



УДК 004.9

DOI 10.18413/2411-3808-2018-45-3-537-546

ИСПОЛЬЗОВАНИЕ АЛГОРИТМОВ КОМПЬЮТЕРНОГО ЗРЕНИЯ ДЛЯ ВЫПОЛНЕНИЯ ГЕМАТОЛОГИЧЕСКОГО АНАЛИЗА НА ОСНОВЕ КРИВОЙ ПРАЙС-ДЖОНСА

USING COMPUTER VISION ALGORITHMS FOR IMPLEMENTATION OF HEMATOLOGICAL ANALYSIS BASED ON PRICE-JONES CURVE

В.Г. Синюк¹, Д.С. Батищев², Е.С. Сойникова³, В.М. Михелев²
V.G. Sinyuk¹, D.S. Batishchev², E.S. Soynikova³, V.M. Mikhelev²

¹ Белгородский государственный технологический университет имени В.Г. Шухова,
Россия, 308012, г. Белгород, ул. Костюкова, 46

² Белгородский государственный национальный исследовательский университет,
Россия, 308015, г. Белгород, ул. Победы, 85

³ ООО «Технологии Надежности», Россия, 308000, г. Белгород, Свято-Троицкий бульвар, 11

¹ Belgorod State Technological University named after V.G. Shukhov,
46 Kostyukova St, Belgorod, 308012, Russia

² Belgorod State University, 85 Pobeda St, Belgorod, 308015, Russia

³ LLC «Technologies Reliability», 11 Svyato-Troitskiy bul'var, Belgorod, 308000, Russia

E-mail: vgsinuk@mail.ru, batishchev@bsu.edu.ru, 831468@bsu.edu.ru, mikhelev@bsu.edu.ru,

Аннотация

В данной статье рассматривается задача постановки возможного диагноза по гематологическому анализу цифрового изображения эритроцитов. Описываются шаги по предварительной обработке изображения для уменьшения шумов и увеличения точности сегментации объектов клеток на классы. Для каждого этапа приведены примеры работы фильтров. Заключительным шагом является построение гистограммы распределения площадей объектов клеток – кривая Прайс-Джонса. По форме гистограммы можно предположить о наличии заболевания у человека, у которого взяли препарат крови на исследование. В статье приводятся описания типичных заболеваний, которые могут быть выявлены с помощью такого рода анализа. Программное обеспечение гематологического анализа крови реализовано на языке Python 2.7 с использованием библиотек OpenCV и Seaborn.

Abstract

This article is about implementing a hematological analysis through computer vision algorithms. This type of analysis is one of the basic analyses providing huge amount information about patient and his state. We propose a pipeline with a few steps for image preprocessing thus image become more contrast and noiseless. At first image color space converting – so we separate a luminance channel and ignore other channels (due to source image features). Then we blur image with Gaussian filter and apply CLAHE filter for contrast improvement, so background pixels form more homogenous areas and become less bright in comparison to cell's pixels. The next step is background removal and image binarization based on Otsu algorithm for border pixel luminance level detection. Afterwards we extract an array of contours from binary image and use this array as an input source for Watershed algorithm. As a result, we have a color image where every single class of object has its own color and an array of object. This array then used as a source for cells diameters distribution histogram – a Price-Jones curve. All described steps implemented in Python 2.7 with OpenCV and Seaborn libraries.

Ключевые слова: гематологический анализ, клетки крови, сегментация изображения, компьютерное зрение, кривая Прайс-Джонса

Keywords: hematological analysis, blood cells, image segmentation, computer vision, Price-Jones curve.



Введение

В настоящее время интенсивно развиваются интеллектуальные системы, предназначенные для автоматической обработки медицинских изображений. Автоматизированная обработка и анализ медицинских изображений являются универсальным инструментом медицинской диагностики [Томакова и др., 2013; Борисовский, 2012]. Современные системы гематологического анализа позволяют выполнить предварительную классификацию клеток крови. Например, такие системы эффективно сортируют лейкоциты (8–10% ошибок) и обеспечивают более детальную информативность анализов крови [Беляков, 2014].

Классификация клеток крови на микроскопическом изображении представляет собой (в терминах компьютерного зрения) задачу распознавания объектов. Стандартный подход к решению этой задачи предусматривает следующие два этапа:

- разделение (сегментацию) изображения на области, соответствующие объектам и фону;
- распознавание объектов, включающее в себя выделение характерных признаков объектов и распределение объектов в соответствии с их признаками по классам.

Примерно за пять десятилетий развития компьютерного зрения как науки предложено множество методов сегментации изображений, формирования векторов признаков объектов и их классификации [Соколинский, 2009; Jambhekar, 2011]. Такое разнообразие ставит перед исследователем непростую проблему выбора методов, наиболее адекватных специфике конкретной задачи.

Специфика задачи сказывается на формировании векторов признаков объектов, на выборе метода сегментации, а также классификатора.

Таким образом, при выборе стратегии решения данной задачи особое внимание должно быть уделено изучению опыта применения различных методов сегментации микроскопических изображений и характерным признакам, используемым для классификации клеток крови.

Основным результатом сегментации микроскопического изображения мазка крови является выделение объектов интереса (клеток крови) с целью их дальнейшей классификации. Качество сегментации – ключевой фактор для получения адекватных значений характерных признаков объекта [Грибков, 2005]. Так, в работе [Соколинский, 2009] отмечено, что более половины ошибок классификации лейкоцитов были обусловлены неправильной сегментацией.

Основные причины, ведущие к ошибкам сегментации микроскопических изображений: перекрывание одной клетки другой, сильная вариация клеток по форме и размеру, воздействие разных факторов на внешний вид клетки, слабая контрастность изображений, зашумленность и артефакты на снимке препарата. Также влияет окраска препаратов крови: часто после окраски контрастность контуров структурных элементов внутри клетки превышает контрастность границ самой клетки, это может вызвать пересегментацию изображения, либо потерю части пикселей внутри объекта клетки [Biggs, MacMillan, 1948].

Описание метода

Предлагаемый нами [Сойникова и др., 2016; Рябых и др., 2017; Батищев, Михелев, 2016] метод предварительной обработки изображений предусматривает следующие основные этапы.

Смена цветового пространства – перевод изображения в цветовое пространство HSL (HLS в представлении OpenCV) с выделением отдельно канала L – светимость конкретного пикселя. Специфика исходных данных (изначально изображения представлены в оттенках серого) позволяет на первом этапе выделить один канал и с ним дальше работать. Большинство методов OpenCV, необходимых для обработки изображения, тоже работают с одним каналом.

В общем виде перевод цветового пространства изображения из RGB в HSL можно записать как:

$$\begin{aligned}
 V_{\max} &\leftarrow \max(R, G, B) \\
 V_{\min} &\leftarrow \min(R, G, B) \\
 L &\leftarrow \frac{V_{\max} + V_{\min}}{2} \\
 S &\leftarrow \begin{cases} \frac{V_{\max} - V_{\min}}{V_{\max} + V_{\min}}, L < 0.5 \\ \frac{V_{\max} - V_{\min}}{2 - (V_{\max} + V_{\min})}, L \geq 0.5 \end{cases} \\
 H &\leftarrow \begin{cases} \frac{60(G - B)}{S}, V_{\max} = R \\ 120 + \frac{60(B - R)}{S}, V_{\max} = G \\ 240 + \frac{60(R - G)}{S}, V_{\max} = B \end{cases}
 \end{aligned}$$

Так как в изображении в оттенках серого каждый пиксель в RGB кодируется одинаковыми значениями для каждой компоненты, достаточно выделить один любой канал в качестве L-канала HSL. Однако правильное разложение по каналом оставлено на тот случай, если выходные данные будут представляться полноцветными изображениями.

Удаление заднего фона на изображениях: в некоторых местах есть области переосвечивания пикселей и, наоборот, затемнения, из-за которых происходит неправильное контурирование клетки. Так как изначально изображение зашумлено, необходимо сгладить перепады между группами соседних пикселей.

Сглаживание изображения просто сделать через размытие по Гауссу. Ядра размером 3×3 вполне достаточно, чтобы убрать неравномерности в шумах фона, но в то же время не потерять границы объектов. Пример сглаживания представлен на рисунке 1.

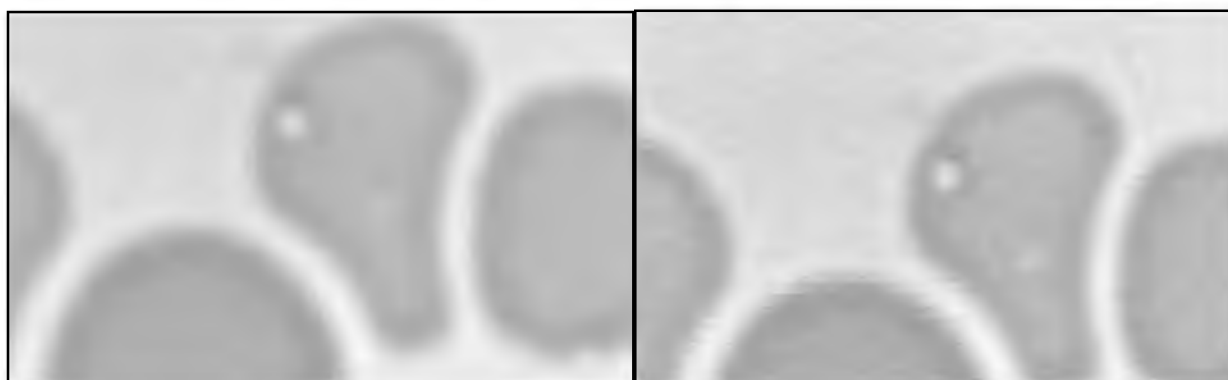


Рис. 1. Применение размытия Гаусса: слева – размытое изображение, справа – оригинал
 Fig. 1. Gaussian blur application: on the left – the blurred image, on the right – source one

После выравнивания фоновых пикселей изображения необходимо увеличить контрастность изображения, чтобы при удалении фоновых пикселей оставить пиксели объектов, которые примерно равны по светимости фоновым.

Для увеличения контрастности изображения используется алгоритм CLAHE (Contrast Limited Adaptive Histogram Equalization) [Sasi, Jayasree, 2013]. Данный алгоритм,

как и все алгоритмы с выравниванием гистограммы, использует функцию плотности вероятности (1) и кумулятивную функцию плотности (2) для приведения гистограммы интенсивности светимости пикселей к нужному виду. Если принять, что N – количество пикселей в изображении, L – общее количество оттенков серого (уровней интенсивности) на изображении, а n_k – это общее количество пикселей со светимостью l_k , то функция плотности вероятности и кумулятивная функция плотности будут иметь вид:

$$f_i(i_k) = \frac{n_k}{N}, \quad (1)$$

$$F_k(i_k) = \sum_{j=0}^k f_i(i_k). \quad (2)$$

Отличие CLAHE от других алгоритмов в том, что он выполняет эквализацию на ограниченной области изображения с предварительно ограниченной гистограммой светимости. Таким образом, он более устойчив к порождению шума в гомогенных областях.

Результат применения алгоритма CLAHE на исследуемом изображении приведен на рисунке 2.

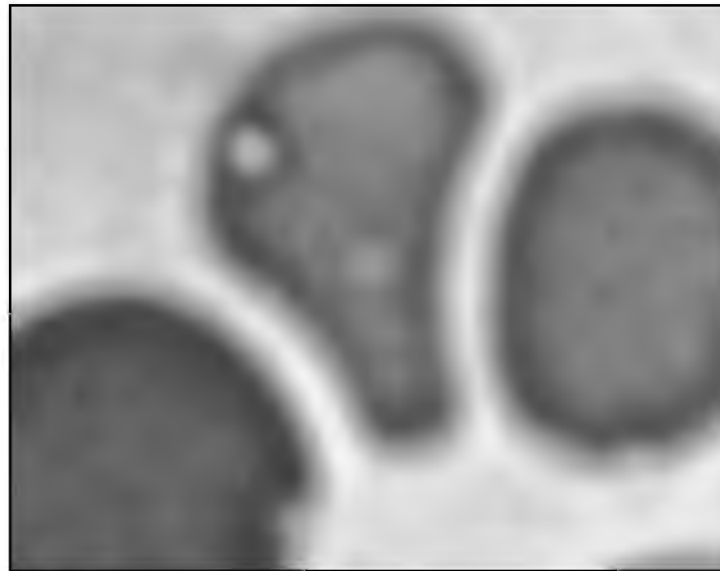


Рис. 2. Применения алгоритма CLAHE
Fig. 2. CLAHE filter application

Теперь, когда клетки (объекты) сильно контрастируют с фоном, можно удалить фоновые пиксели с изображения.

$$x_i = \begin{cases} 255, & x_i \geq \text{mean}(X) * 0.9 \\ x_i, & \text{в остальных случаях} \end{cases}$$

где x_i – светимость i -го пикселя изображения, X – множество всех пикселей.

То есть, в том случае, если светимость пикселя больше или равна 90% средней светимости по всему изображению, ему присваивается значение 255 – белый цвет, остальные пиксели неизменны. Значение в 90% было подобрано эмпирически. При таком пороге удаляются практически все фоновые пиксели, при этом остаются неизменными пиксели объектов. Пример удаления фона приведен на рисунке 3.



Рис. 3. Результат удаления фоновых пикселей
Fig. 3. Background removal result

Бинаризация изображения – приведение изображение к виду, когда каждый пиксель кодируется либо единицей, либо нулем. Этот шаг необходим, так как некоторые последующие функции OpenCV используют бинарное изображение как один из аргументов.

Выбор порога сегментации выполняется по методу Оцу. Так как после предыдущего шага на изображении присутствуют два класса пикселей: фоновые и объектные, – метод Оцу подходит для определения границы бинаризации лучше: выше будет межклассовая дисперсия [Image Thresholding, 2018]. Пример бинаризации изображения приведен на рисунке 4.



Рис. 4. Бинаризация изображения
Fig. 4. Image binarization

Определение контуров объектов на бинарном изображении. В OpenCV используется алгоритм топологического структурного анализа бинарных изображений, предложенный Сатоши Сузуки и Кейчи Эйбом [Satoshi, Keiichi, 1985]. Алгоритм предполагает нахождение контуров с учетом вложенности, то есть способен определить, когда в контур одного объекта вложен другой. В предметной области исследования это может возникать при наличии на снимке здоровых двояковогнутых эритроцитов. При этом при засветке вогнутость эритроцита не фиксируется матрицей камеры, таким образом, объект получается с «дыркой» и, с точки зрения алгоритма Suzuki85, содержит два объекта, а полную площадь можно посчитать, сложив площади самого объекта и его «дырки».

В OpenCV данный режим для `cv::findContours` называется `CV_RETR_CCOMP`. Он извлекает все контуры и организует их в двухуровневую иерархию. На верхнем уровне существуют внешние границы компонентов. На втором уровне есть границы отверстий. Если в отверстии подключенного компонента есть еще один контур, он все еще находится на верхнем уровне.

Пример выделения контуров объектов представлен на рисунке 5.



Рис. 5. Определение контуров объектов
Fig. 5. Objects contours detection

Выделение каждой клетки в свой класс. Классификация областей: пиксели, которые соответствуют номеру класса, красятся в соответствующий цвет.

Далее для наглядности и удобства расчета площади объекта найденные контуры заливаются случайными неповторяющимися цветами.

Алгоритм работает с изображением как с функцией от двух переменных $f = I(x, y)$, где x, y – координаты пикселя.

Значением функции может быть интенсивность или модуль градиента. Для наибольшего контраста можно взять градиент от изображения. Если по оси O_z откладывать абсолютное значение градиента, то в местах перепада интенсивности образуются хребты, а в однородных регионах – равнины. После нахождения минимумов функции f идет процесс заполнения «водой», который начинается с глобального минимума. Как только уровень воды достигает значения очередного локального минимума, начинается его заполнение водой. Когда два региона начинают сливаться, строится перегородка, чтобы предотвратить объединение областей. Вода продолжит подниматься до тех пор, пока регионы не будут отделяться только искусственно построенными перегородками [Watershed approaches for color image segmentation, 2018; Beucher, Meyer, 1992].

В данном случае информацией о перегородках выступают контуры объектов с предыдущего шага, а значение пикселей на контурах – высотой плато. Таким образом, каждая равнина (группа пикселей, заключенных в контур) является уникальным классом. За счет иерархичности контуров правильно (в рамках предметной области, то есть не артефакты и ошибки сегментации) заливают и вложенные объекты.

Исходное изображение мазка крови представлено на рисунке 6.

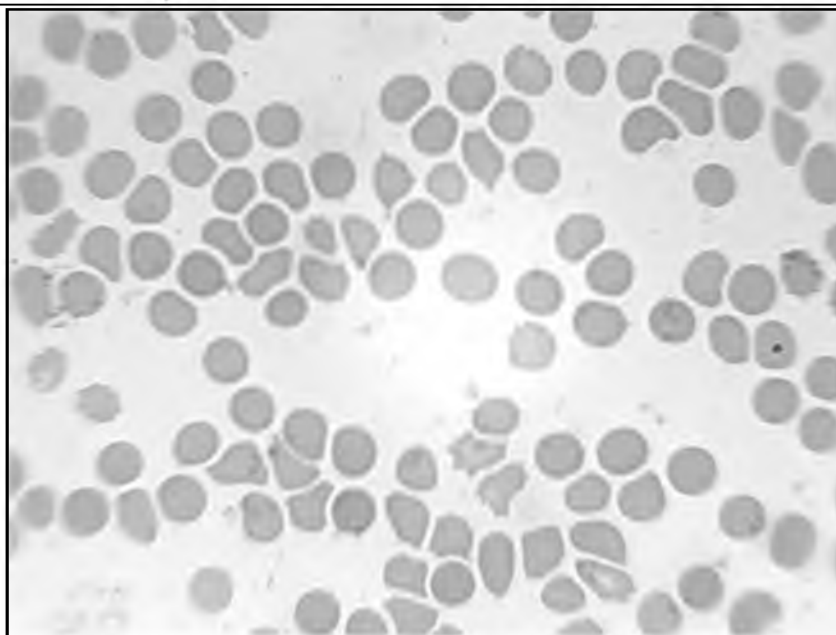


Рис. 6. Исходное изображение мазка крови
 Fig. 6. Source image of blood probe

В результате обработки исходного изображения согласно вышеприведенному методу мы получаем изображение, представленное на рисунке 7.

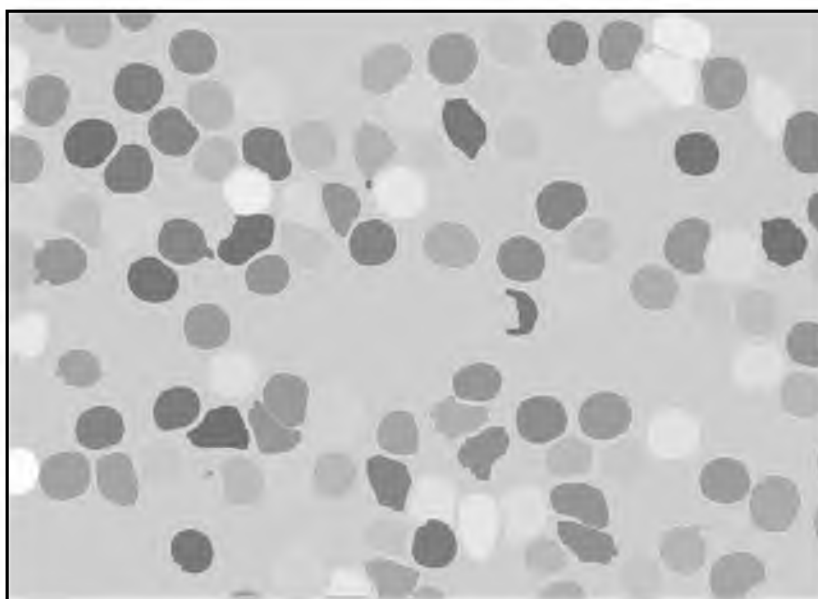


Рис. 7. Изображение мазка крови после предварительной обработки и сегментации
 Fig. 7. Source image after preprocessing and segmentation

После предварительной обработки изображения появляется возможность определить площади клеток крови по наиболее тонким и четким контурам.

Из площади клеток крови вычисляем диаметр каждой клетки и строим гистограмму распределения диаметров клеток, которая представлена на рисунке 8. Это возможно, так как мы знаем, что в нашем случае каждый пиксель на изображении занимает 0,154 мкм.

Кривая Прайс-Джонса может быть представлена как кривая плотности вероятности вхождения клетки в определенный класс (рис. 9), где по оси абсцисс откладывают величину площади эритроцитов (в мкм), а по оси ординат – вероятность вхождения клетки в свой класс, класс в данном случае – диаметр клетки.

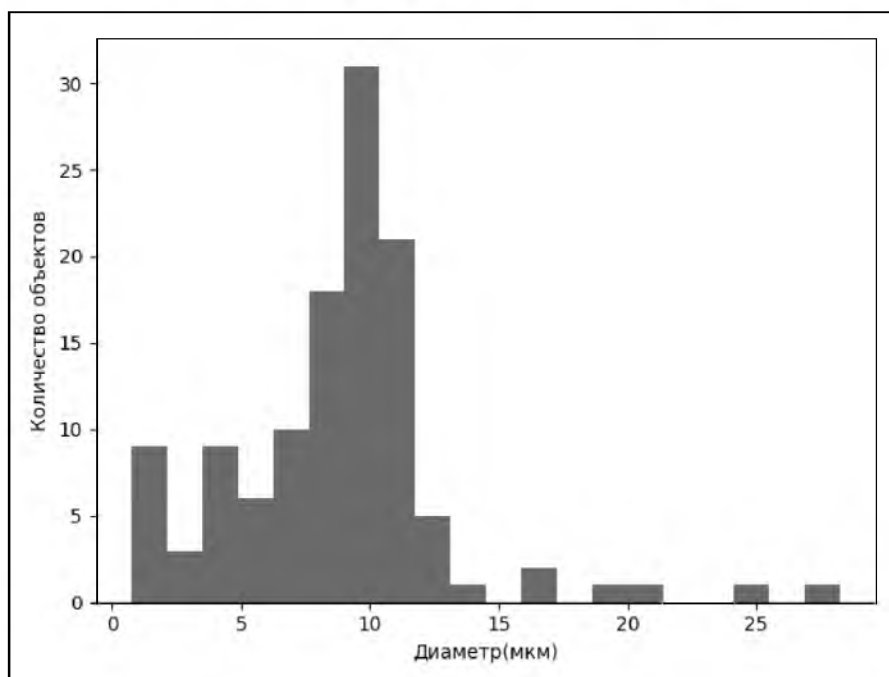


Рис. 8. Гистограмма распределения диаметров эритроцитов
Fig. 8. An erythrocytes diameters distribution histogram

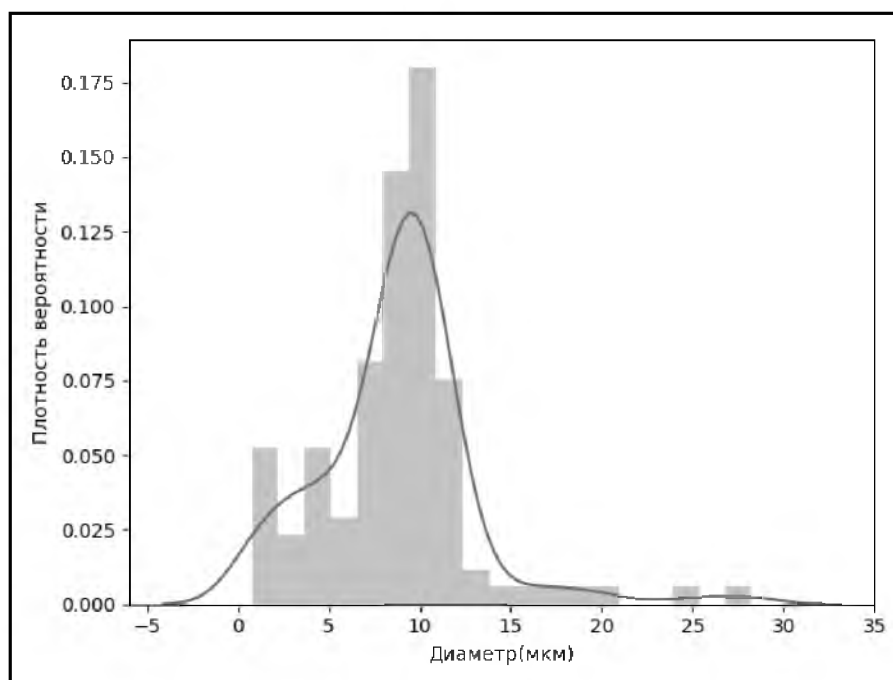


Рис. 9. Кривая плотности вероятности вхождения клетки в класс
Fig. 9. A curve of probability density of cell class affiliation

Кривая Прайс-Джонса у здоровых людей имеет правильную треугольную форму с высокой вершиной и узким основанием. При этом преобладают эритроциты с диаметром 6-8 мкм, которые составляют 70-75% всех эритроцитов. Микроцитов (клетки диаметром меньше 6 мкм) и макроцитов (диаметр более 8 мкм) приблизительно одинаковое количество: они составляют 12-15%; ширина кривой отражает степень анизоцитоза, а положение максимума – средний диаметр эритроцита [Sasi, Jayasree, 2013; Методическое руководство, 2018].

При микроцитозе (характерном, например, для железодефицитной анемии) эритроцитометрическая кривая сдвигается влево, кривая становится ассиметричной, ширина ее увеличивается [Hawksley et al., 1934].

При макроцитозе (например, сопровождающем В12 и фолиеводефицитную анемию) кривая Прайс-Джонса сдвигается вправо, уплощается, основание ее расширяется [Bessman, Feinstein, 1979; Липунова, Скоркина, 2004].

Заключение

Построение кривой Прайс-Джонса вручную – чрезвычайно трудоемкая процедура. Поэтому разработанная система гематологического анализа позволяет в автоматическом режиме быстро и с высокой точностью построить кривую Прайс-Джонса, что дает возможность увидеть процентное соотношение всех видов красных кровяных телец. Это позволяет достаточно быстро выявить анизоцитоз у больного и подобрать соответствующие методы его лечения.

Вычислительные эксперименты по проведению гематологического анализа клеток крови выполнялись с использованием программного обеспечения, реализованного на языке Python 2.7 с использованием библиотек OpenCV и Seaborn. Результаты вычислительных экспериментов продемонстрировали работоспособность и эффективность разработанных алгоритмов компьютерного зрения для выполнения гематологического анализа на основе кривой Прайс-Джонса.

Работа выполнена при финансовой поддержке гранта РФФИ 16-07-00435 А.

Список литературы References

1. Батищев Д.С., Михелев В.М. 2016. Инфраструктура высокопроизводительной компьютерной системы для реализации облачных сервисов хранения и анализа данных персональной медицины. Научные ведомости Белгородского государственного университета. Серия: Экономика. Информатика, 2 (223): 88–92.

Batishhev D.S., Mihelev V.M. 2016. Infrastruktura vysokoproizvoditel'noj komp'juternoj sistemy dlja realizacii oblachnyh servisov hranenija i analiza dannyh personal'noj mediciny. Nauchnye vedomosti Belgorodskogo gosudarstvennogo universiteta. Serija: Jekonomika. Informatika, 2 (223): 88–92. (in Russian)

2. Беляков В.К., Сухенко Е.П., Захаров А.В., Кольцов П.П., Котович Н.В. 2014. Об одной методике классификации клеток крови и ее программной реализации. Программные продукты и системы, 4 (108): 46–56.

Beljakov V.K., Suhenko E.P., Zaharov A.V., Kol'cov P.P., Kotovich N.V. 2014. Ob odnoj metodike klassifikacii kletok krovi i ee programmnoj realizacii. Programmnye produkty i sistemy, 4 (108): 46–56. (in Russian)

3. Борисовский С.А. 2012. Гибридные модели и алгоритмы для анализа сложно структурированных изображений в интеллектуальных системах медицинского назначения. Дис. ... канд. техн. Наук. Курск, ЮЗГУ.

Borisovskij S.A. 2012. Gibridnye modeli i algoritmy dlja analiza slozhno strukturirovannyh izobrazhenij v intellektual'nyh sistemah medicinskogo naznachenija. Dis. kand. tehn. nauk. Kursk, JuZGU. (in Russian)

4. Грибков И.В., Захаров А.В., Кольцов П.П., Котович Н.В., Кравченко А.А. 2005. Сравнительное исследование методов анализа изображений. Автоматика и радиоэлектроника, НИИСИ РАН.

Gribkov I.V., Zaharov A.V., Kol'cov P.P., Kotovich N.V., Kravchenko A.A. 2005. Sravnitel'noe issledovanie metodov analiza izobrazhenij. Avtomatika i radioelektronika, NIISI RAN. (in Russian)

5. Липунова Е.А., Скоркина М.Ю., 2004. Система красной крови: сравнительная физиология, НИУ «БелГУ».

Lipunova E.A., Skorkina M.Ju. 2004. Sistema krasnoj krovi: sravnitel'naja fiziologija, NIU «BelGU». (in Russian)



6. Методическое руководство: Общий анализ крови (трактовка результатов исследований, выполненных на гематологических анализаторах), 2. Ставропольский государственный медицинский университет.

Methodicheskoe rukovodstvo: Obshhij analiz krovi (traktovka rezul'tatov issledovaniy, vypolnennyh na gematologicheskikh analizatorah), 2. Stavropol'skij gosudarstvennyj medicinskij universitet. (in Russian)

7. Рябых, М.С., Батищев Д.С., Михелев В.М. 2017. Метод анализа и обработки медицинского изображения клетки крови для определения типа лейкоза. В кн.: Научные тенденции: вопросы точных и технических наук. Сборник научных трудов по материалам XI Международной научной конференции, 23–29.

Rjabyh M.S., Batishhev D.S., Mihelev V.M. 2017. Metod analiza i obrabotki medicinskogo izobrazhenija kletki krovi dlja opredelenija tipa lejkoza. In.: Nauchnye tendencii: voprosy tochnyh i tehniceskikh nauk. Sbornik nauchnyh trudov po materialam XI Mezhdunarodnoj nauchnoj konferencii, 23–29. (in Russian)

8. Сойникова Е.С., Рябых М.С., Батищев Д.С., Синюк В.Г., Михелев В.М. 2016. Высокопроизводительный метод обнаружения границ на медицинских изображениях. Научный результат. Информационные технологии, 1 (3): 4–9.

Sojnikova E.S., Rjabyh M.S., Batishhev D.S., Sinjuk V.G., Mihelev V.M., 2016. Vysokoproduktivnyj metod obnaruzhenija granic na medicinskih izobrazhenijah. Nauchnyj rezul'tat. Informacionnye tehnologii, 1 (3): 4–9. (in Russian)

9. Соколинский Б.З., Демьянов В.Л., Медный В.С., Парпара А.А., Пятницкий А.М. 2009. Автоматическая сортировка лейкоцитов мазка крови с использованием методов обучаемых нейронных сетей и watershed. Методы микроскопического анализа, 128–132.

Sojnikova E.S., Rjabyh M.S., Batishhev D.S., Sinjuk V.G., Mihelev V.M. 2016. Vysokoproduktivnyj metod obnaruzhenija granic na medicinskih izobrazhenijah. Nauchnyj rezul'tat. Informacionnye tehnologii, 128–132. (in Russian)

10. Томакова Р.А., Филист С.А., Жилин В.В., Борисовский С.А. 2013. Программное обеспечение интеллектуальной системы классификации форменных элементов крови. Фундаментальные исследования, 10 (2): 303–307.

Tomakova R.A., Filist S.A., Zhilin V.V., Borisovskij S.A. 2013. Programmnoe obespechenie intellektual'noj sistemy klassifikacii formennyh jelementov krovi. Fundamental'nye issledovaniya, 10 (2): 303–307. (in Russian)

11. Bessman J.D., Femstein D.I. 1979. Quantitative Anisocytosis as a Discriminant Between Iron Deficiency and Thalassem. Blood, 53. Date Views 1.04.2018 www.bloodjournal.org/content/bloodjournal/53/2/288.full.pdf?ssoc-checked=true.

12. Beucher S., Meyer F., 1992. Optical Engineering. New York: Marcel Dekker Incorporated.

13. Biggs R., MacMillan R.L., 1948. The errors of some hematological methods as they are used in a routine laboratory. J Clin Pathol, 1. Date Views 1.04.2018 jcp.bmj.com/content/jclinpath/1/5/269.full.pdf.

14. Hawksley J.C., Lightwood R., Bailey U.M. 1934. Iron-deficiency anaemia in children: Its association with gastro-intestinal disease, achlorhydria and hemorrhage. Archives of disease in childhood, 9. Date Views 1.04.2018 pdfs.semanticscholar.org/6a86/f416daf9c3d90217db7e25cb86273bb1be42.pdf.

15. Image Thresholding. Date Views 01.04.2018 docs.opencv.org/trunk/d7/d4d/tutorial_py_thresholding.html.

16. Jambhekar N. Red blood cells classification using image processing. Science Research. 2011. Date Views 1.04.2018 studyres.com/doc/17754179/red-blood-cells-classification-using-image.

17. Price-Jone, S. Lond M.B., 1910. The variation in the sizes of reb blood cells. British Medical Journal, 2. Date Views 1.04.2018 digitalcommons.ohsu.edu/cgi/viewcontent.cgi?article=1062&context=hca-cac.

18. Sasi N.M., Jayasree V.K. 2013. Contrast Limited Adaptive Histogram Equalization for Qualitative Enhancement of Myocardial Perfusion Images. Engineering, 5. Date Views 1.04.2018 file.scirp.org/pdf/ENG_2013110109155688.pdf.

19. Satoshi S., Keiich A. 1985. Topological Structural Analysis of Digitized Binary Images by Border Following. Computer vision, graphics, and image processing, 30. Date Views 1.04.2018 download.xuebalib.com/xuebalib.com.17233.pdf.

20. Watershed approaches for color image segmentation. Date Views 1.04.2018 www.gipsa-lab.grenoble-inp.fr/~jocelyn.chanussot/publis/ieee_nsip_99_chanuss_watershed.pdf.