

Многомерный анализ изображений в медицине и биологии

Беляев И.А.¹, Кучерявский С.В.¹, Родионова О.Е.², Померанцев А.Л.²

¹Алтайский государственный университет

²Институт химической физики РАН

Дан развернутый обзор методов получения, обработки, анализа и классификации изображений в медицине и биологии. Предложен общий подход, основанный на применении методов анализа многомерных данных. Приведены результаты использования данного подхода для решения задач обнаружения и корректировки неправильной сегментации биочипов и классификации белых клеток крови.

Введение

Получение и анализ изображений является важным инструментом в биомедицинских исследованиях и диагностике. При этом класс задач, которые могут быть решены при помощи этих методов, достаточно широк и включает в себя такие важные проблемы, как определение раковых клеток, распознавание клеток крови и костного мозга, выявление нарушений тканей и так далее. Как правило, объекты исследования на медицинских изображениях имеют резко неоднородную, сложную структуру и требуют от специалистов большого опыта для выявления тех или иных особенностей, позволяющих поставить определенный диагноз. Использование вычислительной техники и математических методов в этой отрасли позволяет автоматизировать этот процесс, повысить точность результатов исследований, а также провести сопоставление между особенностями изображений и результатами других методов исследования (биохимических, биофизических, физиологических и прочих), сделать результаты исследований более объективными. Кроме того, ввод изображений в компьютер не только позволяет хранить и сортировать, но и обмениваться ими с другими исследователями, используя для этого современные средства компьютерных телекоммуникаций.

Существенную роль в выявлении особенностей изображений, а так же их сопоставлении с результатами других методов исследования могут сыграть методы и подходы разработанные в такой смежной области как хеометрика. Хеометрика — это синтетическая дисциплина, находящаяся на стыке химии и математики [1]. По определению Д. Массарту (D. Massart) [2]: — это химическая дисциплина, применяющая математические, статистические и другие методы, основанные на формальной логике, для построения или отбора оптимальных методов измерения и планов эксперимента, а также для извлечения наиболее важной информации при анализе экспериментальных данных. Важной особенностью хеометрического подхода является рассмотрение многофакторных данных в совокупности, без выделения предварительно одного или

нескольких «существенных» факторов или переменных. В результате построенные многомерные эмпирические модели позволяют обнаружить новые связи или явления, так как учитывают скрытые объективно существующие связи внутри объекта.

В настоящее время задача получения качественных изображений биологических объектов не представляет научного интереса в связи с широким распространением систем получения цифровых снимков, и особое внимание следует уделять методам их математической обработки и анализа.

Одним из наиболее важных этапов здесь является предварительная обработка изображений, которая позволяет повысить их качество, а также изменить некоторые характеристики таким образом, чтобы при количественном анализе изображений выделение интересующих объектов в дальнейшем могло бы быть произведено наилучшим образом.

Для преобразования изображений используются различные методы фильтрации, которые можно условно разделить на четыре группы [3, 4].

1. Изменение яркости и контрастности изображений. При использовании этих методов преобразования меняются яркостные характеристики изображения, что отражается на гистограмме яркости.
2. Фильтры выравнивания, обеспечивающие выравнивание изображений с точки зрения яркости и цвета, что необходимо в тех случаях, когда, например, имеются дефекты осветительной системы или оптики регистрирующей аппаратуры, выражающиеся в том, что фон изображений становится неравномерным.
3. Фильтры сглаживания, обеспечивающие очистку изображений от шумов.
4. Фильтры детализации, обеспечивающие пограничные преобразования, что может выражаться (в зависимости от примененного фильтра) в усилении резкости, выделении границ объектов, детализации объектов и т.п.

К настоящему времени математический аппарат теории фильтрации изображений проработан достаточно глубоко, так что задача исследователя сводится лишь к определению искажающих факторов изображения и подбору наиболее подходящего способа его улучшения. Однако, по возможности, следует стремиться к качественным снимкам уже на этапе их получения.

Для последующего изучения объектов изображения их выделяют относительно фона или посторонних элементов. Зачастую выбор способа выделения объектов зависит

как от способа получения изображений, так и от типа объекта исследования. Заключительным этапом является собственно анализ выделенных объектов, их классификация.

Спектр задач медицинских и биологических исследований, решаемых с применением анализа изображений, охватывает все уровни организации органической материи: молекулярный, клеточный, тканевый, органнй, организменный, популяционный.

Живые организмы устроены крайне сложно и содержат большое количество взаимодействующих систем. Основную роль в управлении жизнедеятельностью играют гены — участки молекулы ДНК, в которых хранится информация об устройстве молекул, вовлеченных в различные процессы в живой клетке.

Биологам и медикам необходимо знать реакцию больших каскадов взаимозависимых и взаимообуславливающих генов на то или иное изменение внешних условий, например в ответ на введенное лекарство.

До последнего времени в значительной степени отсутствовали возможности для получения, хранения и обработки столь значительных массивов данных. Благодаря прогрессу компьютерной индустрии были созданы как технологии для одновременного экспериментального получения информации о работе большого числа генов в клетке, так и методы обработки этой информации, позволяющие сделать на ее основе простые и однозначные выводы (например, поставить точный диагноз какого-либо заболевания). Современная экспериментальная техника позволяет создать анализирующую матрицу (называемую также биочипом) размером несколько сантиметров, при помощи которой можно получить данные о состоянии всех генов организма.

Применению анализа изображений в технологии биочипов посвящено много работ, например [5–9]. Сюда входят задачи автоматического выделения из шума сигналов разной интенсивности (фильтрация), выделения элементарных ячеек анализирующей матрицы (сегментация) и их кластеризация. Кластеры генов, ведущих себя схожим образом в разных условиях или в разные моменты времени, служат исходной точкой для заключений биологического характера. Анализ функций генов является одним из направлений цитогенетических исследований.

Цитогенетика — наука, изучающая строение хромосом — ведет свое начало с 1956г., когда было доказано, что в человеческих клетках (кроме половых) содержится 46 хромосом-носителей генетической информации. В конце 60-х — начале 70-х годов были

разработаны методы, позволяющие однозначно различить каждую хромосому и выявить ее тонкое строение, после чего выяснилось, что множество тяжелых врожденных заболеваний определяется нарушениями хромосом или их количества в клетке. Исследование количества и строения хромосом называют кариотипированием.

В процессе проведения кариотипирования вначале исследуемые клетки (лимфоциты крови, клетки ворсинок хориона, клетки, содержащиеся в амниотической жидкости и др.) специальным образом культивируются, затем обрабатываются с тем, чтобы получить максимальное количество делящихся клеток, находящихся в стадии метафазы митоза, когда хромосомы достигают оптимальных размеров для исследования под микроскопом. После специального окрашивания, препараты анализируются, и производится раскладывание хромосом по порядку, определяемому международной классификацией.

Еще один метод, относящийся к кариотипированию — *FISH (Fluorescence In Situ Hybridization)*, позволяющий выявлять тонкое строение отдельных районов определенных хромосом, используется в случаях, когда необходима более тонкая оценка строения конкретного участка данной хромосомы.

Цифровой анализ изображений позволяет полнее определять хромосомные особенности ДНК в метафазе и интерфазе для каждой клетки в колонии [10]. В настоящее время разработаны и внедряются в практику методы, позволяющие окрашивать каждую хромосому в свой цвет, что позволяет автоматизировать хромосомный анализ [11]. Существуют программные комплексы, позволяющие во многих случаях проводить автоматическое кариотипирование. Например, программные комплексы производства фирм «Applied Imaging» и «Видеотест».

Одним из наиболее распространенных и мощных инструментов биологических исследований является оптическая микроскопия. Начиная с 70-х годов прошлого века, предпринимались попытки компьютерного анализа изображений поля зрения микроскопа. Получивший распространение термин «автоматизированная микроскопия» (*Automated microscopy*) соответствует системе микроскопии, в которой изображение поля зрения преобразуется в цифровую форму, позволяющую использовать компьютер для автоматизации микроскопического анализа и управления микроскопом [12, 13]. Однако создание полностью автоматизированных систем микроскопии стало возможно лишь в последнее десятилетие благодаря бурному прогрессу в области вычислительных средств, техники видеосъемки, математических методов анализа.

Одна из главных задач построения систем автоматизированной микроскопии – выделение объектов на изображениях. Так, например, выделение гистологических объектов на цифровых изображениях затруднено их существенной вариабельностью и слабой контрастностью [12]. Кроме того, границы самих микрообъектов и границы их внутренних элементов изначально, по своей природе, нечеткие. В настоящее время предложено большое число подходов к решению этой проблемы. Предлагаются методы сегментации с использованием различных фильтров-детекторов границ [14], сегментация методом k -средних [15], использование искусственных нейронных сетей [16], метод роста регионов [17] и другие [18, 19]. Универсальные подходы к выбору алгоритма для сегментации произвольного изображения гистологических объектов неизвестны.

Для построения вектора признаков, используемого для классификации обнаруженных объектов, применяют как морфологические признаки, так и признаки, основанные на текстурных и цветовых особенностях. В качестве классификаторов используются разного рода нейронные сети [20, 21].

Некоторые из упомянутых методов получили реализацию в так называемых автономных системах автоматизированной микроскопии (АСАМ) [12]. АСАМ автоматически перемещает и фокусирует препарат, выбирает траекторию просмотра в зависимости от распределения биоматериала, контролирует качество освещения и окраски, обнаруживает и записывает в базу данных изображения объектов заданных типов. Выполняется измерение и анализ автоматически собранной выборки объектов. Кроме автоматизированного микроскопа и системы анализа изображений в состав АСАМ входит роботизирующий программный компонент, управляющий автоматизированным микроскопом и заменяющий при перемещении препарата глаза и руки лаборанта. Примерами таких систем являются:

- *DiffMaster Octavia* (Швеция) для автоматического подсчета лейкоцитарной формулы по мазку крови. В системе используется система классификации на основе обучающихся нейронных сетей;
- *Rare Event Imaging System* (REIS, США) для автоматического обнаружения инфицированных цитомегаловирусом лейкоцитов периферической крови. Используются иммунофлюоресцентные метки *lower matrix phosphoprotein* (pp65). Морфометрические признаки помеченных лейкоцитов используются в автоматической диагностической системе;

- *IQ-200* (IRIS, США) для автоматического анализа осадка мочи. Система обнаруживает 12 элементов мочи — эритроциты, лейкоциты, лейкоцитарные сгустки, бактерии, грибки, гиалиновые цилиндры, другие патологические цилиндры, клетки плоского эпителия, другие эпителиальные клетки, слизь, сперматозоиды и артефакты;
- *МЕКОС-ЦИ* (МЕКОС, Россия) для автоматического подсчета лейкоцитарной формулы и количественного анализа эритроцитов и ретикулоцитов. В системе используется система классификации на основе обучающихся нейронных сетей.

Изучение структур и функционирования отдельных органов и внутренних систем организма, не нарушая его целостности, стало возможным благодаря появлению рентгенологии. Началом развития мировой рентгенологии принято считать осень 1895 года, когда немецкий физик Вильгельм Конрад Рентген открыл еще неизвестные науке лучи. Заслуга Рентгена заключается не только в открытии неизвестного проникающего излучения, но и в открытии методов рентгенодиагностики. В настоящее время рентгенологию рассматривают как одно из направлений лучевой диагностики.

Лучевая диагностика непосредственно связана с обработкой и анализом изображений. Прямой компьютерный анализ таких изображений затруднен их низкой контрастностью, наличием геометрических искажений, размытостью. В сложнейших по структуре изображениях биологических объектов обилие наложенных друг на друга теней различных органов ухудшает субъективное восприятие деталей малых контрастов в несколько раз (т.н. краудинг-эффект) [22].

Сегментации изображений лучевой диагностики посвящено большое количество исследований [23–27] и др. Задача выделения интересующего объекта может рассматриваться как отдельно, так и рамках более сложных проблем. Так, например, работа [27] посвящена способу получения точных контуров костей на рентгеновских снимках, основанном на вычислении робастной статистики участков изображения. Этот метод получил реализацию в компьютеризированной ортопедической системе устранения смещения фрагментов костей *Fracas*. Разработанная система позволяет соотнести статическое изображение с пространственной моделью фрагмента кости, полученной на основе дооперационной компьютерной томографии и отслеживать его положение во время операции в режиме реального времени.

Задачам определения границ и положения объектов на рентгенограммах посвящена работа [28], где описывается метод автоматического определения положения и

ориентации позвоночника на снимке. Кроме того, получаемые в результате контуры тел позвонков можно использовать в качестве признаков при построении системы автоматической диагностики патологий позвоночника.

Методы анализа многомерных данных

Классический подход, применяемый для анализа медицинских и биологических изображений, может быть схематично представлен в виде последовательности, состоящей из четырех этапов: получение изображения, предварительная обработка, получение вектора признаков и анализ полученных признаков. В качестве признаков наиболее часто используют морфологические и цветовые (яркостные) характеристики изучаемых объектов, а для анализа широкое распространение на сегодняшний момент получили искусственные нейронные сети. Однако не всегда такая комбинация позволяет получить оптимальный результат.

В первую очередь это связано с проблемой поиска релевантных признаков в каждом конкретном случае. Не всегда такие параметры, как морфология и распределение интенсивности непосредственно связаны с изучаемым явлением, а использование нейронных сетей предполагает существенные временные и вычислительные затраты, особенно, если обучающая выборка достаточно велика [29]. Кроме этого, для нейронных сетей характерны проблемы интерпретируемости и размерности. Проблема интерпретируемости заключается в том, что полученную модель (в частности, весовые коэффициенты) практически невозможно проанализировать. Это приводит к риску получить заведомо неправильный результат, например, из-за наличия выбросов в данных. Кроме этого для нейронных сетей характерно очень жесткое ограничение на количество выходных нейронов в сети, на количество рецепторов и на сложность структуры взаимосвязей нейронов. Достаточно сказать, что количество выходных нейронов в реальных нейронных сетях, реализуемых на базе известных программных пакетов, обычно не превышает несколько сотен, а чаще всего составляет единицы или десятки [30].

Этих проблем можно избежать, позаимствовав для анализа методы широко используемые в хемометрике. Это метод главных компонент (МГК) [31] и метод проекций на латентные структуры (ПЛС) [32]. Основная идея этих методов — замена исходные данные их проекцией на подпространство в системе координат исходных переменных, которое ориентировано таким образом, что позволяет оставить только те вариации, которые непосредственно связаны с рассматриваемой проблемой. Такое преобразование носит название факторной компрессией, а векторы, которые задают ориентацию данного подпространства — факторами. Факторная компрессия делит исходные данные, а точнее

их вариацию, на две части — структурную часть, которая непосредственно влияет на требуемый результат и шум — вариацию данных, которая никак с ним не связана. При этом в большинстве случаев, размерность такого подпространства существенно ниже размерности исходного пространства.

Для оптимизации факторного подпространства, в частности определения его размерности и ориентации, метод главных компонент использует принцип наибольшей протяженности [31]. Так первый вектор, определяющий факторное подпространство, направлен вдоль наибольшей вариации данных в исходном пространстве (первая главная компонента). Вторым — перпендикулярен первому и так же ориентирован вдоль второго по значению направления вариации данных (вторая главная компонента) и т.д.

Помимо понижения размерности и выделения структурной части в данных, МГК позволяет провести предварительный анализ, выявить возможные выбросы или нетипичные образцы, определить имеется ли среди образцов тенденция к образованию отдельных групп и т.д.

Для непосредственно установления взаимозависимостей между исходными данными (векторами признаков изучаемых изображений) и интересующими нас характеристиками объектов, представленных на них, наилучшие результаты дает метод проекций на латентные структуры (ПЛС) [32]. При оптимизации факторного подпространства ПЛС учитывает не только вариацию исходных данных ($\mathbf{X}$), но и вариацию переменных, значения которых нам необходимо предсказывать ($\mathbf{Y}$), моделируя только ту структурную часть $\mathbf{X}$, которая непосредственно коррелирует с $\mathbf{Y}$. С точки зрения математической статистики, ПЛС максимизирует ковариацию между $\mathbf{X}$ и $\mathbf{Y}$.

Одним из наиболее наглядных инструментов и МГК и ПЛС являются графики счетов и нагрузок. График счетов показывает расположение проекций образцов на факторное подпространство (пространство главных компонент). Так, например, близость двух точек на графике счетов означает схожесть характеристик соответствующих им объектов, т.е. положительную корреляцию между ними. Точки, удаленные вдоль разных осей (образующие прямой угол относительно начала координат), являются некоррелированными, а расположенные диаметрально противоположно — имеют отрицательную корреляцию.

График нагрузок показывает, каким образом исходные переменные влияют на положение пространства главных компонент и, соответственно, на положение проекций образцов в этом пространстве. Анализируя его аналогично графику счетов, можно понять, какие переменные связаны, а какие независимы.

Многомерный анализ изображений

Один из способов построения вектора признаков изображения для анализа с помощью МГК и ПЛС состоит в использовании цветовых компонент пикселей изображения. Многомерный анализ изображений (*Multivariate Image Analysis — MIA*) самый известный пример такого подхода [33]. В MIA изображение рассматривается как трехмерная $W \times H \times K$ матрица, где W и H — ширина и высота изображения, а K — это значение переменной, описывающей каждый пиксель изображения. В простейшем случае этой переменной является интенсивность цветового компонента пикселя (значения R , G и B), но MIA наиболее эффективен применительно к гиперспектральным изображениям, т.е. каждый пиксель имеет интенсивность в различной полосе длин волн [34]. Такие вспомогательные изображения собирают, используя, например, мультиспектральный сканер.

Гиперспектральные изображения также могут получаться в результате аналитического эксперимента в лаборатории, при использовании спектрометрических методов исследования (таких как БИК-спектроскопия, масс-ионная спектрометрия и т.д.). Количество длин волн a , следовательно, длина K -вектора для каждого пикселя в этом случае будет изменяться от десятков до сотен. Гиперспектральные изображения часто применяются в биологии и медицине, например, для определения поддельных лекарств [35].

Однако многомерный анализ изображений может успешно использоваться даже для анализа изображений, представленных в градациях серого. Проиллюстрируем это на примере решения задачи анализа изображений, получаемых с помощью биочипов.



Рис. 1. Пример изображения биочипа, содержащего 480 элементарных ячеек.

На Рис. 1 показан пример биочипа, включающего $16 \times 30 = 480$ элементарных ячеек. Анализ такого изображения начинается с сегментации — разделения его на элементы, соответствующие одной ячейке. Для этого на изображение биочипа накладывается прямоугольная маска, выделяющая отдельные элементы. Однако по самым разным причинам изображение некоторых пятен может оказаться сдвинутым относительно центра ячейки. На Рис. 2 показаны два примера: идеальный случай позиционирования (а) и случай, в котором пятно сдвинуто влево и вверх, так, что в маску попадают даже два кусочка от соседних ячеек (б). Очевидно, что неправильная сегментация ведет к некорректной интерпретации изображений биочипов.

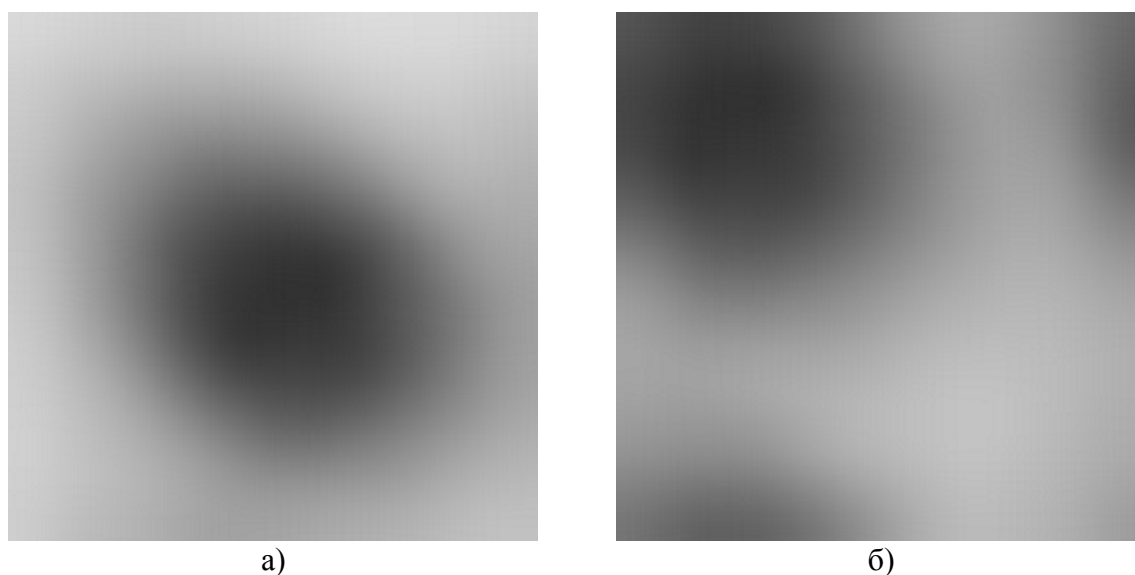


Рис. 2. Изображение правильно выделенного элементарного элемента (а) и смещенного, относительно центра (б).

Целью исследования явилось автоматическое определение неверной сегментации и ее корректировка. Для этого изображение каждого пятна было оцифровано с разбиением на $14 \times 14 = 196$ точек, охарактеризованных интенсивностью — величиной градации серого цвета, представляемой числом от 0 (черный цвет) до 255 (белый цвет). В результате такой оцифровки была получена матрица свойств X размерностью $I = 480$ строк (изображения элементарных ячеек) на $J = 196$ столбцов (последовательность градаций серого цвета пикселей этих изображений).

Анализ полученной матрицы проводился методом главных компонент, включающим предварительное центрирование и шкалирование [32]. Результаты анализа показали, что четыре главных компоненты объясняют 87% вариации в данных. Для интерпретации смысловой нагрузки первой главной компоненты была построена зависимость средней интенсивности каждого пятна от величины ГК1 (Рис. 3а).

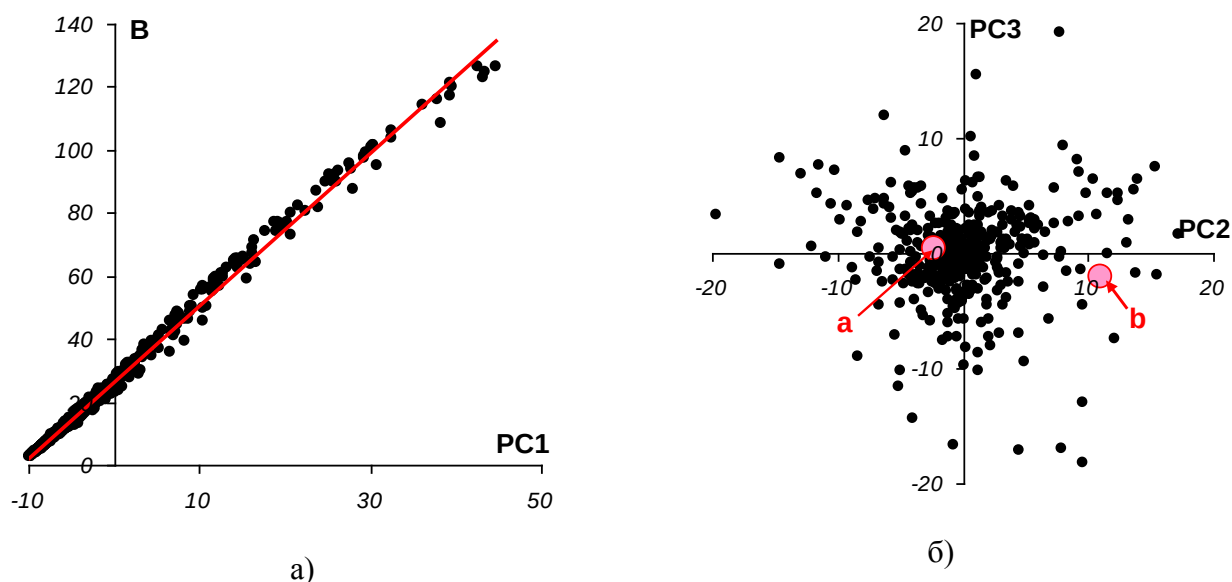


Рис. 3. Зависимость средней интенсивности яркости элемента от значения ГК1 (а).

График счетов ГК2–ГК3 (б).

Очевидно, что первая компонента объясняет именно эту характеристику элемента. График счетов ГК2–ГК3, на котором отмечены счета, соответствующие двум элементам представленным на Рис. 2а и Рис. 2б, показывает, что вторая и третья главная компонента отвечает за положение пятна относительно маски. Дальнейший анализ графиков счетов и нагрузок позволил установить окончательные закономерности:

1. Первая главная компонента отвечает за среднюю интенсивность элемента.
2. Вторая компонента отвечает за смещение центра пятна в направлении *NW-SO*, а третья — за смещение в направлении *W-O*.
3. Четвертая главная компонента отвечает за структуру пятна, т.е. за распределение интенсивности яркости внутри элемента.

Таким образом, метод многомерного анализа изображений дал возможность построить алгоритм, который выявляет случаи неверной сегментации биочипов, и разработать процедуру корректировки, состоящую в сдвиге маски в направлении, определяемом величинами второй и третьей главной компонент.

Метод среднеугловых спектров

Не смотря на то, что применение МГК и ПЛС непосредственно к изображениям очень часто позволяет получить требуемый результат, иногда необходимо построение отдельного вектора признаков изображения при помощи специальных методов (таких как статистический анализ, текстурный анализ, преобразование Фурье и т.д.). Такой вектор позволяет получать различные интегральные, масштабно-зависимые и другие типы

признаков изображения. Методы получения этих характеристик быстры и нетребовательны к ресурсам, что делает привлекательными их применение в медицинской диагностике.

Одним из таких методов, позволяющим, в частности, проводить классификацию клеток крови на низкоконтрастных и зашумленных изображениях является метод среднеугловых спектров (*Angle Measure Technique — AMT*) [36]. Метод АМТ был разработан в 1994 году Робертом Андерле для описания сложных геоморфных линий, в частности русел рек, границы береговой линии и т.п. Он преобразует изображение в одномерный вектор, характеризующий изменение типичных масштабов объектов на изображении.

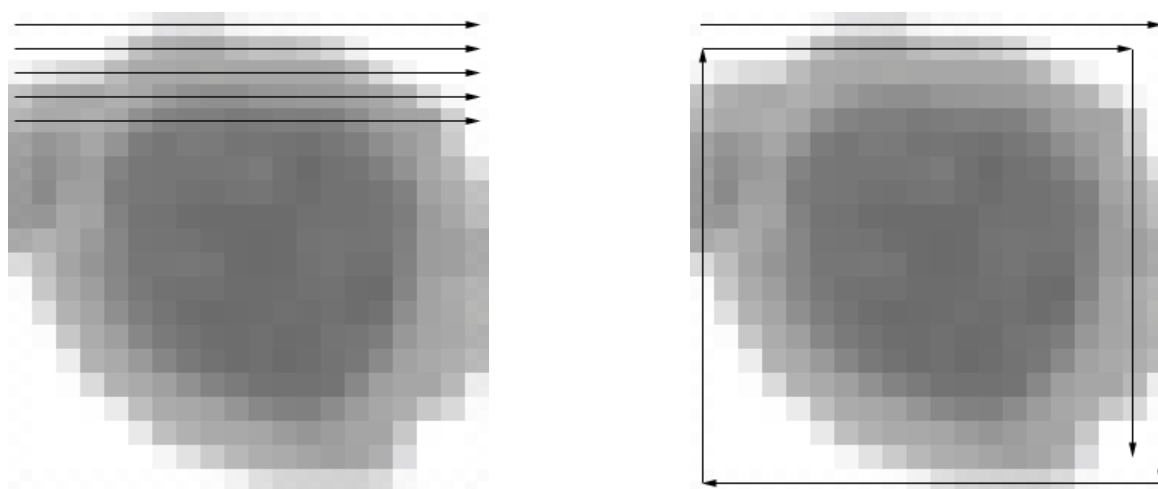


Рис. 4. Схема построчной (а) и спиральной (б) развертки изображений.

Сам процесс преобразования состоит из двух этапов. На первой стадии исходное изображение разворачивается в одномерный сигнал или профиль. При этом могут быть использованы разные способы развертки (Рис. 4). В простейшем случае, применяется построчная развертка, которая используется в большинстве случаев при анализе текстур или других гомогенных изображений.

После выполнения процедуры развертки вдоль полученного одномерного сигнала случайным образом выбираются точки, число которых составляет 5–10% от всех точек изображения. Затем строятся окружности с центрами в этих точках и радиусом s . На следующем шаге находят точки пересечения окружности и профиля. Схема процедуры показана на Рис. 5. Здесь A — случайно выбранная на профиле точка, C и B — найденные точки пересечения профиля и окружности радиуса s . Вычисляется угол, дополняющий CAB , после чего сохраняется среднее значение найденного угла для всех точек. Среднее

значение угла отражает особенности профиля и, следовательно, изображения на масштабе, совпадающем с радиусом s . Повторяя эти вычисления для различных масштабов, строится сложный спектр, который может рассматриваться как вектор признаков изображения на наборе масштабов.

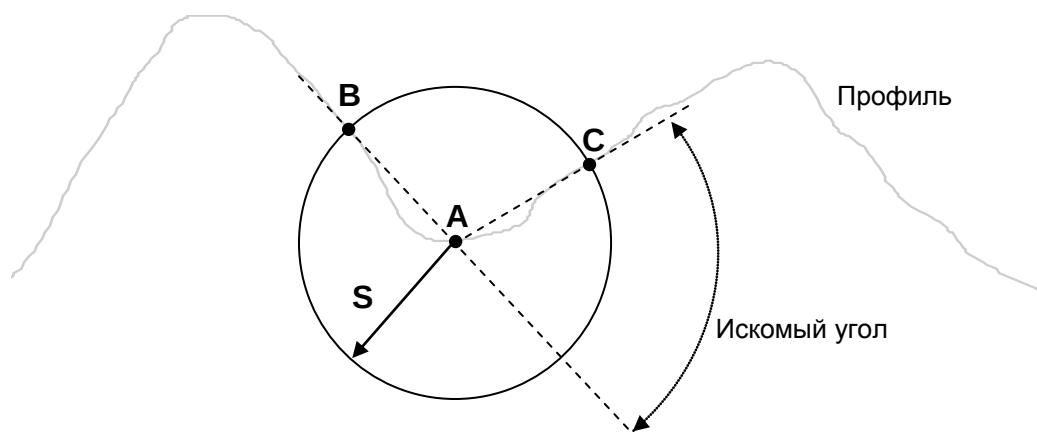


Рис. 5. Схема нахождения среднеуглового спектра

Изначально Андрле использовал для анализа изображений графическое представление полученных спектров. Наиболее важной информацией здесь является присутствие точек перегиба, пиков, локальных экстремумов и других особенностей на спектральной кривой. Но во многих случаях визуальный анализ АМТ-спектров не дает прийти к требуемым результатам, потому что кривые перекрываются. В этом случае наилучший результат дает применение для анализа и классификации МГК или ПЛС к среднеугловым спектрам. Рассмотрим один из примеров такой комбинации, которая была использована для решения задачи распознавания клеток крови [37].

Как известно, кровь состоит из плазмы и клеток, которые делятся на белые и красные клетки. Среди белых клеток принято выделять такие, как лимфоциты, нейтрофилы, моноциты и т.п. Используя хорошее оборудование и реактивы для приготовления мазков крови, можно получать высококачественные изображения, где кровяные клетки могут очень хорошо различаться, что и используется в дорогостоящих программно-аппаратных комплексах. Однако качество изображений, получаемых при помощи обычного оптического микроскопа и недорогой цифровой камеры, весьма далеко от требований коммерческих программ распознавания клеток.

Именно такие изображения, полученные при помощи оптического микроскопа и цифровой VGA камеры, стали предметом эксперимента с использованием метода среднеугловых спектров. В качестве предварительной обработки использовалась кластеризация в цветовом пространстве *Lab* для сегментации белых кровяных клеток из исходных изображений. В процессе сегментации клетки помещались на изображение с белым фоном и разрешением 128×128 точек (Рис. 6).



Рис. 6. Изображения клеток крови после сегментации

Все изображения были разделены случайным образом на обучающий и тестовый наборы. Обучающий набор состоял из 60 образцов лимфоцитов и нейтрофилов, которые были взяты от разных пациентов в разные дни, а тестовый набор состоял из 96 образцов, взятых в другие дни и от других людей, в отличие от калибровочного набора.

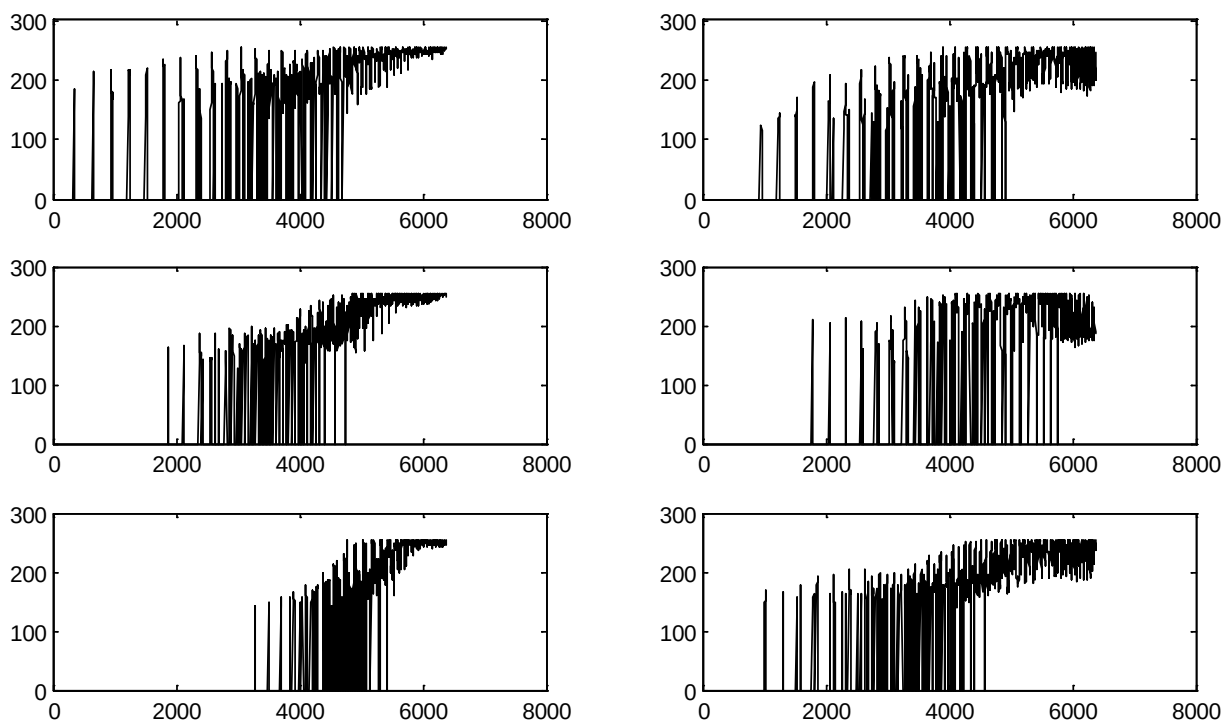


Рис. 7. Профили изображений лимфоцитов (слева) и нейтрофилов (справа) после спиралевидной развертки.

Для каждого изображения клетки вычислялись векторы признаков с помощью АМТ-алгоритма и спиралевидной развертки. Затем был использован метод главных компонент для предварительного анализа и метод проекций на латентные структуры для калибровки моделей и предсказания классов новых «неизвестных» образцов из тестового набора.

Развертка по спирали была выбрана в первую очередь, основываясь на радиальной симметрии большинства клеток. На Рис. 7 показаны примеры профилей изображений различных клеток, развернутых по спирали. Профили в левой части рисунка соответствуют лимфоцитам, а в правой — нейтрофилам. Очевидно, что размер профилей коррелирует с размером клеток и их ядер, но что более важно, видна общая тенденция в формах профилей для каждого типа клеток.

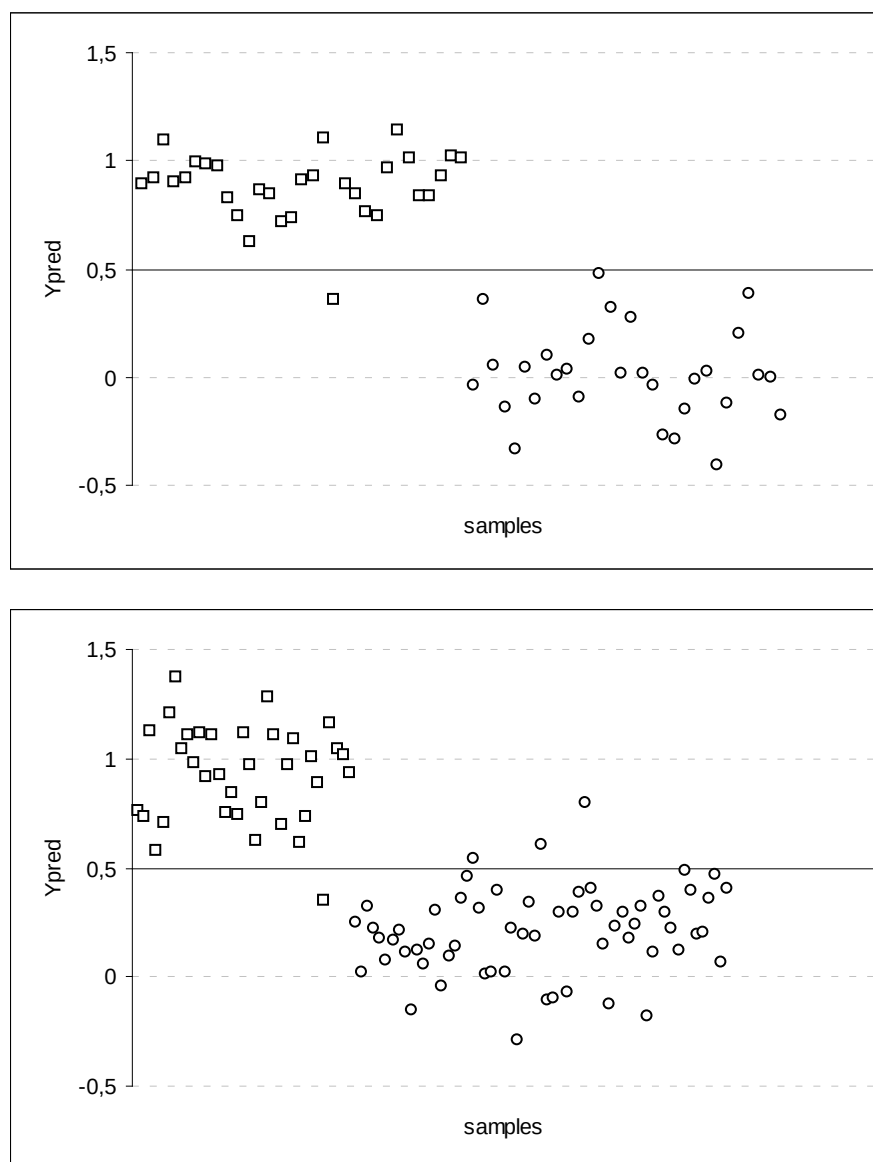


Рис. 8. Результаты ПЛС-классификации изображений клеток крови для калибровочного (вверху) и тестового (внизу) наборов данных.

Качество распознавания определялось как отношение количества неправильно распознанных клеток к общему количеству образцов каждого типа клеток, выраженное в процентах. Результаты анализа методом проекций на латентные структуры представлены на Рис. 8. В качестве отклика использовалась качественная переменная y , принимающая значения 0 для нейтрофилов и 1 — для лимфоцитов. Таким образом, объекты, имеющие предсказанное значение $-0.5 \leq y < 0.5$ (лежащие выше прямой $y = 0.5$) классифицируются, как нейтрофилы, а при $0.5 \leq y < 1.5$, как лимфоциты. На графиках лимфоциты изображены квадратами, а нейтрофилы — окружностями. График классификации в верхней части Рис. 8 соответствует обучающему набору, а классификационный график в нижней части этого рисунка — тестовому набору. Точность распознавания неизвестных образцов составила 97% для одного класса и 96% для другого. Таким образом, симбиоз АМТ и методов анализа многомерных данных позволил распознавать клетки даже на некачественных изображениях мазков крови, полученных на обычном, недорогом оборудовании. Очевидно, что даже увеличение разрешения цифровой камеры от VGA до 1–2М позволит получить еще более надежные результаты и распознать более тонкие отличия в кровяных клетках.

Заключение

Обработка и анализ изображений стали важной частью медицинской практики и диагностики. В данной статье дан обзор положения дел в этой области, рассмотрены как традиционные методы анализа, широко распространенные в медицине, так и новый подход, использующий математический аппарат хеометрики, т.е. проекционных методов анализа многомерных данных.

Как правило, традиционные методы при анализе и классификации используют в первую очередь информацию о морфологии объекта исследования, что ведет к жестким требованиям к качеству изображений и, следовательно, потребность в дорогом оборудовании. В свою очередь использование проекционных методов как непосредственно в применении к исходным изображениям, так и в сочетании с алгоритмами извлечения различных признаков позволяет получить хорошие результаты даже для изображений среднего и низкого качества. Все это позволяет говорить о высоком потенциале описанного подхода.

Литература

1. Родионова О.Е., Померанцев А.Л. "Хеометрика: достижения и перспективы", *Успехи химии*, **75** (4) 302-317 (2006)

2. Massart D.L. Chemometrics: a textbook. Elsevier, N. Y., 1988.
3. C.A. Glasbey and G.W. Horgan. Image Analysis for the Biological Sciences. John Wiley & Sons Ltd, 1995. 218p.
4. Форсайт, Дэвид А., Понс, Жан. Компьютерное зрение. Современный подход.: Пер. с англ. – М.: Издательский дом «Вильямс», 2004. – 928с.
5. Norbert Brändle, Horst Bischof, Hilmar Lapp. Robust DNA microarray image analysis. Machine Vision and Applications (2003) 15: 11–28.
6. Shuanhu Wu, Chuangcun Wang and HongYan. Mean Shift and Morphology Based Segmentation Scheme for DNA Microarray Images. D.S. Huang, X.-P. Zhang, G.-B. Huang (Eds.): ICIC 2005, Part II, LNCS 3645, pp. 41 – 50, 2005.
7. Luis Rueda and Li Qin. An Improved Clustering-Based Approach for DNA Microarray Image Segmentation. A. Campilho, M. Kamel (Eds.): ICIAR 2004, LNCS 3212, pp. 17–24, 2004.
8. Luis Rueda and Li Qin. A New Method for DNA Microarray Image Segmentation. M. Kamel and A. Campilho (Eds.): ICIAR 2005, LNCS 3656, pp. 886–893, 2005.
9. Luis Rueda and Vidya Vidyadharan. A New Approach to Automatically Detecting Grids in DNA Microarray Images. M. Kamel and A. Campilho (Eds.): ICIAR 2005, LNCS 3656, pp. 982–989, 2005.
10. T. Cremer, P. Lichter, J. Borden, D. C. Ward, and L. Manuelidis. Detection of chromosome aberrations in metaphase and interphase tumor cells by in situ hybridization using chromosome-specific library probes. Human Genetics, Vol. 80, №3 / November, 1988
11. Sabine Langer, Jürgen Kraus, Isabell Jentsch & Michael R. Speicher. Multicolor chromosome painting in diagnostic and research applications. Chromosome Research 12: 15–23, 2004.
12. Медовый В.С., Пятницкий А.М. и др. Автоматизированная микроскопия биоматериалов. «Здравоохранение и медицинская техника», №4, 2005 г.
13. C.A. Glasbey, Problems in digital microscopy (1996), Abstracts of Invited Papers at XVIIIth International Biometric Conference, Amsterdam, 183-200.
14. S. Alen, E. Cernadas et al. Comparison of Region and Edge Segmentation Approaches to Recognize Fish Oocytes in Histological Images. A. Campilho and M. Kamel (Eds.): ICIAR 2006, LNCS 4142, pp. 853–864, 2006.

15. Kyungsu Kim, Jeonghee Jeon, WanKyoo Choi, Pankoo Kim, Yo-Sung Ho. Automatic Cell Classification in Human's Peripheral Blood Images Based on Morphological Image Processing. . M. Brooks, D. Corbett, and M. Stumptner (Eds.): AI 2001, LNAI 2256, pp. 225–236, 2001.
16. Roberto Rodríguez, Teresa E. Alarcón and Juan J. Abad. Blood Vessel Segmentation via Neural Network in Histological Images. . Journal of Intelligent and Robotic Systems 36: 451–465, 2003.
17. Ongun, G.; Halici, U.; Leblebicioglu, K.; Atalay, V.; Beksac, M.; Beksac, S. An automated differential blood count system. Engineering in Medicine and Biology Society, 2001. Proceedings of the 23rd Annual International Conference of the IEEE Volume 3, 2001, p. 2583 — 2586 vol.3.
18. Alison Todman and Ela Claridge. Cell segmentation in histological images of striated muscle tissue – a perceptual grouping approach. Proceedings of Medical Image Understanding and Analysis '97 CJ Taylor, AJ Noble and JM Brady (Eds.), BMVA, 101-104.
19. Miguel A. Guevara, Augusto Silva, Helena Oliveira, Maria de Lourdes Pereira, and Fernando Morgado. Segmentation and Morphometry of Histological Sections Using Deformable Models: A New Tool for Evaluating Testicular Histopathology. A. Sanfeliu and J. Ruiz-Shulcloper (Eds.): CIARP 2003, LNCS 2905, pp. 282–290, 2003.
20. Swolin B., Simonson P., Backman S., Lofqvist I., Bredin I., Johnsson M. Differential counting of blood leukocytes using automated microscopy and a decision support system based on artificial neural networks – evaluation of DiffMaster Octavia. Clin. Lab. Haem. Vol. 25, 2003, p. 139–147. .
21. Б.З.Соколинский, В.Л.Демьянов, В.С.Медовый, А.А.Парпара, А.М.Пятницкий. Автоматическая сортировка лейкоцитов мазка крови с использованием методов обучаемых нейронных сетей и watershed. «Здравоохранение и медицинская техника», №4, 2005.
22. Блинов Н.Н. Методы компьютерной томографии в медицине. Здравоохранение и медицинская техника, №3 (17) 2005 с.10-11.
23. Shitong Wang, Duan Fu, Min Xu, and Dewen Hu. Applying Advanced Fuzzy Cellular Neural Network AFCNN to Segmentation of Serial CT Liver Images. L. Wang, K. Chen, and Y.S. Ong (Eds.): ICNC 2005, LNCS 3612, pp. 1128 – 1131, 2005.

24. Sven Lončarić and Domagoj Kovačević A Method for Segmentation of CT Head Images. LNCS 1311, pp. 388-395, 1997.s.
25. Marie-Hélène Roy Cardinal, Jean Meunier, Gilles Soulez, Éric Thérasse, and Guy Cloutier. Intravascular Ultrasound Image Segmentation: A Fast-Marching Method. . R.E. Ellis and T.M. Peters (Eds.): MICCAI 2003, LNCS 2879, pp. 432–439, 2003.
26. S. Rahnamayan, H.R. Tizhoosh, and M.M.A. Salama. Automated Snake Initialization for the Segmentation of the Prostate in Ultrasound Images. M. Kamel and A. Campilho (Eds.): ICIAR 2005, LNCS 3656, pp. 930–937, 2005.
27. Z. Yaniv, L. Joskowic, A. Simkin, M. Garza-Jinich, and C. Milgrom MD Fluoroscopic Image Processing for Computer-Aided Orthopaedic Surgery. . W.M. Wells et al. (Eds.): MICCAI'98, LNCS 1496, pp. 325-334, 1998.
28. V. P. Dinesh Kumar and Tessamma Thomas, Automatic Estimation of Orientation and Position of Spine in Digitized X-rays using Mathematical Morphology. Journal of Digital Imaging, Vol 18, No 3 (September), 2005: pp 234Y241.
29. С. Осовский. Нейронные сети для обработки информации. Финансы и статистика, 2004 г. 344 стр.
30. J. Stader. 1992: Applying Neural Networks; available as AIAI-IR-11, 1992.
31. S. Wold, K. Esbensen, P. Geladi. Principal component analysis. Chemom. Intell. Lab. Syst., 2, 37 (1987).
32. К. Эсбенсен. *Анализ многомерных данных*, сокр. пер. с англ. под ред. О.Родионовой, Из-во ИПХФ РАН, 2005 [К.Н. Esbensen. *Multivariate Data Analysis – In Practice 4-th Ed.*, CAMO, 2000]
33. T. Lied, P. Geladi, K. Esbensen, J. Chemom. 14 (2000) 585– 598.
34. D. Landgrebe, IEEE Signal Process. Mag., 19 (2002) 17–28.
35. O.Ye. Rodionova, L.P. Houmøller, A.L. Pomerantsev, P. Geladi, J. Burger, V.L. Dorofeyev, A.P. Arzamastsev "NIR spectrometry for counterfeit drug detection", *Anal. Chim. Acta*, 549, 151-158 (2005)
36. R. Andrlé, Math. Geol., 16 (1996) 83–79.
37. S. Kucheryavski. Using hard and soft models for classification of medical images. *Chemom. Intell. Lab. Syst.* In press.