|  |  |
| --- | --- |
| Gerb-BMSTU_01 | **Министерство науки и высшего образования Российской Федерации**  **Федеральное государственное бюджетное образовательное учреждение**  **высшего образования**  **«Московский государственный технический университет**  **имени Н.Э. Баумана**  **(национальный исследовательский университет)»**  **(МГТУ им. Н.Э. Баумана)** |

|  |  |
| --- | --- |
| ФАКУЛЬТЕТ | ИНФОРМАТИКА, ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И СИСТЕМЫ УПРАВЛЕНИЯ |
| КАФЕДРА | СИСТЕМЫ ОБРАБОТКИ ИНФОРМАЦИИ И УПРАВЛЕНИЯ (ИУ5) |

**РАСЧЕТНО-ПОЯСНИТЕЛЬНАЯ ЗАПИСКА**

***К НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ***

***НА ТЕМУ:***

|  |
| --- |
| ***Создание пайплайна сегментации и таксации*** |
| ***облака точек с пользовательским интерфейсом*** |
|  |
|  |
|  |
|  |

Студент \_\_ИУ5-33М\_\_\_\_ **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_П.К. Лейба\_\_\_**

(Группа) (Подпись, дата) (И.О.Фамилия)

Руководитель **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_Ю.Е. Гапанюк \_\_**

(Подпись, дата) (И.О.Фамилия)

Консультант **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_**\_\_\_\_\_\_\_\_\_\_\_\_**\_\_**

(Подпись, дата) (И.О.Фамилия)

*2024 г.*

**Министерство науки и высшего образования Российской Федерации**

**Федеральное государственное бюджетное образовательное учреждение**

**высшего образования**

**«Московский государственный технический университет имени Н.Э. Баумана**

**(национальный исследовательский университет)»**

**(МГТУ им. Н.Э. Баумана)**

УТВЕРЖДАЮ

Заведующий кафедрой \_\_\_\_\_\_\_\_\_\_

(Индекс)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(И.О.Фамилия)

« \_\_\_\_\_ » \_\_\_\_\_\_\_\_\_\_\_\_ 20 \_\_\_\_ г.

**ЗАДАНИЕ**

**на выполнение научно-исследовательской работы**

по теме \_\_\_«Создание пайплайна сегментации и таксации облака точек с пользовательским интерфейсом»\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Студент группы \_\_\_\_\_\_ИУ5-33М\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Лейба Павел Константинович\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(Фамилия, имя, отчество)

Направленность НИР (учебная, исследовательская, практическая, производственная, др.)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_учебная, исследовательская\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Источник тематики (кафедра, предприятие, НИР) \_\_\_\_\_\_\_\_\_\_\_кафедра\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

График выполнения НИР: 25% к \_\_\_ нед., 50% к \_\_\_ нед., 75% к \_\_ нед., 100% к \_\_\_ нед.

***Техническое задание:***\_\_\_« Создание пайплайна сегментации и таксации облака точек с пользовательским интерфейсом»\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

***Оформление научно-исследовательской работы:***

Расчетно-пояснительная записка на \_\_\_\_\_ листах формата А4.

Перечень графического (иллюстративного) материала (чертежи, плакаты, слайды и т.п.)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Дата выдачи задания « 10 » сентября 2024 г.

**Руководитель НИР**  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_Ю.Е. Гапанюк\_\_\_\_

(Подпись, дата) (И.О.Фамилия)

**Студент** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_П.К. Лейба\_\_\_\_\_\_

(Подпись, дата) (И.О.Фамилия)

СОДЕРЖАНИЕ

[1 СОДЕРЖАНИЕ 3](#_Toc186206797)

[3 ВВЕДЕНИЕ 4](#_Toc186206798)

[4 ИССЛЕДУЕМЫЕ ДАННЫЕ 5](#_Toc186206799)

[5 МЕТОДЫ ОБРАБОТКИ ДАННЫХ 7](#_Toc186206800)

[Удаление точек рельефа поверхности 8](#_Toc186206801)

[Фильтрация по порогу интенсивности 9](#_Toc186206802)

[Сегментация участка леса на малые участки 9](#_Toc186206803)

[Кластеризация стволов 10](#_Toc186206804)

[6 ИНТЕРФЕЙС СИСТЕМЫ НА ОБЛАЧНОЙ ПЛАТФОРМЕ 13](#_Toc186206805)

[7 ЗАКЛЮЧЕНИЕ 16](#_Toc186206806)

[8 СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ 17](#_Toc186206807)

ВВЕДЕНИЕ

В настоящее время обработка и анализ данных представляют собой ключевые элементы множества научных и прикладных исследований, включая задачи лесного хозяйства. С развитием технологий дистанционного зондирования и вычислительной техники появилась возможность автоматизировать процессы анализа больших массивов данных, что позволяет не только повысить точность исследований, но и значительно сократить временные затраты на выполнение трудоемких задач.

Одной из актуальных задач современной лесной науки является определение координат деревьев и сегментация точек облаков, полученных при помощи лазерного сканирования. Эти данные находят применение в мониторинге состояния лесов, оценке их ресурсного потенциала, а также в разработке карт лесного хозяйства, необходимых для планирования устойчивого использования лесных ресурсов.

Целью настоящей научно-исследовательской работы является анализ алгоритма, позволяющего определять координаты деревьев и выполнять сегментацию точек облаков, что обеспечивает привязку точек к каждому отдельному дереву.

Актуальность темы определяется потребностью в точных, оперативных и автоматизированных методах анализа данных для обеспечения рационального природопользования, охраны лесов от негативных воздействий и повышения их продуктивности. Разработка и внедрение алгоритмов для анализа данных дистанционного зондирования открывают новые перспективы в создании высокоточных систем управления лесными ресурсами.

ИССЛЕДУЕМЫЕ ДАННЫЕ

Облака точек представляют собой набор данных, в котором каждая точка описывается координатами X, Y, Z, формируя трехмерное цифровое представление поверхности или объекта. Такие данные получают с помощью лазерного сканирования, которое используется для точного описания пространственной структуры объектов. В лесном хозяйстве облака точек находят применение для анализа состояния лесного покрова, оценки его ресурсного потенциала, а также для расчета дендрометрических параметров деревьев, таких как высота, диаметр и объем древесины.

Процесс формирования облаков точек осуществляется с использованием технологии LiDAR (англ. Light Detection and Ranging). Данный метод основан на измерении времени или фазового сдвига лазерного импульса, который отражается от поверхности объекта. Лазерные сканеры могут быть установлены на наземных платформах или беспилотных летательных аппаратах, что позволяет сканировать местность с разной детализацией. Результатом таких измерений является плотное облако точек, содержащее информацию о положении объектов в пространстве.

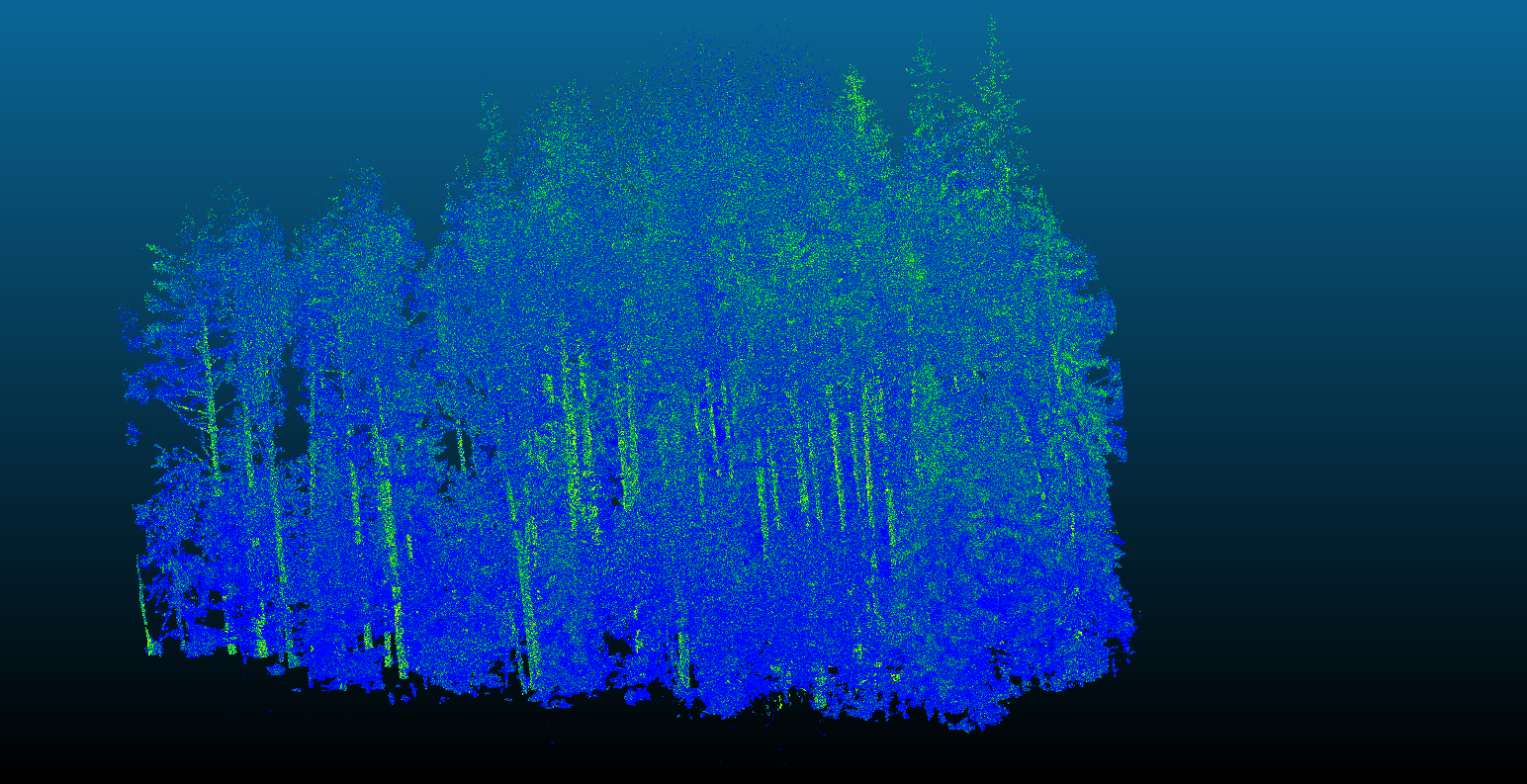


Рисунок 1 – Плотное облако точек

Одним из важных параметров облака точек является интенсивность отражения лазерного луча. Этот параметр зависит от физико-химических свойств поверхности, таких как материал, угол наклона, шероховатость и влажность. Интенсивность позволяет выделить особенности структуры поверхности, например различия между листвой и стволами деревьев, что делает анализ данных более точным. Однако её значение подвержено влиянию внешних факторов, что требует учета этих условий при обработке.

МЕТОДЫ ОБРАБОТКИ ДАННЫХ

Первым этапом при выполнении задачи автоматизированной таксации леса является обнаружение и фиксация местоположения дерева. Вторым этапом - сегментация отдельных деревьев. Третьим этапом - расчет параметров и характеристик дерева [**Unsupported source type (Journal Article) for source an-efficient-technique-for-determining-tree-coordinates-using-lidar-data-via-deep-learning.**] [**Unsupported source type (Journal Article) for source procedure-for-locating-trees-and-estimating-diameters-using-lidar-data.**].

Вкратце рассмотрим алгоритм определения координат расположения деревьев. Первым этапом обработки является удаление точек, соответствующих поверхности земли. Этот процесс включает фильтрацию данных для исключения шума, вызванного рельефом местности, что позволяет сосредоточиться исключительно на объектах лесного покрова. Далее осуществляется выделение слоя данных на основе высоты, что позволяет исключить точки, превышающие заданную высоту над поверхностью земли, ненужные для определения координат деревьев.

После этого данные проходят фильтрацию по порогу интенсивности, где из облака точек исключаются элементы с интенсивностью отражения ниже заданного значения. Это помогает проредить облако точек, убрав слабые сигналы, которые часто могут быть шумом.

Для уточнения анализа выполняется сегментация участка леса на малые участки для упрощения последующего применения алгоритмов кластеризации. Этот этап реализуется с использованием методов, таких как заливка области (flood fill) и Вороного (Voronoi Tesselation), что позволяет локализовать данные в пределах небольших пространственных областей.

На следующем этапе сегментируются отдельные деревья с использованием алгоритмa  DBSCAN [**Unsupported source type (Conference Proceedings) for source a-density-based-algorithm-for-discovering-clusters-in-large-spatial-databases-with-noise.**], что позволяет определить точки облака к конкретным деревьям.

Заключительным этапом является вычисление координат расположения стволов деревьев, который совпадает по X, Y с центроидой каждого кластера и по Z находится рядом с ближайшей точкой рельефа поверхности. Опционально, в качестве дополнительной очистки, опционально производится классификация каждого облака на основе глубокого обучения для определения, являеется ли облако деревом.

Удаление точек рельефа поверхности

Процесс удаления точек, соответствующих рельефу поверхности, начинается с разбиения облака точек с использованием kd-дерева. Эта структура данных эффективно делит многомерное пространство на ячейки, последовательно разрезая его гиперплоскостями. В результате каждое подпространство содержит подмножество точек, что ускоряет поиск ближайших соседей. Используя kd-дерево, можно легко выделить локальные наборы соседей для каждой точки, что упрощает анализ геометрических характеристик поверхности.

На следующем этапе для каждой точки вычисляются нормали, которые характеризуют локальную ориентацию поверхности. Для этого применяется ковариационный анализ, позволяющий определить главные оси распределения ближайших соседей. Матрица ковариации рассчитывается для группы ближайших точек, определённых с помощью kd-дерева, а её собственные векторы используются для определения направления нормалей.

После вычисления нормалей выполняется фильтрация точек по нескольким критериям. Один из них – угол наклона нормали: точки, чьи нормали имеют наклон, превышающий установленный порог, обозначаются как элементы рельефа. Другой критерий – близость точки к поверхности земли. Точки, расположенные ниже определённого уровня по высоте (оси Z), также обозначаются как рельеф.

Описанный процесс является важным для предподготовки данных к определению координат деревьев и последующему анализу их структуры и характеристик.

Фильтрация по порогу интенсивности

Фильтрация по порогу интенсивности осуществляется для удаления точек с характеристиками, не подходящими для определения координат деревьев. Интенсивность отраженного лазерного сигнала зависит от физических свойств поверхности, таких как материал, угол наклона, шероховатость и содержание влаги, что позволяет использовать этот параметр для предварительной очистки данных. Низкая интенсивность чаще всего соответствует слабым сигналам, которые могут быть вызваны объектами с плохой отражательной способностью или шумом измерений.

Исключение точек с низкой интенсивностью позволяет устранить элементы, которые не несут значимой информации, например, слабые сигналы от листвы, травы или других мелких объектов.

Более высокие значения интенсивности чаще связаны с объектами, находящимися под прямым углом к источнику лазера, такими как стволы деревьев или крупные ветви. Отбор таких точек улучшает точность сегментации и классификации, сохраняя ключевую информацию о структуре объектов и снижая общий объём данных для обработки.

Для определения координат абсолютно не нужно учитывать любые точки выше какого-то порога (если задача стоит в определении координат стволов, без учета кроны). Для этого, точки с координатой Z выше 3 метров удаляются.

Сегментация участка леса на малые участки

При большом количестве стволов и высоком разрешении облака точек кластеризация становится крайне ресурсозатратной задачей. Для уменьшения вычислительной нагрузки облако точек разбивается на более мелкие сегменты, что позволяет анализировать отдельные участки поочередно, упрощая выполнение алгоритмов кластеризации. Такой подход сокращает объем одновременно обрабатываемых данных и увеличивает общую производительность системы.

Одним из методов разбиения является тесселяция Вороного. В её основе лежит разделение пространства на области, каждая из которых соответствует определенной исходной точке — генератору. Генераторы представляют собой набор заранее заданных точек, вокруг которых формируются области. Все точки пространства, находящиеся ближе к своему генератору, чем к другим, включаются в соответствующую область. Это деление на полигоны Вороного позволяет разложить весь объем данных на независимые части.

Алгоритм тесселяции начинается с определения позиций генераторов. Они могут быть выбраны равномерно, с учетом плотности точек или в зависимости от трека съемки. Далее, границы между областями строятся на основе серединных перпендикуляров, проведенных между соседними генераторами. Эти границы образуют полигоны, определяющие области Вороного. Каждая ячейка включает точки, ближайшие к своему генератору, что обеспечивает локальную обработку данных. Такое разбиение снижает сложность кластеризации, позволяет учитывать локальные особенности структуры леса и повышает точность последующих этапов анализа.

Кластеризация стволов

Для каждой области, полученной в результате тесселяции Вороного, применяется алгоритм DBSCAN (Density-Based Spatial Clustering of Applications with Noise) [**Unsupported source type (Conference Proceedings) for source a-density-based-algorithm-for-discovering-clusters-in-large-spatial-databases-with-noise.**], который позволяет выделить группы точек, соответствующие стволам деревьев, и удалить шум. Основное преимущество DBSCAN заключается в его способности обрабатывать данные произвольной формы и учитывать различия в плотности, что делает его особенно подходящим для работы с облаками точек лесных массивов.

Алгоритм представляет собой итеративный вариант обхода точек и выполняется следующим образом:

1. Выбирается случайная точка, которая не была ранее посещена. Для этой точки определяется её eps-окрестность — множество точек, находящихся на расстоянии не больше eps.

2. Если в eps-окрестности найдено меньше MinPts точек, точка переносится в список возможных выбросов. Затем выбирается другая непосещённая точка, и процесс повторяется.

3. Если количество точек в eps-окрестности превышает MinPts, точка помечается как ядровая, и вокруг неё формируется кластер. Затем алгоритм проверяет соседей этой точки. Если соседние точки также являются ядровыми, их окрестности добавляются к текущему кластеру, и процесс расширения продолжается. Этот шаг выполняется рекурсивно до тех пор, пока все точки, подходящие для добавления в кластер, не будут проверены. В ходе этого:

* Выбранная точка исключается из списка точек, которые нужно обойти.
* Точка помечается новой меткой и становится частью текущего кластера.
* Проверяются её соседи, чтобы определить, должны ли они быть добавлены в кластер. Если соседняя точка является краевой, она остаётся на границе кластера; если точка подходит для добавления, она становится частью текущего обхода соседей.

4. Алгоритм возвращается к выбору следующей произвольной непосещённой точки и повторяет шаги 1–3 до тех пор, пока не будут рассмотрены все точки в наборе данных.

5. После завершения обхода всех точек, точки из списка возможных выбросов проверяются повторно. Если они не соответствуют условиям для включения в кластеры, они помечаются как шумовые точки. В противном случае они присоединяются к наиболее близкому кластеру.

Алгоритм DBSCAN, применяемый отдельно для каждой области Вороного, возвращает набор кластеров, центры которых определяются как координаты стволов деревьев.

ИНТЕРФЕЙС СИСТЕМЫ НА ОБЛАЧНОЙ ПЛАТФОРМЕ

Google Colab [**Unsupported source type (web\_page) for source welcome-to-colab---colab.**] представляет собой облачную платформу для разработки, анализа и выполнения вычислений на основе Jupyter Notebook. Платформа особенно полезна при реализации сложных систем, таких как системы анализа лесных массивов, благодаря своей интеграции с мощными облачными вычислительными ресурсами и поддержке многочисленных библиотек машинного обучения и анализа данных.

Одним из главных преимуществ Google Colab является предоставление вычислительных ресурсов, включая графические процессоры (GPU) NVIDIA Tesla K80 с 12 ГБ памяти и доступ к TPU. Это позволяет значительно ускорять вычисления, что особенно важно при работе с большими наборами данных. Для стандартных задач в бесплатной версии Colab доступно до 12 ГБ оперативной памяти.

Кроме того, Colab позволяет легко интегрироваться с Google Drive, что даёт возможность эффективно загружать, хранить и обрабатывать большие массивы данных, такие как облака точек. Пользователи могут сохранять свои проекты и работать над ними совместно, что делает платформу удобной для командных исследований.

С использованием Google Colab был реализован интерфейс для анализа облаков точек лесных массивов. Автоматически скачивается репозиторий с кодом проекта, устанавливаются зависимости и предлагается подключить Google Drive.

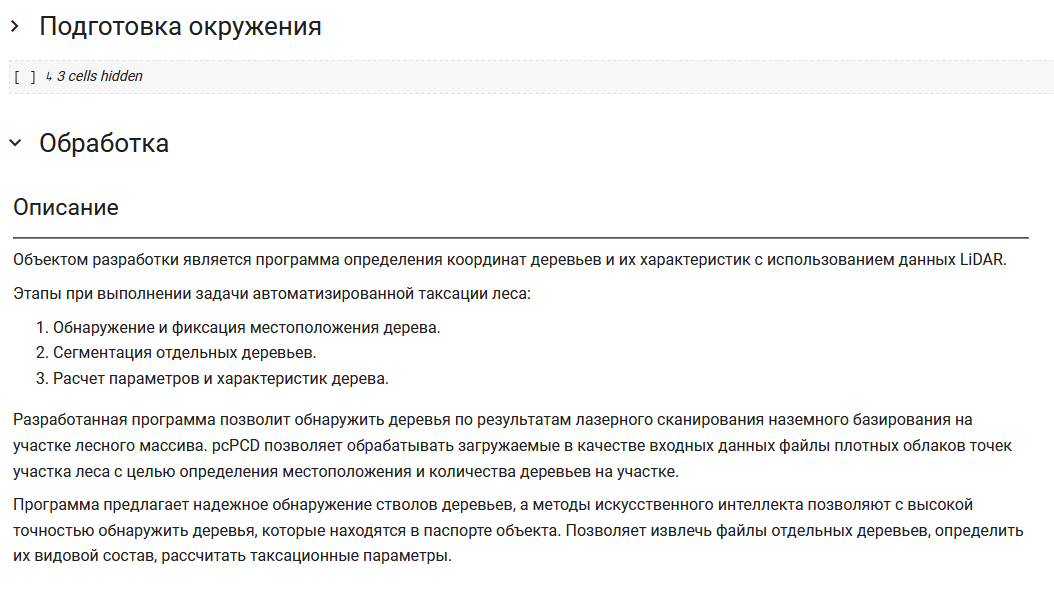


Рисунок 2 – Интерфейс в Google Colab

Интерфейс реализован, держа конечного пользователя в голове, который может быть не знаком с языком Python и программированием в целом. Поэтому, на основе Markdown созданы формы, позволяющие менять значения переменных, даже не открывая код.

Интерфейс состоит из нескольких блоков, содержащих общие настройки (пути до необходимых файлов в Google Drive), а также настроек пайплайнов определения координат и сегментации.



Рисунок 3 – Форма установки общих настроек проекта

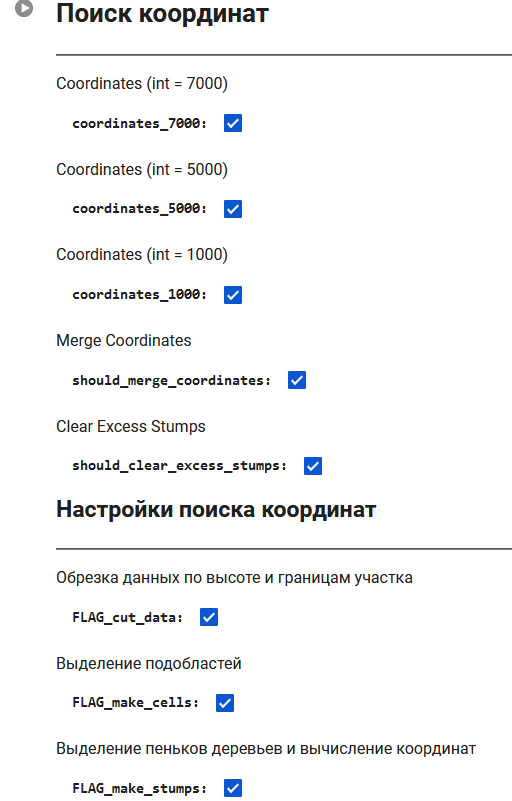


Рисунок 4 – Настройка параметров алгоритма поиска координат

ЗАКЛЮЧЕНИЕ

В данной работе рассмотрены методы анализа данных для задач, связанных с исследованием лесных массивов. Были изучены современные подходы к обработке облаков точек, полученных с помощью технологии LiDAR, включая их сегментацию, фильтрацию и кластеризацию. Облака точек представляют собой ценный источник данных для мониторинга состояния лесных насаждений, оценки их ресурсов и создания высокоточных моделей структуры деревьев.

Особое внимание было уделено алгоритмам обработки данных, таким как удаление точек рельефа поверхности, фильтрация по интенсивности отраженного сигнала и тесселяция Вороного. Эти методы позволяют оптимизировать процесс анализа больших массивов данных, выделяя значимые объекты, такие как стволы деревьев, и удаляя шум. Применение тесселяции Вороного для разбиения облаков точек на участки снижает вычислительную сложность, а алгоритм DBSCAN обеспечивает точное определение координат стволов деревьев.

Кроме того, в работе продемонстрировано, как использование облачной платформы Google Colab облегчает выполнение сложных вычислений и реализацию интерфейсов для анализа лесных массивов. Возможности интеграции с облачными хранилищами данных, использование GPU и создание интерактивных форм делают Google Colab подходящим инструментом для реализации подобных систем.

Таким образом, разработанная система анализа данных на основе облаков точек открывает новые возможности для автоматизированного мониторинга лесных массивов. Она сочетает в себе высокую точность, масштабируемость и удобство использования, что делает её перспективным решением для экологического мониторинга и управления лесными ресурсами.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ