

## АЛГОРИТМЫ РАСПОЗНАВАНИЯ КЛЕТОК КРОВИ

Б.И. Шахтарин<sup>1</sup>, С.А. Панов<sup>1</sup> К.С. Калашников<sup>2</sup>

<sup>1</sup>МГТУ им. Н.Э. Баумана, Москва, Российская Федерация  
e-mail: panovstepan@gmail.com

<sup>2</sup>ООО “НПП “Цифровые решения”, Москва, Российская Федерация  
e-mail: kalash@dsol.ru

*Рассмотрена структура системы анализа медицинских изображений. Приведены принципиальная схема установки для проведения автоматизированных микроскопических исследований мазков крови, а также алгоритм работы системы распознавания клеток крови. Сформулированы основные задачи, решаемые при проведении морфологического анализа крови. Определены требования к алгоритму при определении лейкоцитарной формулы и обнаружении форменных элементов крови на мазке. Предложена модель цветоярких характеристик для описания типичных изображений мазка крови. Определены пороговые значения размеров объектов при поиске клеток. Исследована гистограмма яркости типичного поля зрения. Описан двухэтапный алгоритм обнаружения клеток крови, а также алгоритм построения разделяющей прямой на плоскости относительных цветов. Приведены результаты экспериментов на реальных препаратах. Рассмотрены причины возникновения ошибок обнаружения.*

**Ключевые слова:** анализ мазков крови, автоматизированная микроскопия, обнаружение клеток крови, машинное зрение.

## ALGORITHMS OF BLOOD CELLS RECOGNITION

B.I. Shakhtarin<sup>1</sup>, S.A. Panov<sup>1</sup> K.S. Kalashnikov<sup>2</sup>

<sup>1</sup>Bauman Moscow State Technical University, Moscow, Russian Federation  
e-mail: panovstepan@gmail.com

<sup>2</sup>ООО Scientific Production Enterprise “Digital Solutions”,  
Moscow, Russian Federation  
e-mail: kalash@dsol.ru

*The structure of the medical image analysis system is considered. A schematic diagram of the system for automated microscopic studies as well as the algorithm of recognition of blood cells are shown. The main task to be solved during the morphological analysis of blood is formulated. The requirements are specified for the algorithm used for determining leukocyte counts and detecting blood cells. A model of color and brightness characteristics to describe typical images of a blood smear is offered. The threshold values of the object size at searching the cells are specified. The luminance histogram of a typical field of view is studied. Two-step algorithm for detecting blood cells as well as the algorithm for constructing a separating line are described. The results of the experiments on real samples are given. The causes of detection errors are considered.*

**Keywords:** analysis of blood smears, automated microscopy, detection of blood cells, machine vision.

Визуальная оценка морфологических характеристик клеток крови является неотъемлемой частью анализа крови человека. Даже с развитием технологий проточной цитометрии невозможно сформировать полную картину без анализа морфологии клеток. Поэтому остро

стоит задача автоматизации процесса визуальной микроскопии. Для решения подобных задач обычно используются компьютерные анализаторы изображений на базе микроскопов проходящего света. На рис. 1 представлена принципиальная схема компьютерного анализатора изображений, который состоит из моторизованного микроскопа с блоком управления, видеокамеры и компьютера с программным обеспечением.

Алгоритм работы системы для анализа клеток крови представлен на рис. 2. После установки препарата на микроскоп запускается программа анализа препарата. Блок управления по команде с компьютера инициирует перемещение предметного столика, а видеокамера захватывает изображения полей зрения. Алгоритм оценки фокуса анализирует полученное изображение и принимает решение о фокусировке либо переходе на следующий кадр. Полученные кадры поступают на вход алгоритма обнаружения объектов. В зависимости от требуемого числа объектов, система переходит к следующему полю зрения или проводит сегментацию и классификацию объектов.

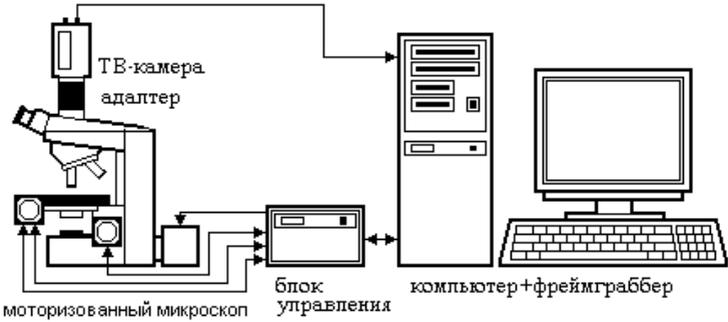


Рис. 1. Принципиальная схема компьютерного анализатора изображений

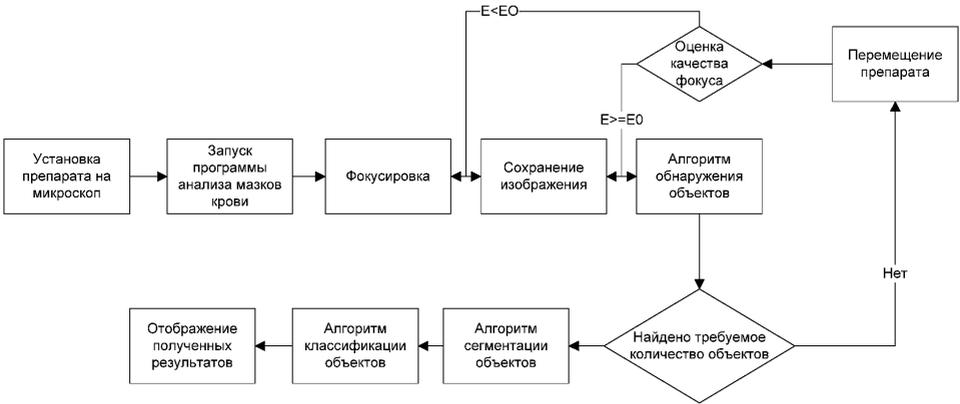


Рис. 2. Алгоритм работы системы

На практике при обнаружении клеток крови возникает типичная для медицинских приложений проблема варибельности окраски препаратов. Требования к окраске размыты и относительно, необходимо использовать адаптивные алгоритмы для обнаружения объектов. Типичное описание окраски препаратов выглядит следующим образом: эритроциты — розовато-бурые, ядра лейкоцитов — черно-фиолетовые или пурпурно-фиолетовые, цитоплазма лимфоцитов — голубая, цитоплазма моноцитов — голубая или сиренево-голубая. Цитоплазма нейтрофилов должна содержать ясно видимую нейтрофильную зернистость розоватого цвета, эозинофилы должны содержать гранулы оранжевого цвета [1].

**Требования к алгоритму обнаружения при определении лейкоцитарной формулы.** Анализ мазка крови должен выполняться в соответствии с общепринятыми методическими указаниями [2], поэтому подсчет лейкоцитарной формулы с помощью компьютерного анализатора изображений в простейшем случае состоит из четырех независимых этапов [3]:

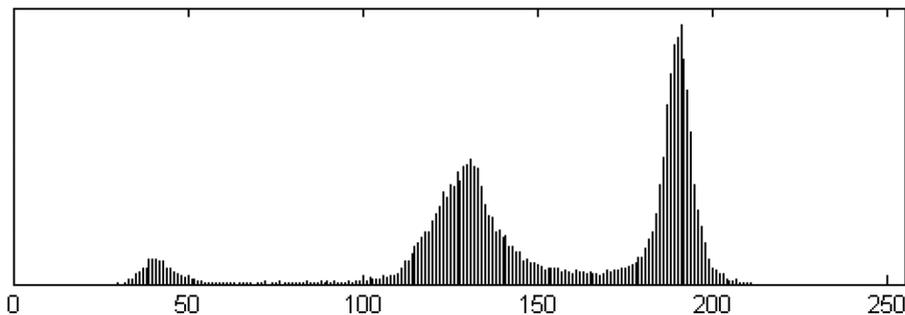
1) обнаружение лейкоцитов — из большого кадра, соответствующего одному полю зрения, вырезаются маленькие фрагменты, каждый из которых содержит один лейкоцит;

2) сегментация — в полученных малых фрагментах выделяются четыре области: ядро, цитоплазма лейкоцита, возможно, прилегающие эритроциты и фон изображения;

3) измерение геометрических, цветовых и яркостных характеристик ядра и цитоплазмы;

4) классификация лейкоцитов.

**Гистограмма яркости типичного поля зрения в мазке крови.** В простейшем случае на гистограмме яркости типичного кадра (независимо от того, как определяется яркость) выделяются три пика, соответствующие ядрам лейкоцитов, эритроцитам и фону (рис. 3).



**Рис. 3.** Гистограмма яркости типичного поля зрения в мазке крови в простейшем случае: видны три изолированных пика — ядра лейкоцитов, эритроциты, фон (слева направо)

В правильно окрашенном препарате оптическая плотность ядер по зеленой компоненте  $ODG = \ln(G_0/G) = 0,6 \dots 2,0$ ; оптическая плотность эритроцитов в тонком слое  $ODG = 0,2 \dots 0,5$ . На практике возникают следующие проблемы:

- малое число пикселей, соответствующих ядрам лейкоцитов: маленькое ядро или часть ядра на краю кадра;
- “светлые” (слабо окрашенные) ядра лейкоцитов;
- “темные” эритроциты могут иметь оптическую плотность, близкую к оптической плотности ядер лейкоцитов вплоть до 1,1.

Эти причины могут привести к тому, что на гистограмме яркости ядрам лейкоцитов и эритроцитам будет соответствовать один общий пик. Число пиков в гистограмме яркости оказывается непостоянным, а меняется от 1 до 5. Пик эритроцитов может быть слабо выражен или вообще отсутствовать: в кадре имеются только лейкоциты и фон. Если в эритроцитах имеется выраженное просветление в центре клетки, то пик эритроцитов может раздваиваться. Иногда из-за неоднородного освещения может раздваиваться даже пик фона. Поэтому одномерный алгоритм обнаружения ядер, основанный на гистограмме яркости, не будет устойчивым. Ему следует предпочесть двухмерные, а в идеале — трехмерные алгоритмы, использующие все цвета одновременно.

**Общая схема алгоритма.** При обнаружении лейкоцитов используются два метода сегментации “первичных объектов”. В начале скрининга информация о цветах лейкоцитов и эритроцитов в данном препарате предполагается неизвестной. Поэтому используется одномерный итеративный метод, основанный на изучении пиков гистограммы яркости. После того, как с его помощью найдено несколько ядер лейкоцитов и накоплена информация о цветах эритроцитов, на плоскости относительных цветов  $fR, fB$  строится прямая, разделяющая цвета ядер и цвета эритроцитов. Если граница между цветами эритроцитов и ядер проведена “с запасом” (ошибки первого и второго рода малы), она начинает использоваться для сегментации. Итак, при выделении первичных объектов (возможно, ядер) происходит переход с одномерного метода на двухмерный: вместо сегментации гистограммы яркости проводится сегментация плоскости цветов. Последний вариант более устойчив: нет опасности пропуска светлых ядер, проще работать с плохо сфокусированными кадрами, не нужно точно определять положение пика фона [4].

Алгоритм выделения фрагментов состоит из четырех шагов.

1. Выделение первичных объектов — возможно ядер. При этом используются два разных метода сегментации — по гистограмме яркости или по цветовой плоскости. Второй метод предпочтительней, но на начальной стадии накопления выборки используется первый.

2. Проверка первичных объектов на соответствие с уже накопленной выборкой ядер, что возможно, если число накопленных объектов

более 5. Первичные объекты классифицируются на фрагменты ядер и артефакты.

3. Фрагменты ядер объединяются друг с другом, причем делается попытка достраивания вокруг них цитоплазмы. В результате получаем прямоугольник, внутри которого находится один лейкоцит.

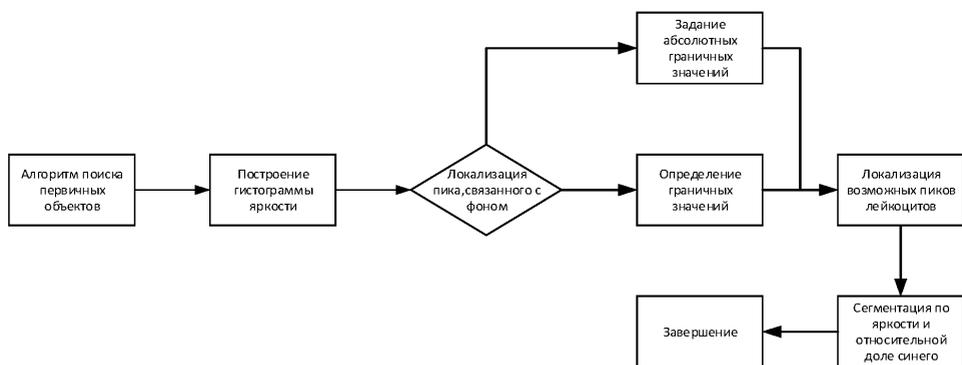
4. Если для сегментации первичных объектов использовалась гистограмма яркости, причем на ней имелось несколько пиков, возможно соответствовавших ядрам, то число найденных пикселей ядер сравнивается с предполагаемым (числом пикселей в пике). В результате проверки гипотеза о том, какой пик следует считать пиком ядер, может измениться и алгоритм запуститься с первого шага заново.

Как правило, необходимости в итерациях не возникает, алгоритм состоит из трех шагов: первичные объекты  $\implies$  фрагменты ядер  $\implies$  лейкоциты.

**Алгоритм сегментации, основанный на гистограмме яркости.** Предлагаемый алгоритм обнаружения лейкоцитов состоит из двух этапов, которые могут повторяться несколько раз для одного и того же кадра. На первом этапе (на основе изучения гистограммы яркости кадра и предыстории поиска) выбираются пороговые значения для яркости  $G$  и доли синего цвета  $fB$ . На втором этапе множества пикселей, удовлетворяющие этим условиям (первичные объекты), исследуются на предмет того, можно ли их рассматривать в качестве ядер лейкоцитов. Если общее число пикселей в этих фрагментах существенно меньше заранее предполагаемого числа, то выбор пороговых значений признается неудовлетворительным и алгоритм запускается еще раз и т.д.

**Первый этап** — это изучение гистограммы, выбор пороговых значений. Первый шаг состоит в локализации пика, связанного с фоном изображения, который в дальнейшем будет являться опорным, как при вычислении оптической плотности, так и при определении относительных цветов для остальных пикселей. Абсолютные значения приходится использовать только, если пик фона не локализуется. В качестве пика фоновых пикселей выбирается крайне правый пик с достаточно малой дисперсией: среднеквадратическое отклонение меньше 10 разрядов [5].

Далее составляется список максимумов (пиков), которые могли бы соответствовать ядрам лейкоцитов. Для этого их оптическая плотность должна быть достаточно большой (эмпирически найденная граница  $> 0,6$ ) и среднее значение доли синего  $fB = B/(B + G + R)$  для пикселей в этом максимуме должно превосходить аналогичное значение для фона на 0,03 (эмпирически найденная граница). Если таких подозрительных максимумов найдено несколько, то они выбираются последовательно друг за другом (при этом происходит вызов второй стадии алгоритма), начиная с самого светлого. Проводится пороговая



**Рис. 4.** Схема алгоритма обнаружения первичных объектов

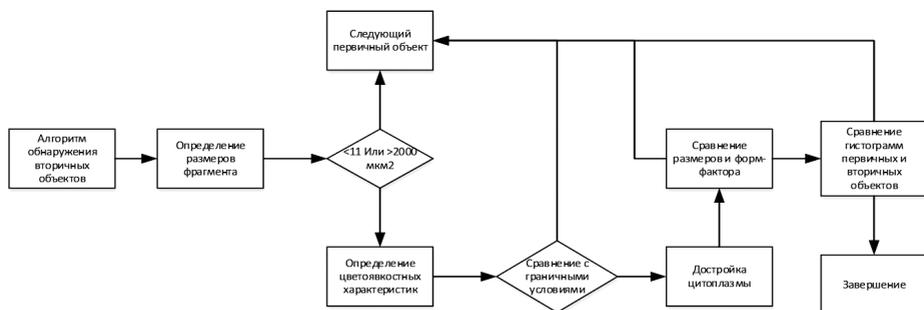
сегментация по яркости и относительной доли синего цвета  $G < G_{\max}$ ;  $fB = B/(B + G + R) < fB_{\max}$ .

Полученные первичные объекты сравниваются с уже существующими ядрами. Если они не будут квалифицированы как ядра, а это возможно при наличии оптически плотных и синеватых эритроцитов, пятен краски, крупных тромбоцитов, то будет проведен выбор следующего максимума. Если лейкоцитарный пик вообще не выделяется, то используются пороговые значения, основанные уже не на текущей гистограмме, а на предыстории поиска, а если таковой нет, то на априорных значениях. Схема алгоритма представлена на рис. 4.

**Второй этап** — исследование полученных фрагментов. Эта часть не зависит от того, каким методом были получены первичные объекты. Алгоритм проверки выделенных объектов состоит из трех циклов. В первом цикле отбрасываются слишком большие ( $A > 2000 \text{ мкм}^2$ ) и слишком маленькие ( $A < 11 \text{ мкм}^2$ ) объекты.

Далее измеряются оптическая плотность и цветовые характеристики. Если имеется предыстория поиска, то по критерию “трех сигм” отбрасываются чрезмерно светлые объекты с малой оптической плотностью и вызывается процедура проверки цветов. Оставшиеся объекты помещаются в класс условных ядер лейкоцитов.

Во втором цикле по оставшимся объектам происходит достройка цитоплазмы вокруг ядер. К цитоплазме относят связанное множество близлежащих пикселей, которые с большой вероятностью (более 0,95) не являются пикселями эритроцитов или фона. Построенное множество отвергается, если оно слишком велико (более  $2000 \text{ мкм}^2$ ) или форм-фактор его внешней границы (квадрат периметра/площадь) превышает достаточно большое значение (более 50). Достаточно часто при плотном прилегании к лейкоциту близких по окраске эритроцитов цитоплазму не удается достроить таким достаточно простым образом. Тогда ядро или его фрагмент помещаются внутрь прямоугольника с добавленными рамками в 15 мкм. После достраивания цитоплазмы



**Рис. 5.** Схема алгоритма поиска вторичных объектов

проводится агглютинация фрагментов ядер. Это необходимо, так как ядра нейтрофилов обнаруживаются в виде нескольких фрагментов. В таком случае проводится разъединение клеток, оказавшихся близко друг к другу. Это возможно, если клетки лежат в несвязных друг с другом островах, окруженных фоном.

После второго цикла полученные объекты рассматриваются уже как отдельные лейкоциты. В третьем цикле снова проверяются размеры этих объектов, и слишком большие объекты отбрасываются.

Если при сегментации использовалась гистограмма яркости, то в конце делается проверка согласованности предположений и полученных результатов. Для этого найденное число пикселей ядер лейкоцитов сравнивается с числом пикселей в том пике гистограммы, который предполагался соответствующим ядрам. Если различия существенны (превышают  $> 50\%$ ) и левее пика предполагаемых лейкоцитов был еще один, то приведенный алгоритм запускается еще раз. На рис. 5 представлена схема алгоритма обнаружения вторичных объектов.

Алгоритм проверки первичного объекта на принадлежность к группе ядер лейкоцитов. Ядра лейкоцитов не составляют однородную группу. Поэтому не стоит строго относиться к проверке на принадлежность нового объекта к двумерному, нормальному распределению, даже для средних значений относительных цветов. Кроме того, в процессе скрининга мазка желательно использовать алгоритм проверки как можно раньше, когда число накопленных объектов мало. Поэтому предлагаемый далее алгоритм носит эвристический характер. В его основу положены следующие положения.

1. Каждый новый объект сравнивается с двумя группами: группой ядер и группой эритроцитов.

2. При определении вероятности принадлежности новой точки к уже существующей группе вероятность подсчитывается дважды. В первом случае вероятность  $P_1$  вычисляется до прибавления точки к уже существующей статистике, а при втором —  $P_2$ , после такого прибавления. Очевидно, что  $P_2 > P_1$ . Подобное вычисление сразу двух

вероятностей необходимо, если решения принимаются на основе малых выборок, а число накапливаемых без проверки объектов обычно как раз мало — порядка 5. Если число объектов в группе превышает несколько десятков, то разница между  $P_1$  и  $P_2$  практически исчезает.

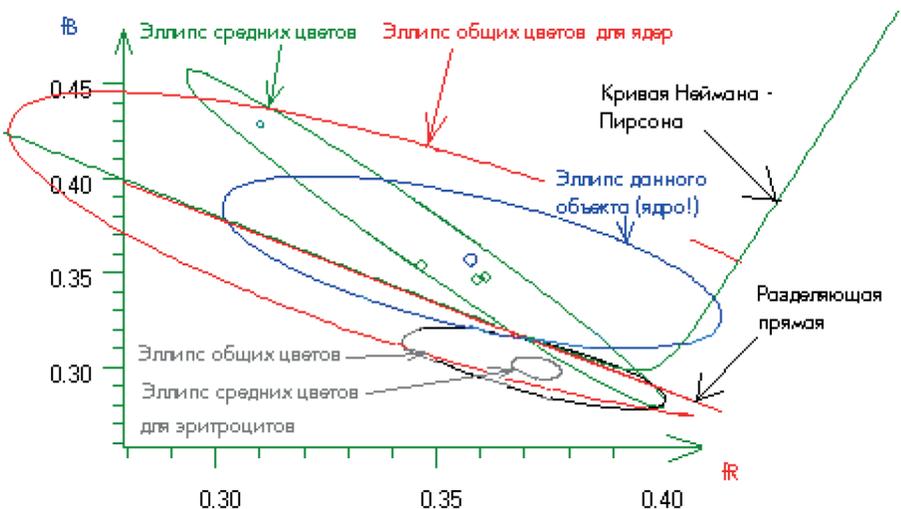
3. Если исследуемый объект более “сине-красный”, чем уже накопленные ядра, то он принимается в любом случае. Иными словами, возможность существования выбросов в эту “сине-красную” сторону не учитывается. Наоборот, если объект более “зеленый” чем эритроциты, то он в любом случае отвергается.

4. Если вероятность  $P_2$ , подсчитанная относительно группы эритроцитов, больше, чем соответствующая вероятность, подсчитанная относительно ядер, то объект отвергается.

5. Если на основе предыдущих пунктов решение не принято, то оно принимается с учетом трех вероятностей. Вероятности  $P_1$  и  $P_2$  оценивают отклонение среднего значения, а вероятность  $P_3$  — отклонение площади эллипса рассеяния от характерных для группы ядер. Объект отвергается, если  $P_2 < 0,01$  или  $P_3 < 0,01$ .

На рис. 6 показан пример проверки первичного объекта (“Эллипс данного объекта” и его центр имеют синий цвет). Средние цвета ранее принятых в качестве ядер объектов изображены зелеными кружками. Верхний из них на самом деле соответствовал выбросу — пятну краски, но был принят в силу своей “сине-красной” природы. Это исказило эллипс рассеяния средних цветов ядер, вытянув его по диагонали. Для данного первичного объекта соответствующие вероятности  $P_1 = 0,53$ ,  $P_2 = 0,65$ ,  $P_3 = 0,45$  велики, поэтому он будет верно квалифицирован как ядро и добавлен к статистике ядер.

**Алгоритм построения разделяющей линии на плоскости относительных цветов.** Для решения задачи классификации объектов це-



**Рис. 6.** Пример проверки данного объекта на принадлежность к выборке ядер

лесообразно проводить пересчет условных вероятностей принадлежности объектов к уже существующим группам. При этом получается больше информации, а время, потраченное на дополнительные вычисления, мало [5]. Действительная ценность простой и проведенной “с запасом” разделяющей границы так, чтобы практически все цвета пикселей ядер оказались по одну ее сторону, а цвета эритроцитов по другую, — это использование ее для сегментации объектов первого уровня (“возможно ядер”) на следующих кадрах. Как правило, это оказывается возможным.

В настоящее время проводится только одна разделяющая линия между двумя однородными группами — ядрами лейкоцитов и эритроцитами. Дело в том, что из остальных объектов (тромбоцитов, пятен краски и других загрязняющих мазок пятен) трудно сформировать однородные группы. Для двух нормальных случайных распределений с различными ковариационными матрицами оптимальной разделяющей кривой является линия Неймана–Пирсона (кривая второго порядка, обычно эллипс или гипербола), на которой отношение соответствующих плотностей вероятностей постоянно. Величину этого отношения можно выбрать, задавая вероятность ошибки первого рода — отнести объект первого типа (ядро) к объектам второго типа (эритроцитам). При этом вероятность ошибки второго рода (принять эритроцит за ядро) оказывается наименьшей. Напомним, что частным случаем кривой Неймана–Пирсона (при равенстве ковариационных матриц) является линейный дискриминант Фишера [6]. В нашем случае им нельзя пользоваться, так как эритроциты всегда намного однородней ядер в цветовом отношении, хотя их эллипсы рассеяния часто ориентированы почти параллельно.

Однако при построении разделяющей кривой Неймана–Пирсона по вновь поступающим на вход программы выборочным данным можно столкнуться с явлением неустойчивости: функциональный вид кривой (эллипс или гипербола) и ее положение на плоскости могут скачкообразно изменяться (обычно при малом объеме выборки для  $n < 10 \dots 15$ ). Возможный выход — замена теоретически оптимального решения (кривой второго порядка) на более устойчивое и простое решение — прямую линию. Точное построение этой прямой требует численного решения. В настоящее время используется приближенный способ. Разделяющей прямой объявляется касательная к линии Неймана–Пирсона. Эта касательная проводится в точке, в которой пересекаются линии Неймана–Пирсона и прямая, соединяющая центры двух разделяемых выборок. Подобное решение не является лучшим среди всех возможных прямых, но практически вполне удовлетворительно. Ошибки первого и второго рода определяются уже после проведения прямой. Приведем соответствующие формулы.

Исходным является пропорциональность оценок плотностей вероятностей:  $f_1 = cf_2$ , или в развернутом виде

$$\frac{1}{(2\pi)^{n/2} (\det \text{Cov}^{(1)})^{1/2}} \exp \left( -\frac{1}{2} a_{ij}^{(1)} (x_i - x_{i0}^{(1)}) (x_j - x_{j0}^{(1)}) \right) =$$

$$= \frac{c}{(2\pi)^{n/2} (\det \text{Cov}^{(2)})^{1/2}} \exp \left( -\frac{1}{2} a_{ij}^{(2)} (x_i - x_{i0}^{(2)}) (x_j - x_{j0}^{(2)}) \right), \quad (1)$$

где  $x_{i0}^{(1)}, x_{i0}^{(2)}$  — оценки средних, а  $a_{ij}^{(1)}$  и  $a_{ij}^{(2)}$  — оценки элементов обратных ковариационных матриц. В результате получаем уравнение поверхности 2-го порядка (для  $n = 2$  линии 2-го порядка). Это и есть уравнение разделяющей поверхности (линии) Неймана – Пирсона

$$a_{ij}^{(1)} (x_i - x_{i0}^{(1)}) (x_j - x_{j0}^{(1)}) - a_{ij}^{(2)} (x_i - x_{i0}^{(2)}) (x_j - x_{j0}^{(2)}) =$$

$$= \ln \left( \frac{\det \text{Cov}^{(2)}}{\det \text{Cov}^{(1)} \cdot c^2} \right). \quad (2)$$

Рассмотрим прямую линию, соединяющую центры выборок:  $x_i = x_{i0}^{(1)} + (x_{i0}^{(2)} - x_{i0}^{(1)}) t$ , где параметр  $t \in [0; 1]$ . Для  $t = 0$  мы находимся в центре первой выборки, а для  $t = 1$  — в центре второй выборки. Введем сокращенные обозначения:

$$l_i = x_{i0}^{(2)} - x_{i0}^{(1)} \neq 0; \quad C_0 = \ln \left( \frac{\det \text{Cov}^{(2)}}{\det \text{Cov}^{(1)} \cdot c^2} \right); \quad (3)$$

$$\alpha_i = a_{mn}^{(i)} l_m l_n > 0, \quad i = 1, 2.$$

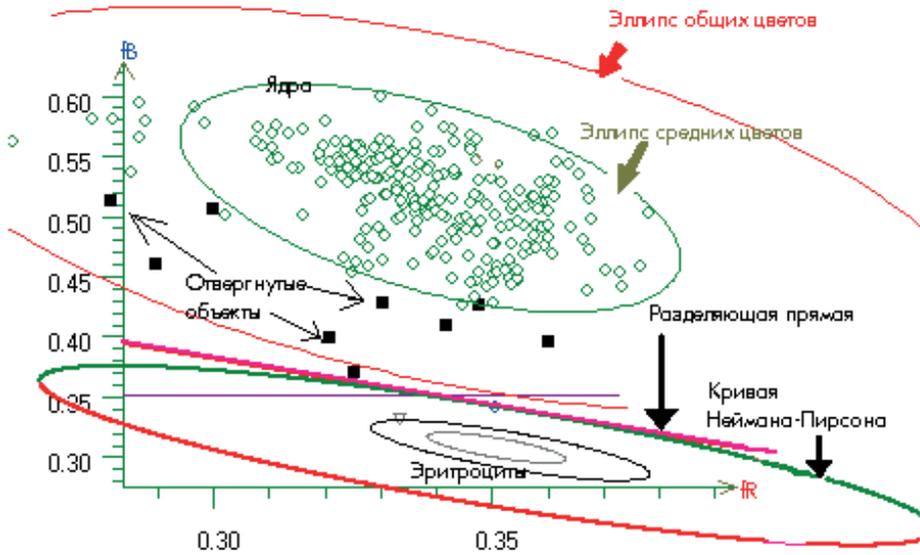
Тогда уравнение для поиска точек пересечения прямой и поверхности Неймана – Пирсона принимает вид

$$\alpha_1 t^2 - \alpha_2 (t - 1)^2 = C_0. \quad (4)$$

Для  $\alpha_1 > 0, \alpha_2 > 0$  в интервале  $[0; 1]$  это уравнение имеет единственный корень  $t^*$  при условии  $-\alpha_2 < C_0 < \alpha_1$ :

$$t^* = \frac{-\alpha_2 + \sqrt{\alpha_1 \alpha_2 + (\alpha_1 - \alpha_2) C_0}}{(\alpha_1 - \alpha_2)}; \quad -\alpha_2 < C_0 < \alpha_1. \quad (5)$$

Подставляя полученное значение в уравнение прямой линии, получаем точку, в которой проводится искомая касательная плоскость (для  $n = 2$  прямая линия):  $x_i^* = x_{i0}^{(1)} + l_i t^*$ . В случае совпадения ковариационных матриц знаменатель и числитель дроби равны нулю. Решением при этом является середина отрезка  $t^* = 0,5$  (линейный дискриминант Фишера).



**Рис. 7. Сравнение двух разделяющих линий на плоскости относительных цветов  $fR$ ,  $fB$ :**

линии Неймана–Пирсона (гипербола) и прямой, касательной к линии Неймана–Пирсона

На рис. 7 изображена ситуация, возникшая в начальной стадии обнаружения. Эллипс цветов атипично окрашенных (темно-коричневых) эритроцитов определен все еще недостаточно точно, а эллипсы цветов ядер еще хуже: накоплено всего 6 ядер, причем одно из них — верхнее правое, существенно отличается от остальных. Это вызвало сильную деформацию эллипса средних (причем не только в сторону “выброса”) и, возможно, эллипса общих цветов ядер. Тем не менее, даже для такой крайне неблагоприятной ситуации в существенной части плоскости (справа от правой ветви гиперболы) разделяющая прямая хорошо аппроксимирует гиперболу. Ошибки первого и второго рода в данном случае составляют десятки процентов, поэтому перехода к двумерной сегментации по относительным цветам не произойдет и продолжится накопление информации. Заметим, что на рис. 7 изображен крайний случай: разделяющая прямая проходит почти через центр эллипса цветов эритроцитов. При небольшом его смещении вверх прямую провести бы не удалось.

Более важным, чем вопрос точного построения разделяющей прямой, является выбор одного из двух возможных эллипсов рассеяния (общей статистики по всем пикселям всех объектов или только по средним значениям), который будет использоваться в качестве характеристики группы. Общая цветовая статистика для ядер кажется более неустойчивой. Для эритроцитов в силу их большей цветовой однород-

ности общая статистика достаточно стабильна и ее эллипс рассеяния обычно меньше по размерам, чем эллипс рассеяния средних цветов для ядер. Поэтому при построении линейного дискриминанта вводится асимметрия: используются разные эллипсы рассеяния – средних цветов для ядер и общих цветов для эритроцитов.

Полученная прямая будет иметь практическую ценность, если ее удастся использовать для сегментации пикселей, принадлежащих ядрам лейкоцитов в следующих кадрах. Это возможно при выполнении трех условий.

1. Эллипс общих цветов уже найденных ядер практически целиком лежит по одну сторону прямой ( $> 90\%$  всех уже отнесенных к ядрам пикселей).

2. Ошибка первого рода для средних цветов ядер очень мала ( $< 1\%$ ). Обычно это условие следует из предыдущего.

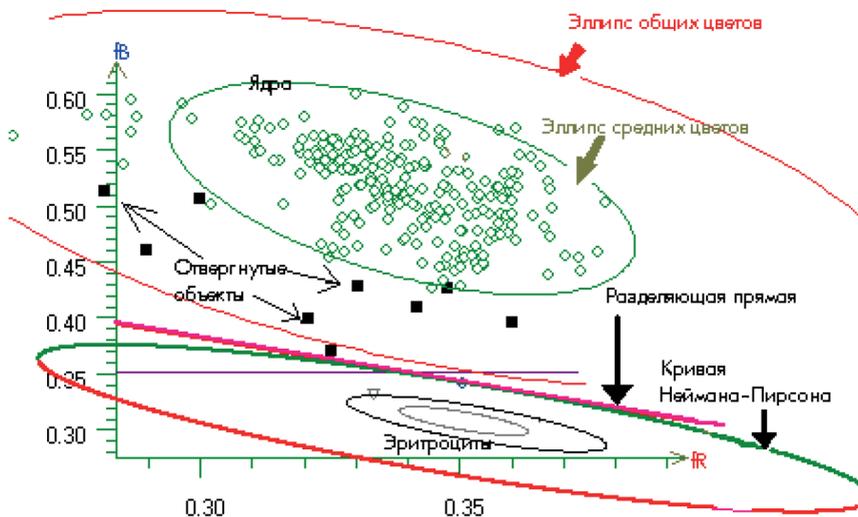
3. Ошибка второго рода мала ( $< 1\%$ ). Это условие обычно выполняется. Однако даже если бы оно и не было выполнено, то обычно пиксели оптически не слишком плотных эритроцитов отсекаются дополнительным условием по оптической плотности.

Критичным является первое условие. И если эллипс пикселей, принадлежащих ядрам, не деформирован за счет загрязнения артефактами (ложными ядрами), то оно выполнено практически для любой окраски. Хотя эритроциты и могут быть окрашены весьма по-разному (иметь розовый, бурый, зеленоватый или синеватый оттенок), они существенно отличаются по цвету от ядер лейкоцитов. Относительное различие в цвете между окрашенным гемоглобином эритроцитов и нуклеиновыми кислотами ядер лейкоцитов оказывается существенным всегда. Поэтому главным препятствием для устойчивого обнаружения ядер лейкоцитов является не вариабельность окраски мазка, а присутствие в нем артефактов.

**Типичные примеры разделения ядер, артефактов и эритроцитов на плоскости относительных цветов.** Далее приведены диаграммы, соответствующие трем типичным случаям разделения объектов на плоскости относительных цветов ( $fR, fB$ ).

1. Артефакты, плохо отделяющиеся от ядер. Возникают ложные тревоги и пропуски с выдачей соответствующего сообщения.

Этот случай приведен на рис. 8. Линия Неймана – Пирсона — это эллипс, верхняя и нижняя половины которого изображены зеленым и красным цветом. Для сегментации была использована разделяющая прямая (касательная к линии Неймана – Пирсона). Видно, что несмотря на хорошее разделение по цветам между эритроцитами и ядрами лейкоцитов (эллипс общих цветов ядер практически весь лежит по одну сторону от разделяющей прямой), имеется много сомнительных объектов (пятна краски и грязи). Средние относительные цвета отвергнутых объектов обозначены черными квадратами, а принятых в



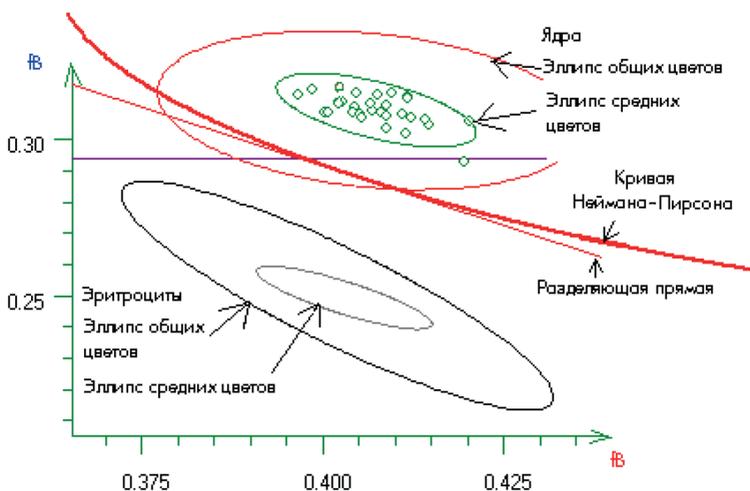
**Рис. 8. Разделение объектов на плоскости относительных цветов:**

артефакты, плохо отделяемые от истинных ядер (а); однородная выборка ядер, артефакты отсутствуют (б); артефакты отделены от объектов (в)

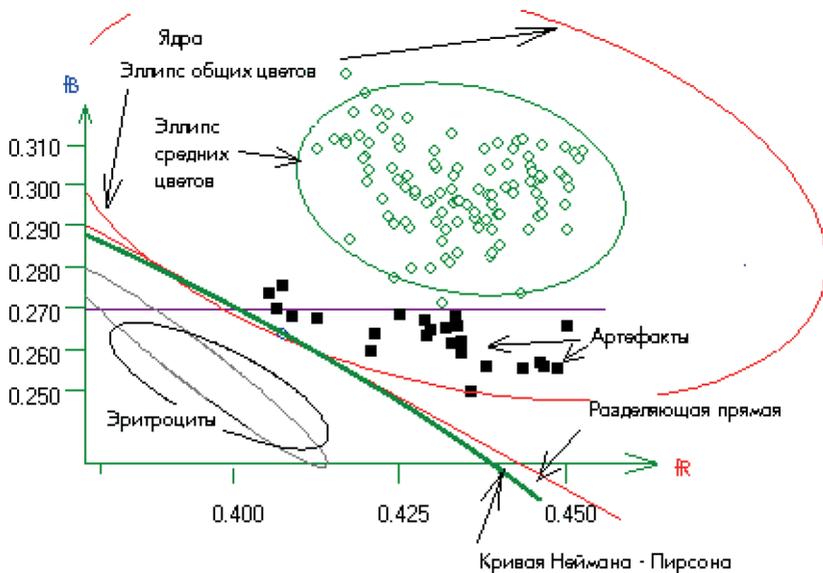
качестве ядер — зелеными кружками. Около 10 отвергнутых в процессе скрининга объектов фактически не отличаются от некоторых объектов принятых за ядра. Поэтому в результате было выдано сообщение о возможных пропусках и ложных тревогах.

2. Артефакты отсутствуют, выборка практически однородна, пропусков и ложных тревог нет.

Для случая, изображенного на рис. 9, артефакты на препарате отсутствуют, все объекты составляют однородную выборку и правильно квалифицированы как ядра — зеленые кружки соответствуют их сред-



**Рис. 9. Разделение объектов на плоскости относительных цветов: однородная выборка ядер, артефакты отсутствуют**



**Рис. 10. Разделение объектов на плоскости относительных цветов: артефакты отделены от объектов**

ним цветам. Отвергнутых объектов не было. Здесь, как и в предыдущем случае, проведение разделяющей линии в плоскости относительных цветов между эритроцитами и ядрами лейкоцитов не составляет труда. Горизонтальная линия соответствует разделению по одному синему цвету: качество разделения при этом было бы несколько хуже. Действительно, если бы для сегментации использовалась не наклонная прямая, а горизонтальная, то одно ядро (эозинофил), по которому проходит горизонтальная линия, возможно, было бы пропущено.

3. Артефакты, в цветовом отношении достаточно отличающиеся от истинных ядер. Программе удастся разделить истинные объекты и артефакты, пропусков и ложных тревог нет.

На рис. 10 видно, что благодаря однородности артефактов (имевших вид темных бурых пятен), все они (25 штук) были успешно отвергнуты программой обнаружения, и при окончательном контроле ни один из них не попал в диапазон цветовых характеристик ядер. Поэтому сообщения о возможных пропусках и ложных тревогах не было.

**Выводы.** В результате проведенного исследования были сформулированы требования к алгоритму обнаружения ядер лейкоцитов, рассмотрены проблемы, возникающие при обнаружении и сегментации клеток крови. Были выбраны яркостные и цветовые характеристики, положенные в основу алгоритмов обнаружения и сегментации. Также были рассмотрены и проанализированы алгоритмы поиска ядер лейкоцитов по гистограмме яркости, по абсолютным и относительным характеристикам. В результате работы можно сделать вывод,

что наиболее устойчивым методом обнаружения ядер клеток является двухпроходный адаптивный к цвету препарата алгоритм поиска. Для сегментации клеток были рассмотрены разделяющие алгоритмы, в их основе лежат относительные оптические плотности ядер и цитоплазмы различных типов клеток. В качестве разделяющей кривой использовалась кривая 2-го порядка Неймана – Пирсона. Для разделения также использовались цвето-яркостные характеристики и размеры объектов, что позволило успешно отделить артефакты от лейкоцитов и эритроцитов.

## ЛИТЕРАТУРА

1. Соколинский Б.З., Демьянов В.Л., Медовый В.С., Парпара А.А., Пятницкий А.М. Автоматическая сортировка лейкоцитов мазка крови с использованием методов обучаемых нейронных сетей и watershed // *Здравоохранение и медицинская техника*. 2005. № 4.
2. *Национальное руководство по лабораторной диагностике*. Т.1,2 / под ред. В.В. Долгова, В.В. Меньшикова.
3. *Обзор методик автоматизированной микроскопии биоматериалов* / В.С. Медовый, А.А. Парпара, А.М. Пятницкий, Б.З. Соколинский, В.Л. Демьянов // *Клиническая лабораторная диагностика* 2006. № 7. С. 15–20.
4. Medovyi V.S., Pyatnitskii A.M. Robotic Microscopy and information technology to increase accuracy, sensitivity and availability of blood cell analyses. Current microscopy contributions to advances in science and technology (Microscopy Book Series, Publisher: Formatex Research Center). Book 5. Vol. 1. P. 775–781, December 2012.
5. Albertini Maria C., Teodori Laura, Piatti Elena, Piacentini Maria P., Accorsi Augusto, Rocchi Marco B.L. Automated analysis of morphometric parameters for accurate definition of erythrocyte cell shape Cytometry. Part A. 52A:12-18, 2003.
6. Прэнтт У. Цифровая обработка изображений: Пер. с англ. М.: Мир, 1982. Кн. 1. 312 с.
7. Пантелеев И., Егорова О., Клыкова Е. Компьютерная микроскопия. Техносфера, Москва, 2005.
8. Bikhet S.F., Darwish A.M., Tolba H.A., Shaheen S.I. Segmentation and classification of white blood cells. Acoustics, Speech, and Signal Processing, 2000. ICASSP'00. Proceedings. 2000. IEEE International Conference on Volume 6, 2000. Vol. 4. P. 2259–2261.
9. Jiang K., Liao Q., Dai S. A novel white blood cell segmentation scheme using scale-space filtering and watershed clustering. Proc. Of the Second Intern. Conf. on Machine Learning and Cybernetics. Xi'an, 2–5 November 2003.
10. Анисимов Б.В., Курганов В.Д., Злобин В.К. Распознавание и цифровая обработка изображений. М.: Высш. шк., 1983. 295 с.

## REFERENCES

- [1] Sokolinskiy B.Z., Dem'yanov V.L., Medovyy V.S., Parpara A.A., Pyatnickiy A.M. Automated Sorting Leukocytes in a Blood Smear Using Techniques of Educable Neural Networks and Watershed. *Zdravoohranenie i medicinskaya tehnika* [Health Service and Medical Technology], 2005, no. 4 (in Russ.).
- [2] Dolgov V.V., Menshikov V.V., eds. *Nacionalnoe rukovodstvo po laboratornoy diagnostike* [National Guidelines for Laboratory Diagnostics], vol. 1, 2.

- [3] Medovyy V.S., Parpara A.A., Pyatnickiy A.M., Sokolinskiy B.Z., Dem'yanov V.L. Review of Methods of Biomaterial Automated Microscopy. *Klinicheskaya laboratornaya diagnostika* [Clinical Laboratory Diagnostics], 2006, no. 7, pp. 15–20 (in Russ.).
- [4] Medovyy V.S., Pyatnitskiy A.M. Robotic Microscopy and information technology to increase accuracy, sensitivity and availability of blood cell analyses. Current microscopy contributions to advances in science and technology (Microscopy Book Series, Publisher: Formatex Research Center), book 5, vol. 1, pp. 775–781, December 2012.
- [5] Albertini M.C., Teodori L., Piatti E., Piacentini M.P., Accorsi A., Rocchi M.B.L. Automated analysis of morphometric parameters for accurate definition of erythrocyte cell shape Cytometry. Part A. 52A:12-18, 2003.
- [6] Pratt W.K. Digital Image Processing. N.Y.–Chichester–Brisbane–Torc. John Wiley & Sons, 1978.
- [7] Panteleev I., Egorova O., Klykova E. Komp'yuternaya mikroskopiya [Computer Microscopy]. Moscow, Tekhnosfera Publ., 2005.
- [8] Bikheth S.F., Darwish A.M., Tolba H.A., Shaheen S.I. Segmentation and classification of white blood cells. Acoustics, Speech, and Signal Processing, 2000. ICASSP'00. Proceedings. 2000 IEEE International Conference. On vol. 6, 2000, vol. 4, pp. 2259–2261.
- [9] Jiang K., Liao Q., Dai S. A novel white blood cell segmentation scheme using scale-space filtering and watershed clustering. *Proc. of the Second Intern. Conf. on Machine Learning and Cybernetics*. Xi'an, 2–5 November 2003.
- [10] Anisimov B.V., Kurganov V.D., Zlobin V.K. Raspoznavanie i tsifrovaya obrabotka izobrazheniy [Image Recognition and Digital Processing]. Moscow, Vyssh. shk. Publ., 1983. 295 p.

Статья поступила в редакцию 8.10.2014

Шахтарин Борис Ильич — д-р техн. наук, профессор кафедры “Автономные информационные и управляющие системы” МГТУ им. Н.Э. Баумана. Автор более 250 научных работ в области радиотехники, статистического анализа, фазовой синхронизации, формирования и обнаружения сигналов.

МГТУ им. Н.Э. Баумана, Российская Федерация, 105005, Москва, 2-я Бауманская ул., д. 5.

Shakhtarin B.I. — D.Sc. (Eng.), Professor of Engineering, Department of Autonomous Information and Control Systems, Bauman Moscow State Technical University, author of over 250 publications in the field of radio engineering, statistical analysis, phase synchronization, formation and detection of signals.

Bauman Moscow State Technical University, 2-ya Baumanskaya ul. 5, Moscow, 105005 Russian Federation.

Панов Степан Александрович — аспирант кафедры “Автономные информационные управляющие системы” МГТУ им. Н.Э. Баумана, заместитель генерального директора ООО “МЕКОС”.

МГТУ им. Н.Э. Баумана, Российская Федерация, 105005, Москва, 2-я Бауманская ул., д. 5.

Panov S.A. — Ph.D. student, Department of Autonomous Information and Control Systems, Bauman Moscow State Technical University, Deputy Director of OAO MECOS. Bauman Moscow State Technical University, 2-ya Baumanskaya ul. 5, Moscow, 105005 Russian Federation.

Калашников Константин Сергеевич — канд. техн. наук, инженер ООО “НПП “Цифровые решения”. Автор более 25 научных работ в области приема и передачи сигналов цифровой радиосвязи, разработки алгоритмов синхронизации и оценки искажений сигналов.

ООО “НПП “Цифровые решения”, Российская Федерация, 111250, Москва, Красноказарменная ул., д. 15.

Kalashnikov K.S. — Ph.D. (Eng.), engineer, ООО Scientific Production Enterprise “Digital Solutions”, author of over 25 research publications in the fields of receiving and transmission of digital radio signals, algorithm development of synchronization and estimation of signal distortion.

ООО Scientific Production Enterprise “Digital Solutions”, Krasnokazarmennaya ul. 15, Moscow, 111250 Russian Federation.

**Просьба ссылаться на эту статью следующим образом:**

Шахтарин Б.И., Панов С.А., Калашников К.С. Алгоритмы распознавания клеток крови // Вестник МГТУ им. Н.Э. Баумана. Сер. Приборостроение. 2015. № 4. С. 49–65.

**Please cite this article in English as:**

Shakhtarin B.I., Panov S.A., Kalashnikov K.S. Algorithms of blood cells recognition. *Vestn. Mosk. Gos. Tekh. Univ. im. N.E. Baumana, Priborostr.* [Herald of the Bauman Moscow State Tech. Univ., Instrum. Eng.], 2015, no. 4, pp. 49–65.