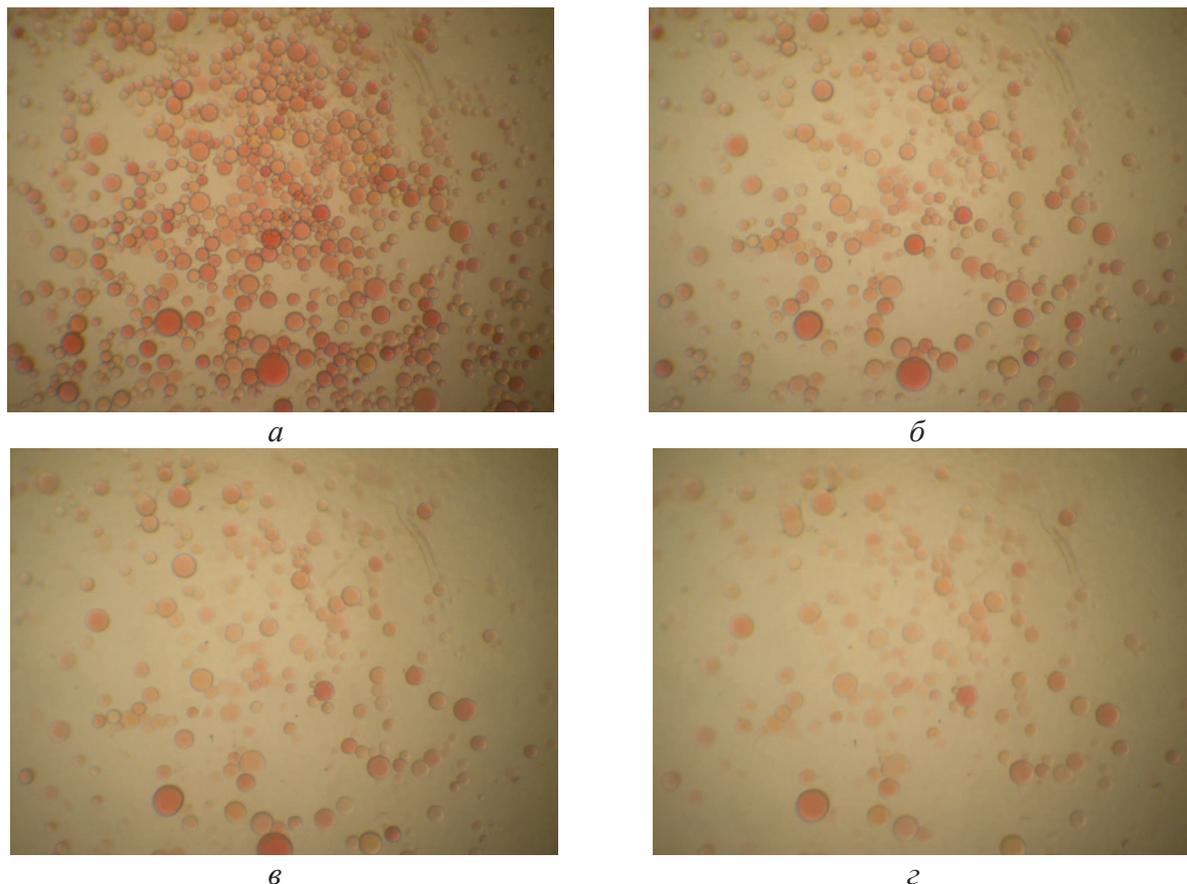


Рис. 7. Тестовые образцы изображений растительных клеток

Таблица 1. Значения параметров алгоритма подсчета клеток

Параметр	ϵ	a_1	a_2	$disTh$	eTh	$dMinTh$
Значение	2	20	160	0,1	0,1	8

В табл. 2 и 3 приведены результаты работы алгоритма для изображений рис. 7, а–г при различных параметрах, используемых в процессе извлечения контуров. В скобках указано отклонение в процен-

тах от результата, полученного подсчетом вручную. На основе экспериментальных данных выбраны оптимальные параметры алгоритма, дающие минимальную по последовательности кадров суммарную ошибку подсчета клеток. При выборе способа разбиения слоев учитывался тот факт, что основные части клеток находятся в диапазоне яркости от 0 до 200. Слой в диапазоне яркости от 201 до 255 содержит фон и мелкие фрагменты клеток.

Таблица 2. Результаты работы алгоритма при различных размерах ядра операторов эрозии и дилатации, применяемых при извлечении контуров

Размер ядра	3×3	5×5	7×7	9×9	Результат подсчета вручную
Изображение (рис. 7)					
а	879 (0,1 %)	817 (7,1 %)	822 (6,5 %)	594 (32,5 %)	880
б	480 (18,2 %)	451 (11,0 %)	391 (3,6 %)	354 (12,8 %)	406
в	370 (37,9 %)	351 (28,1 %)	303 (10,5 %)	291 (6,2 %)	274
г	342 (69,3 %)	324 (60,3 %)	278 (37,6 %)	264 (30,6 %)	202

Таблица 3. Результаты работы алгоритма при различном количестве и толщине слоев яркости, на которые разбивается изображение при извлечении контуров

Кол-во слоев×толщина (пикс.)	2×100	5×40	10×20	12×15	Результат подсчета вручную
Изображение (рис. 7)					
а	508 (42,2 %)	683 (22,3%)	822 (6,5%)	612 (30,4%)	880
б	284 (30,0%)	410 (0,9%)	391 (3,6%)	365 (10,1%)	406
в	367 (33,9%)	322 (17,5%)	303 (10,5%)	287 (4,7%)	274
г	249 (23,2%)	292 (44,5%)	278 (37,6%)	265 (31,1%)	202

В качестве тестовых образцов использовались и изображения размером 1600×1200 пикселей (рис. 7). Алгоритм тестировался на компьютере с процессором Intel Core i3–2310M, 2.1 GHz. Время обработки одного изображения равно 7,3 с. На рис. 8 представлен результат работы алгоритма. Количество обнаруженных клеток – 822.

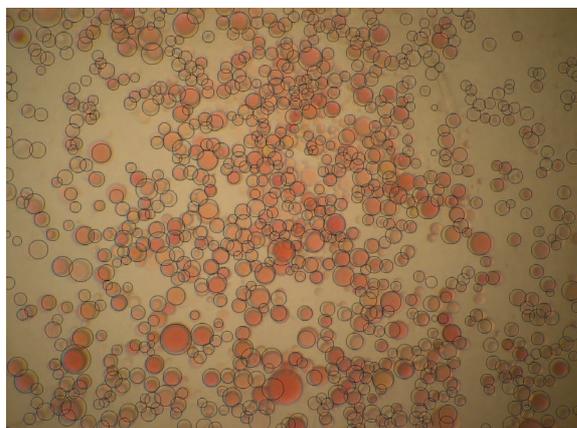


Рис. 8. Пример результата работы алгоритма выделения клеток

СПИСОК ЛИТЕРАТУРЫ

1. Bai X., Sun C., Zhou F. Splitting touching cells based on concave points and ellipse fitting // Pattern Recognition. – 2009. – № 42. – P. 2434–2446.
2. Ковригин А.В. Применение принципов построения систем машинного зрения в задаче анализа изображений клеточных структур // Научный журнал КубГАУ. – 2007. – № 29 (5). URL: <http://ej.kubagro.ru/2007/05/pdf/03.pdf> (дата обращения: 12.03.2012).
3. Тарков М.С., Тихонов Н.В., Половинкин В.Г. Оценивание состава изображений клеточных препаратов для медико-биологических исследований // Известия Томского политехнического университета. – 2012. – Т. 320. – № 5. – С. 130–133.
4. Гонсалес Р., Вудс Р. Цифровая обработка изображений. – М.: Техносфера, 2005. – 1072 с.
5. Danielsson P.-E. Euclidean distance mapping // Computer graphics and image processing. – 1980. – № 14. – P. 227–248.
6. Douglas D., Peucker T. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature // The Canadian Cartographer. – 1973. – V. 10. – № 2. – P. 112–122.
7. Fitzgibbon A., Pilu M., Fisher R.B. Direct Least Square Fitting of Ellipses // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 1999. – V. 21. – № 5. – P. 476–480.
8. Bradski G., Kaehler A. Learning OpenCV. – USA: O'Reilly Media, Inc., 2008. – 576 p.

Поступила 03.03.2013 г.

Заключение

Разработан алгоритм подсчета растительных клеток (вакуолей) на изображениях клеточных препаратов. Алгоритм основан на идее выделения контуров клеток с последующей аппроксимацией контуров эллипсами (окружностями). Известные подходы, основанные на сегментации изображений, неприменимы к подсчету объектов на изображениях растительных клеток, что подтверждено экспериментами. Реализован метод выделения контуров слипшихся клеток. В алгоритме применен ряд дополнительных операций, таких как вычитание фона и коррекция выделения клеток, что позволило значительно повысить качество работы алгоритма.

Экспериментально подобраны параметры, обеспечивающие приемлемое качество работы алгоритма. Алгоритм реализован в среде Visual Studio C++ с использованием процедур обработки изображений библиотеки OpenCV. Алгоритм примечателен тем, что вся работа с исходным изображением заканчивается после выделения контуров, степень четкости границ клеток (если они слипшиеся) не играет роли, так как дальше анализируется только геометрия контуров.

Авторы выражают благодарность В.Н. Нурминскому (СИФИБР СО РАН) за предоставленные изображения клеточных препаратов.