

УДК 004.932.2

Н. Н. Бич, Б. А. Ассанович

Гродненский государственный университет имени Янки Купалы

**РЕАЛИЗАЦИЯ РАСПОЗНАВАНИЯ И ОЦЕНКИ ПАРАМЕТРОВ
БИОМЕДИЦИНСКИХ ИЗОБРАЖЕНИЙ КЛЕТОЧНОГО УРОВНЯ**

Описана методика для экспериментального изучения биомедицинских цифровых изображений, детектирования и анализа конкретных его элементов (областей), построения их геометрических моделей, извлечения их параметров и статистической оценки появляющейся методической погрешности, а также создано специализированное программное обеспечение на основе программных средств широкого применения (библиотеки OpenCV, среды программирования Delphi 7, языка программирования Pascal и системы управления базами данных MySQL).

Ключевые слова: цифровое биомедицинское изображение, детектирование объекта, распознавание объекта, функция-классификатор, библиотека компьютерного зрения.

N. N. Bich, B. A. Assanovich

Yanka Kupala State University of Grodno

**IMPLEMENTATION OF RECOGNITION AND ESTIMATION
OF PARAMETERS BIOMEDICAL IMAGES THE CELLULAR LEVEL**

This paper describes the technique for experimental study of biomedical digital images, detection and analysis of its specific elements (regions), construction of their geometric models, extraction of their parameters and statistical estimation of the emerging methodical error is described, as well as the creation of specialized software based on wide application software (OpenCV libraries, Delphi 7 programming environment, Pascal programming language and MySQL database management system) is described.

Key words: digital biomedical image, detection of an object, recognition of an object, function qualifier, library of computer sight.

Введение. Современные методы медицинской диагностики и биомедицинских исследований, основанные на использовании таких электронно-вычислительных средств, как микроскопы, рентгено- и термографические аппараты, томографы и многое другое, дают возможность получить полную информацию о строении и изменении различных органов и тканей организма человека.

Одной из приоритетных целей применения данных методов является обнаружение (детектирование) и распознавание искомого объекта на цифровом изображении, а в дальнейшем его регистрация, цифровая обработка и анализ. В настоящее время проблеме распознавания отдельной области или конкретного элемента изображения посвящено множество работ. Однако в целом данная задача еще далека от разрешения.

Основные трудности состоят в том, что структура исследуемых объектов на медицинских изображениях является сложной, неоднородной и требует от специалистов владения определенным практическим опытом, а также знаниями в области общепатологических процессов, позволяющих поставить максимально точный диагноз. Использование вычислительной техники и математических методов в этом направлении дает возможность повысить точность результатов исследований, а также провести аналогию между свойствами изображений и

результатами биомедицинских экспериментов. Применение систем управления базами данных для хранения изображений позволяет их сортировать, обмениваться изображениями и результатами их изучения с другими исследователями, используя для этого современные средства компьютерных телекоммуникаций.

В данной работе изучены особенности методов детектирования и распознавания конкретных объектов изображения. Спроектирована и разработана экспериментальная модель для хранения, обработки, детектирования, распознавания, анализа элементов (областей) цифрового снимка биомедицинских изображений в виде информационной системы, использующей данные и программное обеспечение (библиотека компьютерного зрения OpenCV, среда программирования Delphi 7, язык программирования Pascal, система управления базами данных MySQL).

Основная часть. В качестве тестируемого объекта выбран цифровой снимок образца семенного канальца.

Постановка задачи состоит в детектировании объекта (биологической клетки) на цифровом снимке, нахождении его контура, аппроксимации найденного контура геометрическим примитивом, численной обработке выделенных параметров. Метод решения задачи состоит из нескольких этапов.

Первый этап заключается в предварительной подготовке цифрового изображения к анализу и включает следующие шаги: фильтрацию изображения по цвету – настройку цветовой маски; проведение преобразований, связанных с приведением масштабов всех полученных цифровых изображений тестируемого образца к одинаковому масштабу. Процедуры выполняются автоматически отдельным блоком программного приложения. Второй этап заключается в выделении или, другими словами, детектировании на цифровом изображении интересующего исследователя элемента, подразумевает нахождение и прорисовку контуров найденного объекта [1]. Третий этап предусматривает проведение измерительной и статистической обработки результатов [2]. Четвертый – сохранение образцов и полученных результатов в базу данных. Подробно опишем второй этап.

Детектирование объекта – это обнаружение объекта на цифровом снимке с последующим выделением области его расположения. Обнаружение (детектирование) элемента или области на изображении ставит перед собой цель его локализации, определения и выделения в цифровом потоке. Информация о присутствии одного либо нескольких объектов на изображении полезна в системах автоматического учета числа объектов, например при подсчете здоровых и больных клеток в разрезе ткани. Для дальнейшей оценки некоторых количественных характеристик (диаметра) клетки путем подсчета клеток по отдельным полям зрения возникает систематическая погрешность, связанная с учетом пограничных клеток, частично попавших в кадр, поэтому все пограничные клетки следует исключить.

В общем виде алгоритм обнаружения объекта заключается в том, чтобы по данным цифрового снимка на входе системы определить, имеется ли на данном изображении конкретный элемент, если да, то указать, где он находится и каков его размер [3]. Основным требованием, которое предъявляется к подобному классу алгоритмов, является высокое качество изображения, напрямую зависящее от оптической системы ввода. Медицинские электронно-вычислительные средства, используемые для получения изображений исследуемой области, работают в двухмерном режиме. В этом случае распознавание происходит на основе плоского изображения. Двухмерные аппараты весьма чувствительны к уровню освещенности помещения, от этого параметра в значимой мере зависит качество конечного изображения (при плохом свете изображение будет трудноразличимым). Безусловно, конечный результат зави-

сит не только от вышеуказанных характеристик. Большое значение имеют также фокусное расстояние, разрешение, углы обзора и множество других параметров. В частности, использование того или иного микроскопа, а также цифровой видеокамеры сказывается на точности анализа изображения. Поэтому, проводя морфологические исследования посредством компьютерной цифровой микроскопии, необходимо учитывать равномерность освещения, высокие разрешающие способности используемой техники и контрастность изображения.

Для автоматизации процесса распознавания и обработки цифровых изображений используется библиотека алгоритмов компьютерного зрения и обработки изображений с открытым исходным кодом OpenCV [4]. Задача нахождения графических примитивов решается во многих системах компьютерного зрения, в частности, MatLAB Image Processing Toolbox, Juno, Cognitive Passport, Pisoft Image Framework и др. Но перечисленные программные пакеты в отличие от OpenCV дают возможность обработки цифровой картинки удаленно от разрабатываемой под конкретную задачу программы. OpenCV – хранилище алгоритмов из разных областей (вычислительная геометрия, обработка сигнала, машинное обучение). Все алгоритмы делятся на две категории: методы моделирования изображения и методы эмпирического детектирования. Каждый из алгоритмов можно использовать в приложении, написанном на любом языке программирования, в качестве отдельной функции.

Методы моделирования изображения опираются на инструментарий распознавания образов, в них задача обнаружения объектов рассматривается как частный случай общей задачи распознавания. Работа большинства таких алгоритмов заключается в сканировании входного изображения окном, имеющим определенную форму и различный масштаб, и в определении, к какому классу относится изображение внутри этого окна («искомый элемент» либо «неискомый элемент»). Таким образом, задача выделения объекта на изображениях сводится к построению классификатора, эффективно разделяющего классы «искомого элемента» и «неискомого элемента». На основе обучающей выборки строится модель изображения объекта, и задача обнаружения сводится к проверке входного изображения на удовлетворение полученной модели [5].

Данные методы можно назвать весьма производительными, так как все шаги алгоритмов выделения объектов (группирование векторов, вычисление градиента, поиск контуров) требуют лишь нескольких операций на каждый пиксель изображения. К тому же все современные

компьютеры имеют несколько ядер, и цифровой поток может независимо вычисляться на своем ядре, не мешая остальному анализу. Производительность зависит от размера и количества анализируемых изображений, поэтому точную алгоритмическую сложность подсчитать невозможно.

Существует класс изображений, который трудно анализировать в автоматическом режиме путем выбора систем анализа изображений, использующих методы, основанные на вышеуказанных алгоритмах. К таким изображениям относятся снимки микроструктур. Процесс получения таких снимков является, как правило, непростым и многоэтапным, что может сказываться как негативно, так и положительно на качестве полученного изображения. Поэтому в данной работе использован один из эмпирических методов детектирования, который в качестве признака использует цвет области для поиска объекта по контуру [6].

Контур – это внешние очертания (обвод) предмета или объекта. Контурный анализ – это один из методов описания, хранения, распознавания, сравнения и поиска графических образов или объектов. Во время анализа внутренние точки объекта во внимание не принимаются. Проблемы, связанные с выделением контура на изображениях, возникают из-за одинаковой яркости с фоном, объект может не иметь четкой границы или может быть зашумлен помехами, что приводит к невозможности выделения контура. Перекрытие объектов или их группировка приводит к тому, что контур выделяется неправильно и не соответствует границе объекта. Контурный анализ имеет слабую устойчивость к помехам. Любое пересечение или лишь частичная видимость объекта приводит к ложности детектирования. Простота и быстродействие контурного анализа позволяют вполне успешно применять данный подход (при четко выраженном объекте на контрастном фоне при отсутствии помех).

В данной работе для определения контура выбран метод «жука», заключающийся в попиксельном обходе замкнутого контура [7]. Метод рассматривается на цветовом простран-

стве Saturation, в котором насыщенность цвета зависит от расстояния между центром и краем изображения. Варьируется в пределах 0–100 или 0–1. Чем больше этот параметр, тем «чище» цвет, поэтому этот параметр иногда называют чистотой цвета. А чем ближе этот параметр к нулю, тем ближе цвет к нейтральному серому. Изначально «жук» ставится на пиксель по цвету, близкому к нейтральному, и движется до тех пор, пока не попадет на элемент заданного цвета. Как только он дойдет до пикселя заданного цвета, поворачивает налево и переходит на следующий элемент. Если этот элемент нейтрального цвета, то он поворачивает направо, иначе – снова поворачивает налево. Обход контура заканчивается, как только «жук» снова дойдет до первого посещенного темного пикселя. Во время обхода сохраняются координаты каждого перехода между элементами с разной интенсивностью.

К недостаткам этого метода можно отнести потенциальную возможность ложного обнаружения предметов с цветом, близким к цвету детектируемого элемента, и зависимость от цветового баланса соответствующего оборудования и цвета освещения.

На основании выбранного метода детектирования искомого объекта (клетки семенного канальца) на цифровом изображении выделяется его контур. Посредством координат его точек определяется геометрический шаблон формы контура [8] и происходит его обработка в соответствии с шагами предлагаемой методики, обозначенными в блок-схеме, приведенной на рис. 1.

При этом точки контура не дают прямой информации о его форме, что вынуждает применять вероятностные подходы для получения оценки анализируемых измерений, пример выполнения которых будет приведен ниже. Само изображение и результаты его анализа сохраняются в базе данных.

На рис. 2 отражены технология использования и принцип работы программного приложения. В качестве объекта исследования выбран цифровой снимок образца семенного канальца.

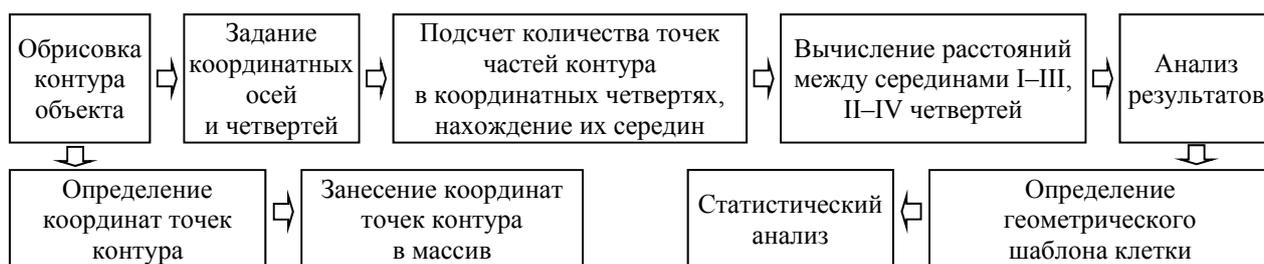


Рис. 1. Схема информационной обработки цифрового изображения объекта поставленной задачи



Рис. 2. Схема принципа работы программы анализа биомедицинских изображений

Все необходимое для нашей разработки имеется в библиотеке открытого компьютерного зрения – OpenCV. Библиотека OpenCV реализует удобные методы для детектирования и манипуляции с контурами изображения.

После соответствующей подготовки к следующему этапу в изображении остались только те пиксели, которые выше или ниже определенного значения. Данная задача решена с помощью функции `cvThreshold()` библиотеки OpenCV. Основная идея алгоритма, заложенного в функцию, – это попадание в конечный массив только тех пикселей, которые выше или ниже заданного порогового значения. Посредством функции `inRange` проделано преобразование цветной картинку в черно-белую маску. В этой маске все пиксели, попадающие в заданный диапазон, становятся белыми. Остальные становятся черными. Каждый пиксель изображения сравнивается с соответствующими значениями изображений `lower` и `upper`. Если значение цвета пикселя больше или равно значению `lower` и меньше, чем `upper`, тогда соответствующее значение устанавливается в `0xff`, иначе в `0`.

Библиотека OpenCV реализует удобный метод «жука» для детектирования и манипуляции с контурами (совокупностями пикселей, составляющих границу объекта) изображения. Для поиска контуров используется функция `cvFindContours()`.

Для выделенной биологической клетки определяется ее геометрическая форма и проводится дальнейший статистический анализ на точность принятого решения. Преобразованное и проанализированное изображение, полученная о нем информация сохраняются отдельным файлом.

Для детектирования формы выделенного на изображении объекта нужно выполнить измерения, геометрические построения и расчеты, приведенные ниже.

Размеры изображения в пикселях, длина s и высота h делятся на два, вследствие чего строится система координат. Определяются точки O_1, O_2, O_3, O_4 , как середины частей контура в

соответствующих четвертях. Строятся отрезки O_1O_3, O_2O_4 (рис. 3).

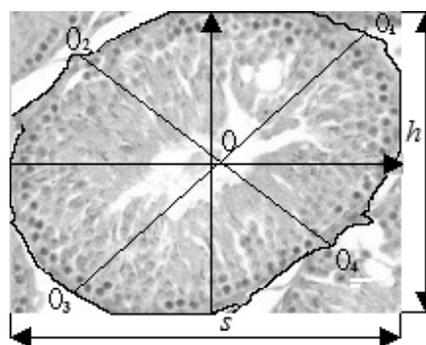


Рис. 3. Вспомогательные построения

Определяется геометрический шаблон клетки.

$$\left| 1 - \frac{|O_1O_3|}{|O_2O_4|} \right| > \varepsilon, \varepsilon \rightarrow 0 - \text{эллипс};$$

$$\left| 1 - \frac{|O_1O_3|}{|O_2O_4|} \right| \leq \varepsilon, \varepsilon \rightarrow 0 - \text{окружность},$$

где ε – заданная точность измерения. Принимается решение о геометрической форме исследуемой биологической клетки. Для рассматриваемого контура (рис. 3)

$$\left| 1 - \frac{|O_1O_3|}{|O_2O_4|} \right| = 0,3958 > \varepsilon = 10^{-5},$$

следовательно, форма исследуемого объекта представляет собой эллипс.

Любые измерения, расчеты, выводы обладают погрешностью, что требует применения некоторых методов ее определения на основании использования модельного описания параметров изучаемого объекта. В данном случае речь идет о случайной погрешности, обусловленной строением биологической клетки. При исследовании случайных погрешностей в силу предельной теоремы вероятностей для анализа обоснованно принять нормальный закон распределения вероятностей.

Статистическая обработка вычислений

n	x (пиксель)	y (пиксель)	ℓ [10]			$\Delta \ell_i^2$
1	4	92	1,1032			0,0084
2	5	94	1,1014			0,0001
...
36	47	151	1,0951			0,0000
37	33	144	1,0712			0,0000
Σ			37,4191			0,0326
$\bar{\ell}$ [10]	$S_{\bar{\ell}}$ [10]	$t(0,05; 36)$	$\Delta \ell = S_{\bar{\ell}} \cdot t$	$\ell = \bar{\ell} \pm \Delta \ell$	$\varepsilon = \Delta \ell \cdot 100\% / x$	$\Delta \ell / \delta = 8$
1,01113	0,0215	2,0315	0,0437	$\delta = 0,05$ (отбрасывается)	(0,9676; 1,055)	4,3216%

Для уменьшения влияния случайных ошибок в уравнение эллипса (окружности, окружность – частный случай эллипса) подставляются координаты n произвольных точек контура $\ell_i, i = \overline{1, n}, n \in N$. На основании проведенных вычислений значений линии контура рассчитывается их средняя величина $\bar{\ell}$. Находится погрешность отдельного расчета $\Delta \ell_i = \mu - \ell_i, \mu = 1$ – истинное значение кривой эллипса. Вычисляются квадраты абсолютных отклонений отдельных расчетов $\Delta \ell_i^2, i = \overline{1, n}, n \in N$. Определяется среднеквадратичная ошибка среднего арифметического $S_{\bar{\ell}}$, характеризующая точность, с которой получено среднее значение измеренной величины.

Для заданной надежности $P = 0,95$ и числа произведенных вычислений n определяется коэффициент Стьюдента t и находится доверительный интервал (погрешность измерения) $\Delta \ell$. Если величина погрешности результата измерения $\Delta \ell$ окажется сравнимой с величиной погрешности $\delta = 1 - P = 1 - 0,05$ строения клетки, то в качестве границы доверительного интервала возьмется $\Delta \ell$, вычисляемая по формуле. Если одна из ошибок окажется меньше другой в три или более раз, то меньшую ошибку следует отбросить. Окончательный результат записывается в виде интервала. Оценивается относительная погрешность результата измерений ε .

Выше в таблице приведена обработка вычислений, полученных в результате исследования контура клетки.

В одних и тех же условиях проделано n измерений, и наиболее вероятным значением измеряемой величины будет ее среднее значение $\bar{\ell}$, стремящееся к значению $\mu = 1$.

На рис. 4 показана аппроксимация контура клетки эллипсом.

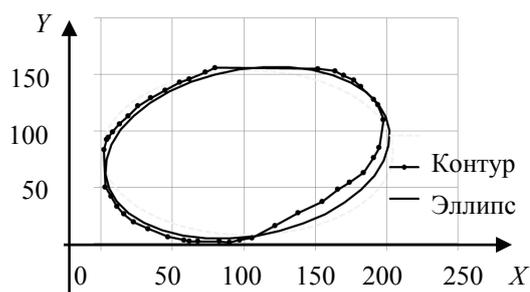


Рис. 4. Графическое сравнение линии анализируемого контура с аналитической кривой эллипса

Заключение. Полученные результаты могут быть, использованы в биомедицинской инженерии, мобильных приложениях для первичной диагностики патологий биотканей, а также как базовый программный комплекс для предварительной обработки цифровых медицинских изображений с выделением их геометрических параметров.

Литература

1. OpenCV шаг за шагом. Нахождение контуров и операции с ними. URL: <http://robocraft.ru/blog/computervision/640.html> (дата обращения: 16.10.2017).
2. Гмурман В. Е. Руководство к решению задач по теории вероятностей и математической статистике М.: Высш. школа, 2010. 400 с.
3. Csink L. Color normalization and object localization // Vierter Workshop Farbbildverarbeitung. Koblenz, 1998. P. 49–57.
4. Шапиро Л., Стокман Д. Компьютерное зрение. М.: БИНОМ, 2006. 752 с.
5. Fleyeh H. Road and traffic sign color detection and segmentation – A Fuzzy Approach // IAPR Conference on Machine Vision Applications. Tsukuba Science City, 2005. P. 124–127.

6. Wang F. Fuzzy-based algorithm for color recognition of license plates // *Pattern Recognition Letters*. 2008. Vol. 29, no. 7. P. 1007–1020.
7. Визильтер Ю. В. *Обработка и анализ изображений в задачах машинного зрения: курс лекций и практических занятий*. М.: Физматкнига, 2010. 672 с.
8. Детектирование объектов – поиск объекта по шаблону (Template matching) URL: <http://robocraft.ru/blog/computervision/3046.html> (дата обращения: 10.11.2017).
9. Воднев В. Т., Наумович А. Ф., Наумович Н. Ф. *Основные математические формулы*. Справочник. Минск: Выш. шк., 1988. 270 с.

References

1. *OpenCV shag za shagom. Nakhozhdenie konturov i operatsii s nimi* [OpenCV step by step. Finding contours and operations with them]. Available at: <http://robocraft.ru/blog/computervision/640.html> (accessed 16.10.2017).
2. Gmurman V. E. *Rukovodstvo k resheniyu zadach po teorii veroyatnostey i matematicheskoy statistike* [The management to the solution of tasks in probability theory and mathematical statistics]. Moscow, Vysshaya shkola Publ., 2010. 400 p.
3. Csink L. Color normalization and object localization. *Vierter Workshop Farbbildverarbeitung*. Koblenz, 1998, pp. 49–57.
4. Shapiro L., Stokman D. *Komp'yuternoe zrenie* [Computer vision]. Moscow, BINOM Publ., 2006. 752 p.
5. Fleyeh H. Road and traffic sign color detection and segmentation – A Fuzzy Approach. *IAPR Conference on Machine Vision Applications*. Tsukuba Science City, 2005, pp. 124–127.
6. Wang F. Fuzzy-based algorithm for color recognition of license plates. *Pattern Recognition Letters*, 2008, vol. 29, no. 7, pp. 1007–1020.
7. Vizil'ter Yu. V. *Obrabotka i analiz izobrazheniy v zadachakh mashinnogo zreniya: kurs lektsiy i prakticheskikh zanyatiy* [Processing and the analysis of images for solving machine vision problems: Course of lectures and practical training]. Moscow, Fizmatkniga Publ., 2010. 672 p.
8. *Detektirovanie ob'ektov – poisk ob'ekta po shablonu (Template matching)* [Object detection – object search by pattern]. Available at: <http://robocraft.ru/blog/computervision/3046.html> (accessed 10.11.2017).
9. Vodnev V. T., Naumovich A. F., Naumovich N. F. *Osnovnye matematicheskie formuly. Spravochnik* [Basic mathematical formulas. Reference book]. Minsk, Vysheyschaya shkola Publ., 1988. 270 p.

Информация об авторах

Бич Наталья Николаевна – кандидат физико-математических наук, инженер-программист кафедры информационных систем и технологий. Гродненский государственный университет имени Янки Купалы (230023, г. Гродно, ул. Ожешко, 22, Республика Беларусь). E-mail: Nadaliya@mail.ru

Ассанович Борис Алиевич – кандидат технических наук, доцент, доцент кафедры информационных систем и технологий. Гродненский государственный университет имени Янки Купалы (230023, г. Гродно, ул. Ожешко, 22, Республика Беларусь). E-mail: bas@grsu.by

Information about the authors

Bich Natal'ya Nikolaevna – PhD (Physics and Mathematics), software engineer, the Department of Information Systems and Technologies. Yanka Kupala State University of Grodno (22, Ozheshko str., 230023, Grodno, Republic of Belarus). E-mail: Nadaliya@mail.ru

Assanovich Boris Alievich – PhD (Engineering), Associate Professor, Assistant Professor, the Department of Information Systems and Technologies. Yanka Kupala State University of Grodno (22, Ozheshko str., 230023, Grodno, Republic of Belarus). E-mail: bas@grsu.by

Поступила 15.05.2018