

Алгоритмы анализа изображения для распознавания контура дрожжевых клеток

Перевод с английского Батлук Н.С., автор оригинальной статьи Кварнстром М.
<http://www.opticsinfobase.org/>

Аннотация: Оценка белкового содержания и внутриклеточного расположения движущих сил от флуоресцентной микроскопии изображений представляет высокий интерес в современной клеточной и молекулярной биологии. Для крупномасштабных исследований клеточных популяций и для времен прошлых исследований, таких как количественный анализ, не могло быть эффективного выполнения без какого-либо автоматизированного инструмента анализа изображения. В статье представлены быстрые алгоритмы для автоматического распознавания контура клеток в хорошо освещенных местах изображения, оптимизированный для модели дрожжевого организма (*Saccharomyces cerevisiae*). Контурные клетки могут быть эффективно использованы для количественной оценки параметров морфологии клеток, а также для определения белкового содержания и внутриклеточной локализации от примененной флуоресценции данных.

Введение

Флуоресцентная микроскопия является исключительно мощным методом изучения белковых процессов в живом организме (*in vivo*). Однако из-за качественного и субъективного характера человеческой интерпретации изображений, количественный анализ данных не является простым. Кроме того, для статистических или измерений с высокой пропускной способностью, ручной анализ занимает очень много времени. Автоматизированный анализ изображения открывает для продвинутых исследований разнообразие клеточных популяций, т.е. характеристика фенотип неоднородности, или какой-то одной клетки исследуется в живом организме. Данный тип исследований в последнее время привлекает большое внимание и содержит ключевые идеи, например, в генной экспрессии шума и регулировании клеточного цикла.

В этой работе представлены новые алгоритмы, предназначенные для автоматического анализа светлого поля и флуоресцентных изображений пивных дрожжей (*Saccharomyces cerevisiae*). Дрожжи являются одним из наиболее важных моделей организмов для клеточной и молекулярной биологии, с преимуществами легкой обработки и генетического изменения. В этом контексте он также приносит пользу из библиотеки хромосомно-помеченных Зеленых Флуоресцентных Протеинов (ЗФП) синтеза белков, которые стали общедоступными в 2003 году. Основной принцип нашего подхода заключается в том, чтобы использовать светлое поле изображения для клеточного распознавания, а затем перевести полученные контуры на соответствующее флуоресцентное изображение для анализа белков. Таким образом, наш метод - не полагаться на клеточные мембраны, и поэтому пригоден для исследования на живом организме.

Автоматическое выявление и распознавание клеток в микроскопии изображений имело место и ранее, но некоторые авторы, рассматривающие весь процесс от изображения до определение точных контуров отдельных клеток. Чэнь и др., использовали подход обучающей машины, чтобы автоматически найти приблизительные границы клетки в графической модели представления. Де Карвалью и др. использовали трансформирование водоразделом, применяемой к топологии на основе сочетания серого масштаба изображения и различных иерархических и геометрических свойств клеток. Немист также использовал водоразделительный подход для разделения дрожжевых клеток до локализации небольших флуоресцентных органелл. Рю и Хасби успешно применили передовые деформируемые шаблонные модели для околосферических клеток, представляющих интерес для выявления флуоресцентно помеченных контуров клеток, даже когда границы в части не видны. Недостатком их схемы является ее огромная сложность вычислений и вычислений, осуществленных с помощью марковских цепей алгоритма Монте-Карло. Мардиа и др. использовали аналогичный подход для выявления частично закупоренных объектов. Программное обеспечение для автоматического анализа флуоресцентной микроскопии изображений дрожжей было представлено другими группами. Miroshita и сотрудники разработали программу анализа изображений для морфологических характеристик клеток генно-разрушенных участков. Тем не менее, их алгоритм требует наличия мембраны клеток для окрашивания флуоресцентным красителем. Открытое исходное обеспечение Cell-ID, с другой стороны, использует яркие области изображения для распознавания клеток. Упомянутые фактические методы анализа изображений до сих пор не опубликованы.

В статье представлены алгоритмы для всего процесса анализа изображений, начиная с пороговой обработки и сегментации изображений до уточнения одного контура клетки. Основным новым вкладом в эту работу является метод уточнения контура клетки. Используется форма аналогичной модели для деформируемой шаблонной модели, использованной Rue и Husby, со связями между тремя последовательными точками на контуре клетки. Эта модель, с другой стороны, дает возможность вычисления со значительно меньшей вычислительной сложностью и, следовательно, более высокой скоростью. Рисунок 1 иллюстрирует основные этапы метода. Первые шаги сегментации используют новый метод адаптивной пороговой обработки, для поиска подходящего кандидата на клеточный центр, см. рис. 1. Заключительным этапом является извлечение контура клеток, где более тонкий контур клетки. Окончательные результирующие контуры показаны на рис. 1 (D).

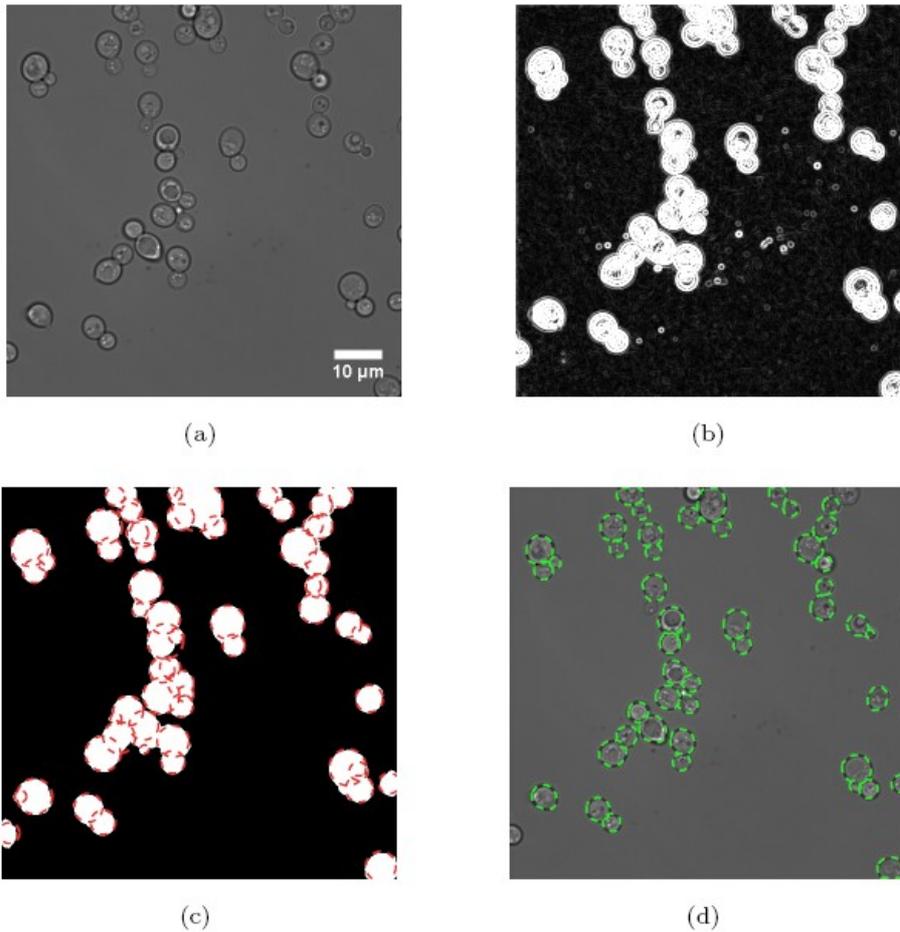


Рис.1. Основные этапы алгоритма

Алгоритм получения контура клетки вычисляет оптимальное решение за полиномиальное время, используя схему динамического программирования. Высокая скорость работы алгоритма делает ее пригодной для исследования, где вовлечены большие наборы данных, таких, как сокращение времени или высокая производительность исследования клеточных популяций. Наш метод приведен для быстроты и, для видов рассмотренных изображений, мы обнаружили, что 96% клеток были удовлетворительно распознаны. Алгоритмы также оказались надежными и адаптивными в том смысле, что они работают без вмешательства человека и в состоянии выявлять клетки с высоким уровнем точности даже в присутствии нескольких соседних клеток, на различных уровнях освещенности или если несколько клеток находятся слегка вне фокуса.

Работа организована следующим образом: первыми представлены экспериментальные и биологические установки, используемые для изготовления изображений в этом документе. Методы анализа изображений описаны в следующем разделе, который является основной частью и направленностью этого документа. Мы потом оцениваем наш метод анализа изображений и приводим пример его использования в биологических приложениях.

2. Методы

2.1. Получение изображения

Изображения были получены с помощью перевернутого эпифлуоресцентного микроскопа (Nikon TE2000E), оснащенного задней подсветкой, электронным умножаемым усилителем КБО камеры (Ixon DV887-DCSBV, Andor технологии). Яркие места изображений, которые используются для метода анализа изображений, были записаны с использованием либо объективов 60X (план Апо НС = 1.40) или 100X (план Апо НС = 1,45). Изображения, использованные в остальной части этого документа, были записаны с 60X объектива.

Для биологического исследования, изображения были записаны в хорошо освещенных местах и флуоресценции в режиме каждые 10 минут в течение 6 часов. Для того, чтобы все ядра были зарегистрированы, клетки были отсняты в трех плоскостях изображения с разделением около 2 мкм. Ртутная лампа была использована для возбуждения флуоресценции и установки фильтра (возбуждение: HQ470/40x выбросов: HQ515/30m, Дихроичные: 505DCLP, Chroma корпорейшн ", США) были использованы для GFP изображений. Флуоресцентные изображения были записаны с интеграцией времени 500 мс с помощью возбуждения интенсивности приблизительно 0,1 W/cm² на образце.

2.2. Деформация и культивация

МСМ4-деформация была получена из GFP-библиотеки. Клетки держали на агар-пластинах с YPD средой, содержащей 1% (W / V) дрожжевого экстракта (Мерк), 2% (W / V) Бакто-пептона (Мерк), и 2% (W / V), глюкозы (Мерк). Для микроскопического исследования клетки были взяты из пластины и помещены в YNB среду, что состоит из 0,17% (W / V), базового дрожжевого азота без аминокислот (Difco), 0,5% (W / V) сульфата аммония (Мерк), 2 % (W / V) глюкозы и аминокислот, дополненных отсутствующим гистидином (BIO101 Инк). РН всех сред был скорректирован до 6,0. Клетки предварительно произведены с использованием орбитальных контейнеров под 30°C. На начальной фазе, 10 мкл суспензии были помещены между двух охватывающих слайдов, разделенных надежной распоркой S24737 из молекулярных проб, созданных для прочного закрытия. Клетки содержались примерно в 20°C в ходе экспериментов.

3. Алгоритм распознавания контура клетки

3.1. Адаптивный метод установления порогов и сегментация

Сегментированное изображение вычисляется через порог на градиентом изображении I_{mag} , который вычисляется из первоначального светлого поля изображения I . Причины для использования I_{mag} носят двоякий характер. Прежде всего, распределение градиента значений пикселей является одномодальным, что делает его легче для нахождения подходящего порога, см. рис. 2.

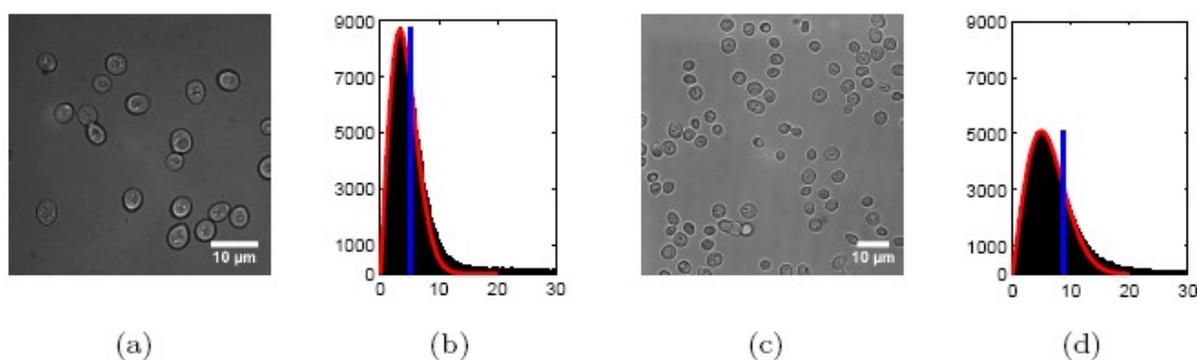


Рис.2. Распределение градиента

Операции на IMAG также имеют возможности обеспечения инвариантной сегментации различий в уровне общей освещенности на изображении. На практике эта инвариантность также проводится для медленно меняющегося уровня освещенности фона. Регионы изображения, где контуры клеток расположены типично, проявляют большую величину градиента. Обратное справедливо для областей фона. Эти отношения между I и I_{mag} можно увидеть, сравнив рис. 1 (а) и (б).

Градиент изображения вычисляется с использованием метод Прюита. Первоначальное светлое поле изображения I фильтруется с помощью двух масок, по одному на каждое направление, где каждая маска усиливает различия в соответствующем направлении. Это приводит к двум новым изображениям, I_x и I_y , каждое из которых является границей пиксельных вариаций в X и Y направлениях, соответственно. После этого, мы положим, что $I_{mag} = (I_x^2 + I_y^2)^{1/2}$ для расположения каждого пикселя. Пиксели с градиентом значения выше определенного порога β теперь будут помечены, как области переднего плана. На рис. 1 (с) мы показываем пороговый градиент изображения после заполнения дыр и запуска морфологических открытий.

Решающая роль в шаговой сегментации заключается в том, чтобы автоматически найти порог β . Это особенно важно для адаптации к уровню освещенности и плотности клеток на изображении. В зависимости от экспериментальной установки, предпочтительный пороговый уровень может быть различным. Методы адаптивного установления порогов, найденные в литературе, например, непараметрический метод Отсу или параметрический метод установки гистограммы по смеси Гаусса, как правило, полагаются на двумодальность гистограмм или на то, что величины пикселей имеют конкретное распределение. Из гистограммы на рис. 2, мы видим, что двумодальность не имеет места в данном случае. Кроме того, распределение для большинства градиентов значений пикселей трудно классифицировать. Хотя это не отображается на рисунке, есть очевидное единообразие большинства значений последовательности вплоть до значений выше 200, где она, наконец, начинает сокращаться.

Максимальная величина этих особых гистограмм является 588 и 770. Метод, представляемый здесь, не делает каких-либо предположений о

распределении градиента значений пикселей больших масштабов, единственно сделанное предположение, что распределение шума в фоновых областях примерно нормальное.

Градиентные значения, ниже медианы, показаны, как вертикальные линии на рис. 2 (б) и (D), которые устанавливаются на Релеевой функции распределения,

$$f_R(x) = \frac{x}{\sigma^2} \exp\left\{-\frac{x^2}{2\sigma^2}\right\}, \quad (1)$$

где σ является параметром, который варьирует, пока наиболее подходящий найден, а X - значение пикселя. Смысл этой процедуры заключается в том, что градиент значения пикселей I_{mag} будут распределены по закону Релея в соответствии с предположениями о том, что такие значения I_x и I_y являются идентичными, независимыми, и, как правило нормально (по закону Гаусса) распределены. Даже если это предположение неверно, мы видим из установленных кривых на рис. 2 (б) и (D), что равенство (1) дает прекрасное описание распределения градиентов значений, ниже медианы.

Параметр σ является хорошей мерой вариации интенсивности пикселей для низкого градиента значений пикселей, то есть для пикселей в фоновой области. Таким образом, пикселями, которые имеют значения градиента гораздо больше, чем σ , нужно пренебречь, как фоном, и вместо этого они будут рассматриваться в качестве фона.

Здесь мы позволим порогу относиться к σ через $\beta = 7.5\sigma$. Это ставит в соответствие объявленным пикселям со значениями градиента более чем на 6 стандартных отклонений больше, чем среднее значение по оценкам распределения фоновых пикселей, пиксели переднего плана. Следует отметить, что не обязательно конкретно использовать медиану, как разделитель для установки Релеевского распределения.

Чтобы получить окончательные сегментированные изображения, дыры (черные области, окруженные белыми областями) в изображении, обработанном пороговым методом, заполнены. Кроме того, в порядке устранения небольших структур из-за шума, мы применяем морфологических открытия с круговой структурой элементов радиусом 4 пикселя. На рис. 3, эти две меры могут быть проиллюстрированы для изображения на рис. 1.

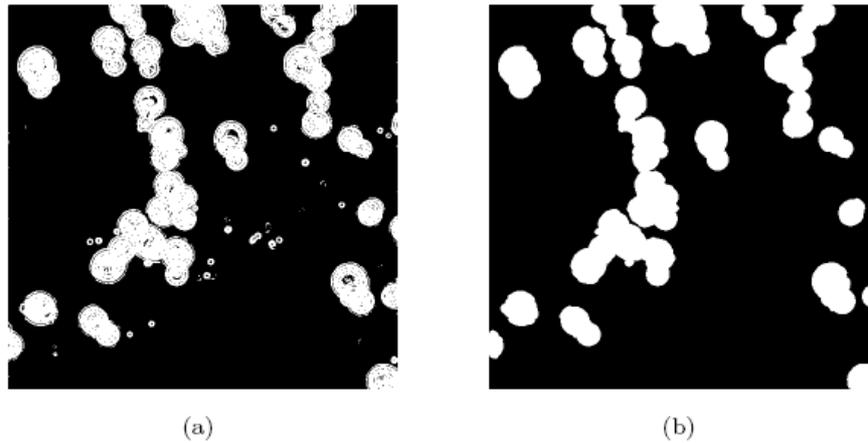


Рис.3. Финальные шаги сегментации

3.2. Поиск предполагаемого центра клетки

Поскольку каждая связная компонента в сегментированном изображении представляет собой предполагаемую совокупность клеток, эти совокупности должны быть разъединены на отдельные клетки. Это достигается с помощью функции поиска по сегменту окружности вдоль внешних контуров объектов в сегментированном изображении, используя вариации кругового преобразования Хоу. Этот принцип показан на рис. 4.

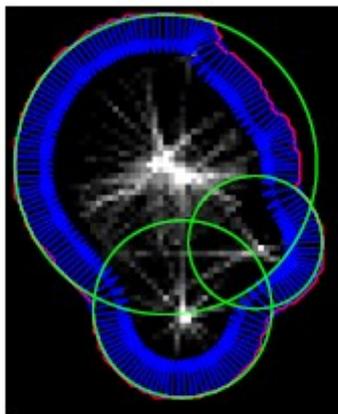


Рис.4. Метода поиска предполагаемого центра клетки

При обходе внешней границы связных областей в сегментированном изображении аппроксимирующая нормаль к границе вычисляется в каждой точке. Вдоль каждой нормали, мы поэтапно идем внутрь и регистрируем расстояния до точки соответствующей границы. Таким образом, накопленная матрица сформирована так, где каждый элемент включает число "попаданий" на нормаль к границе и отслеживает соответствующие расстояния. Когда каждая точка вдоль границы совокупности клеток была пройдена, мы ищем расположение максимума в этой накопленной матрице. Результирующие окружности на рис. 1 (с) отображаются зеленым цветом. Центры кругов

будут выступать в качестве кандидата на центры клеток для получения окончательного контура.

Общий подход к поиску подходящих объектов центров от сегментированных изображений в аналогичных приложениях состоит в том, чтобы искать расположение максимума на расстоянии преобразования сегментированных изображений. Тем не менее, преобразование расстояния работает, только когда клетки немногочисленно сгруппированы, до примерно трех или четырех клеток в каждой совокупности. Хотя несколько большая вычислительная мощность, описанный выше метод работает и для больших совокупностей, применима в более общем плане. Тем не менее, оба подхода зависят от того, что все клетки должны иметь, по крайней мере, часть границы, отнесенной к области фона в сегментированном изображении.

3.3. Получение контура клетки с использованием динамического программирования

Правильные контуры дрожжевых клеток достаточно плавные и, по крайней мере, в определенной степени, выпуклой формы. В порядке нахождения "наиболее подходящего" контура клетки, такая форма ограничения должна быть сбалансирована с данными на изображении, например, градиенты и производные по направлению. В литературе, форму модели, которая адаптирована в соответствии с прилегающими краями, часто называют деформируемым шаблоном. Причиной использования формы модели является то, что она будет руководствоваться решением забрать края, которые лучше всего соответствуют модели, когда информация изображения отсутствует или если она имеет непоследовательные и неоднозначные края. Примерами являются потери контрастности в определенной части границы клетки или вблизи скученных клеток. Что касается информационной части получения клеток на изображении, то оказывается, что темная бахрома дифракции, которая видна вокруг изображения дрожжевых клеток в хорошо освещенных местах, находится примерно на том же расстоянии к истинному контуру клетки, независимо от расстояния до фокальной плоскости. В этой работе мы используем эту особенность, чтобы найти контур клетки, но это можно легко изменить в зависимости от применения.

В принципе, каждая замкнутая последовательность связанных пикселей вокруг предполагаемого центра клетки и, следовательно, пространство возможных контуров огромно. Для того, чтобы урегулировать эту проблему, мы представляем систему отсчета в виде полярных участков по отношению к каждому кандидату на центр. Рисунок 5 (а) выводит производные по направлению I_x и I_y на $M = 32$ луча, выходящих из предполагаемых центров клетки, где каждый луч отобран на $N = 30$ равномерно распределенных радиальных точках. Здесь мы положим подходящий контур точек для тех, которые расположены между позициями с большими производными по направлению в противоположных направлениях. Это согласуется с понятием о том, что истинная граница клеток гранично прилегает к темной «бахроме» вокруг клетки. Формально $F(R, \phi)$ обозначает точку контура критерия

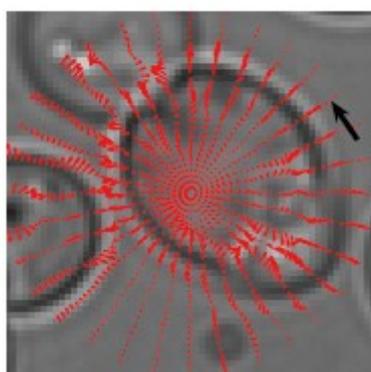
функции за точку на расстоянии R и под углом ϕ будет точка контура. Угол ϕ измеряется против часовой стрелки, начиная с черными стрелками на рис. 5 (А).

Пусть

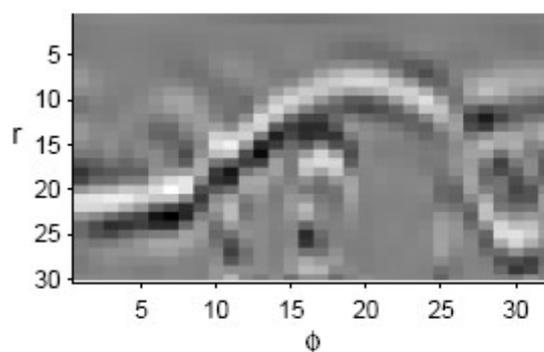
$$f(r_i, \phi_m) = (I_x(r_{i+1}, \phi_m) - I_x(r_{i-1}, \phi_m)) \cos \phi_m + (I_y(r_{i+1}, \phi_m) - I_y(r_{i-1}, \phi_m)) \sin \phi_m \quad (2)$$

являются значениями интерполяции (при необходимости) на расстоянии R и под углом ϕ к предполагаемому центру клетки на производных изображениях I_x и I_y , соответственно. Косинус и синус операции соответствуют проекции вдоль данных лучей.

Это приведет к снижению плотности краев, которые не ориентированы в одном направлении, как луч. На рис. 5 (б), критическая функция $F(R, \phi)$ отображается для предполагаемого центра рис. 5 (А).



(a)



(b)

Рис.5. Процесс оконтуривания клеток

Яркие пиксели на рис. 5 (б) представляют собой высокие значения F и, следовательно, "хорошие" позиции. Идея состоит в настоящее время в выборе пути слева направо в этом полярном участке, обходя каждый столбец один раз. Чтобы найти оптимальный путь, мы используем динамическое программирование. Если просто брать максимум точки R на каждый луч в направлении ϕ , то в общем, не произведет непрерывный и закрытый контур.

Когда выбор пути слева направо в полярном участке, ограничения на радиальных переходах будут влиять на различные формы окончательного контура. Иными словами, ограничения будут определять деформируемый шаблон модели. При использовании динамического программирования, является единственно возможным поставить местные ограничения на этих переходах, они не могут зависеть например, от всего пути до настоящего угла ϕ . Глобальная форма ограничений, как, например, выпуклости, является поэтому не способной применяться напрямую. Невозможно даже гарантировать, что окончательный контур будет замкнут, т.е. то, что она начинается и заканчивается в одной и той же радиальной позиции. В особых

случаях применения выпуклости и замкнутости, есть все-таки методы избежания этой проблемы.

Используется метод, который предусматривает штрафные переходы в полярном участке, что соответствуют правой стороне в координатах исходного изображения, которыми мы окружаем предполагаемые центры против часовой стрелки. Если мы, в дополнение можем гарантировать, что результирующий контур будет замкнут, эти локальные условия выпуклости будут производить глобально выпуклый контур, т. е. контур форм, которая геометрически выпуклая. Замкнутости в свою очередь могут быть исполнены, косвенно запустив алгоритм отдельно для каждой из N возможных отправных точек. Тем не менее, мы используем эвристический алгоритм, который запускает схему динамического программирования на протяжении трех кругов и использует среднюю часть окончательного пути как решение. Подробности относительно схемы динамического программирования и переходного условия приведены в Приложении.

На рис. 6, продемонстрировано получение контура методом экстракции. В верхнем примере мы использовали простые правила перехода, что только позволяет прямые или диагональные последовательные переходы между столбцами в полярных участка. В нижнем примере мы использовали условие о том, что штрафные переходы, которые соответствуют созданию правый стороны, также названы условиями местных выпуклостей. С последним условием, алгоритму удастся получить контур даже несмотря на то, что предполагаемый центр в значительной степени вне центра. Это типичное поведение алгоритма. Используя правило местных выпуклостей, деформируемый шаблона является достаточно гибким, чтобы справиться с весьма нецентральным расположением предполагаемых центров клеток, пока еще соблюдается выпуклость в результирующем контуре. Следовательно, помимо того, что теоретически оно более привлекательно, чем правило перехода, продемонстрированное в верхнем примере рис. 6, которое обычно используется в получении контура с использованием динамического программирования, правило локальных выпуклостей также имеет возросшую устойчивость к ложному позиционированию мнимых центров клеток. В обоих примерах мы использовали три круговых эвристических метода для получения замкнутого контура, где средняя часть (между двумя вертикальными красными линиями) используется в качестве оптимального пути и отображается в соответствующем изображении на правой стороне. Заметим, что только использование одного круга в полярном участке Рис. 6, верхнего примера не приведет к замкнутому контуру, поскольку оптимальный путь отклоняется в сторону повышения от правильного по мере

того, как оно приближается к границе на правой стороне.

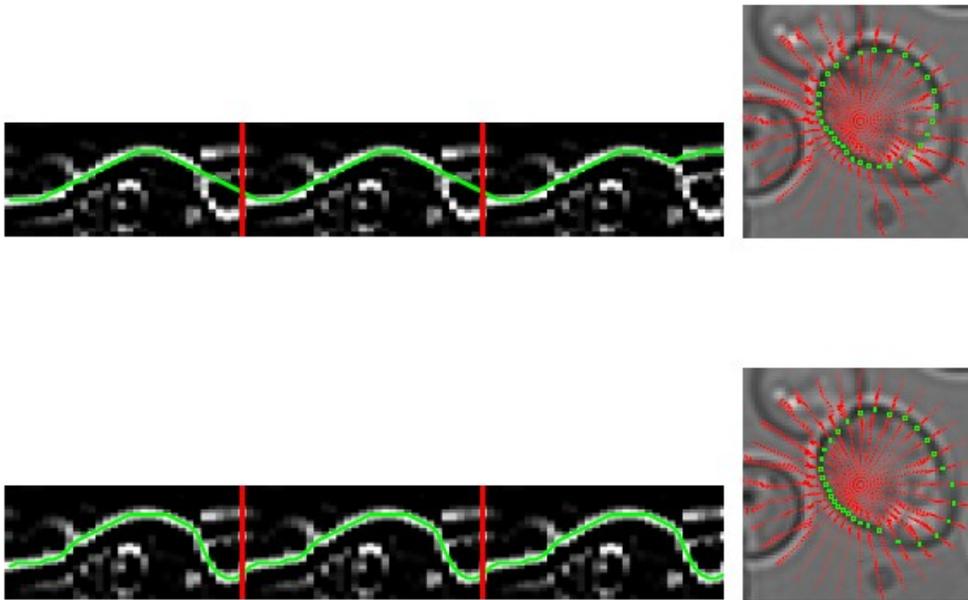


Рис.6. Получение контура методом экстракции

4. Производительность

Алгоритмы распознавания контура клетки были опробованы на 1163 клетках 25 независимых изображений, каждое из которых содержит от 24 до 65 клеток. Типичные изображения характеризовались относительно низкой степенью клеточной кластеризации и плотностью клеток, а также хорошо представлены примерами на рис. 1 и 2. Плотности и кластеризации изображений в исследовании сопоставимы с плотностью и кластеризацией в примере образов литературы. Визуальный осмотр определения контура клеток показали, что 96% были правильно определены, и алгоритмы, было доказано, надежны оба в отношении различий в размерах клеток так же, как и степень внефокусного позиционирования. Оставшиеся 4% клеток были либо упущены, неправильно определены, или были ложными попаданиями. Эти классы показаны на рис. 7 (а) - (D). Клетки упущены в основном в следствие пребывания частью кластера. Даже если они не будут полностью окружены клетками, внешняя граница клеток, расположенных внутри кластеров, может быть слишком короткой для мнимого центра клетки, который будет определен с помощью метода, проиллюстрированного на рис. 4. Это касается пропущенных клеток на рис. 7 (D). Неправильно определенные контуры в основном созданы для зародышевых клеток. Здесь, можно утверждать, что несформированные клетки не должны быть приняты в качестве отдельной клетки вообще. Однако, если они будут приняты, более

узкий диапазон возможных радиальных расстояний в уточненном шаге получения контура должен быть использован для зародышевых клеток. В любом случае, последний шаг мог бы быть довольно легко включен, отказавшись от необоснованных контуров, например, те, которые слишком малы или имеют слишком много совпадений с другими клетками. Таким образом, число ошибочных и ложных контурных попаданий можно было бы сократить.

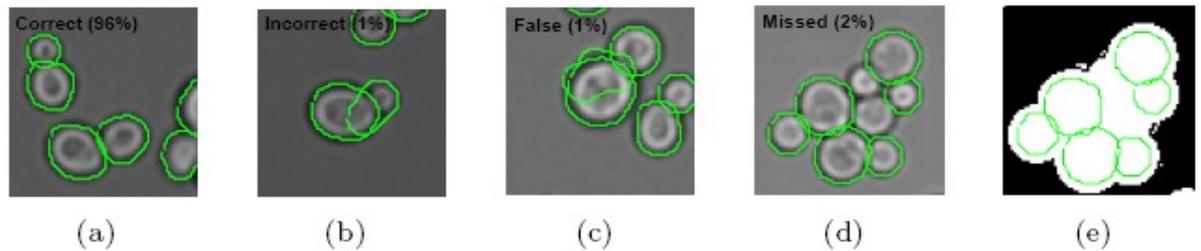


Рис.7. Преобразования процесса распознавания клеток

Для оценки скорости алгоритмов, CPU-время было измерено алгоритмами, работающими на ПК с Intel Duo-Core процессором на 1.66GHz с 2GB RAM, в Windows XP и Matlab R2006a. Критическое время алгоритмов было реализовано в С и связано с использованием Matlab MEX-утилитами. CPU-время было измерено по альтернативному использованию "TIC" и "ТОК" команд в Matlab. Для трех примеров изображений, содержащих 49, 96 и 173 клетки, CPU-время на распознавание клетки на яркой области изображения к области уточненных контуров клеток было измерено примерно в 1,4, 2,0 и 2,8 секунды соответственно. Часть полученного выпуклого контура, используя три эвристических круговых метод, занимает около 10 мс на расчет мнимого центра для $M = 32$ и $N = 30$.