ОБОСНОВАНИЕ ПОДХОДОВ К КОМПЬЮТЕРНОМУ РАСПОЗНАВАНИЮ ГИСТОЛОГИЧЕСКИХ СТРУКТУР

Дементьев Никита Александрович, Тихомирова Александра Александровна, Дохов Михаил Александрович, Стернин Вадим Евгеньевич

Федеральное государственное бюджетное образовательное учреждение высшего образования «Санкт-Петербургский государственный педиатрический медицинский университет», 194100, Санкт-Петербург, ул. Литовская, д. 2

E-mail: tikhomirova@bk.ru

Ключевые слова: клетка, морфологические признаки, изображение, компьютерное зрение

Введение. Информационные технологии в последнее время все сильнее проникают в область медицинских исследований [1], [2]. Одной из форм представления результатов является цифровое изображение. Применительно к клеточной биологии и цитологии такими изображениями будут срезы различных тканей организма [2]. Наиболее часто встречаемыми задачами анализа таких изображений являются подсчет количества клеток и определение их площади. Настоящее исследование направлено на определение основных подходов к решению задач клеточной биологии современными методами обработки цифровых изображений.

Цель исследования. Обоснование подходов к компьютерному распознаванию гистологических структур.

Материалы и методы. Распознавание гистологических структур организма относится к компьютерному зрению — сфере информационных технологий, связанной с оптическим распознаванием объектов. Любое изображение можно определить как двумерную функцию f(x,y), где x и y — координаты на плоскости [3]. Представление изображения в виде некой функции позволяет применять для обработки методы математического анализа. Применительно к цифровому изображению, эта функция является дискретной, что в некоторых случаях упрощает задачу обработки.

Существуют разные подходы к распознаванию объектов, однако все они начинают показывать свою эффективность после предварительной подготовки изображения. Одним из основных видов обработки цифровых изображений является фильтрация, которая подразумевает примитивную работу с растровыми изображениями. К процедурам этого этапа также можно отнести: кадрирование изображения, перевод цветного изображения в градации серого, увеличение контрастности, выделение контуров, бинаризацию и прочие. Использование данных процедур является необходимым, и на начальном этапе должно выполняться с привлечением квалифицированного специалиста-биолога. Особое внимание нужно уделять качеству исследуемого изображения, от него напрямую зависит успех распознавания в целом. В дальнейшем, когда качество получаемых изображений будет стабильным и удовлетворительным, последовательность процедур предварительной обработки можно автоматизировать.

На следующем этапе необходимо создать классификацию объектов, наблюдаемых на изображении. Классификация объектов является одной из важных составляющих процесса распознания образов. Сюда могут входить как целевые объекты поиска (клетки), включая их произвольное расположение (ориентация, угол среза и т. д.), так и побочные объекты (артефакты, межклеточное вещество).

Зная геометрическую форму объекта, можно использовать морфологические операторы для ее анализа и модификации. Также можно выделить дополнительную информацию: средний уровень яркости, площадь, периметр и другие параметры формы объекта [4].

Таким образом, для распознавания клеток на изображении необходимо сформировать базу данных с их морфологическими признаками. Основными признаками являются: форма клетки, включая варианты при различных углах среза; размер клетки; их взаимное расположение. Эти признаки могут сильно различаться в зависимости от вида исследуемой ткани и их выявление в каждом конкретном случае потребуют индивидуального подхода. Таким образом, под каждый вид ткани должна создаваться индивидуальная база данных морфологических признаков. Основные требования к базе данных:

- качество изображений, обеспечивающее возможность выявления признаков клетки;
- примерно одинаковое количество изображений на каждый разрез клетки;
- достаточное количество изображений для отладки алгоритма распознавания.

Дополнительно целесообразно создать базу данных морфологических признаков нецелевых объектов, что позволит проводить операцию распознавания клеток в несколько итерационных этапов. С изображения последовательно будут удаляться однозначно идентифицируемые нецелевые объекты.

На последнем этапе будет проводиться идентификация целевых объектов — гистологических структур с использованием базы данных морфологических признаков.

Реализация перечисленных выше процедур возможна посредством использования открытого программного обеспечения: библиотеки компьютерного зрения OpenCV и языка программирования Python. Данное

программное обеспечение распространяется под свободной лицензией и является кроссплатформенным, что является существенным преимуществом с точки зрения доступности и вариативности реализации под конечную операционную систему.

Результаты.

Изучены подходы к компьютерному распознаванию гистологических структур.

Обоснованы подходы к компьютерному распознаванию гистологических структур.

Предложена возможная среда разработки и сопутствующее программное обеспечение для обоснования подходов.

Выводы. Настоящее исследование показало возможность использования современных методов обработки изображения для решения задач клеточной биологии и цитологии. Развитием работы по этому направлению является практическая реализация разработанного алгоритма для анализа изображений гистологических структур.

Литература:

1. Тихомирова А.А., Котиков П.Е., Дохов М.А. Цифровое здравоохранение в России: современное состояние, проблемы и направления их решения / А.А. Тихомирова, П.Е. Котиков, М.А. Дохов // Медицина: теория и практика. 2019. Т. 4. № S. С. 538–539.
2. Котиков П.Е., Тихомирова А.А., Дохов М.А. О медицинских информационных системах на основе свободного программного обеспечения / П.Е. Котиков, А.А. Тихомирова , М.А. Дохов // Медицина: теория и практика. 2019. Т. 4. № S. С. 276–277.
3. Гонсалес Р., Вудс. Р. Цифровая обработка изображений. Издание 3-е, исправленное и дополненное. Москва: Техносфера, 2012. — 1104 с.
4. Яне Б. Цифровая обработка изображений. Москва: Техносфера, 2007. — 584 с.