(оценка)

Члены комиссии

(оценка) (Ф.И.О)

(подпись) (дата)

(оценка) (Ф.И.О)

(подпись) (дата)

Муром 2020

Руководитель

Абрамова Е.С.

(фамилия, инициалы)

(подпись) (дата)

Студент ПМИ-118

(группа)

Семенов И.А.

(фамилия, инициалы)

(подпись) (дата)

Министерство науки и высшего образования Российской Федерации

**Муромский институт (филиал)**

федерального государственного бюджетного образовательного учреждения высшего образования

**«Владимирский государственный университет**

**имени Александра Григорьевича и Николая Григорьевича Столетовых»**

**(МИ ВлГУ)**

Факультет информационных технологий

Кафедра ФПМ

Курсовая работа

По Специальные главы математики --

(наименование дисциплины)

Тема Реализация генетического алгоритма и его исследование при решении -

задачи кластеризации --

Введение 3

1. Анализ технического задания 4

1.1. Формулировка задания на курсовую работу 4

1.2. Описание предметной области 4

Список использованной литературы 8

# Введение

Цель: создать информационную систему, реализующую генетический алгоритм, которая позволит кластеризовать трехмерные плоскости.

Задачи:

1. Описание задачи кластеризации плоскостей

2. Определение принципа работы генетического алгоритма

3. Определение набора входных данных

4. Определение метрик качества работы алгоритма

5. Завершение проектирования рабочим прототипом информационной системы

# 1. Анализ технического задания

## 1.1. Формулировка задания на курсовую работу

Задание курсовой работы – реализация генетического алгоритма и его исследование при решении задачи кластеризации. Настоящая пояснительная записка курсовой работы отражает выполнение следующих этапов:

1. Описание задачи кластеризации плоскостей

1. Реализация прототипа информационной системы, использующего генетический алгоритм для решения упрощенной задачи

2. Реализация прототипов информационной системы, использующих другие алгоритмы машинного обучения для решения аналогичных задач

3. Сравнение эффективности созданных прототипов, определение необходимости применения генетического алгоритма

4. Подготовка реального датасета

5. Настройка и доработка созданного на первом этапе прототипа для обеспечения эффективной работы с подготовленным датасетом

## 1.2. Описание предметной области

Предметной областью работы является подготовка плоскостей в трехмерном пространстве, которые являются результатом сканирования городской среды. Создание таких компьютерных моделей является важным этапом проектирования новых зданий и реконструкции уже существующих.

На текущем этапе развития технологий наиболее прогрессивные средства такого моделирования позволяют с высокой точностью производить сканирование, но последующая обработка (ручная или автоматическая) сырых данных требует больших вычислительных ресурсов. Поэтому существует задача оптимизации алгоритмов, позволяющих эффективно выделять из сырых данных сканирования важную информацию.

Любая трехмерная плоскость (модель) может быть представлена в виде графа. Тогда задача выделения существенных данных из сырых данных сканирования является задачей удаления подграфов несвязного графа и создания ребер связности. Таким образом, очевидна задача кластеризации графов.

Из [1]:

"Применение кластерного анализа предполагает следующие этапы:

Отбор выборки для кластеризации.

Определение множества переменных, по которым будут оцениваться объекты в выборке.

Вычисление значений той или иной меры сходства между объектами.

Применение метода кластерного анализа для создания групп сходных объектов.

Проверка достоверности результатов кластерного решения."

Выборкой входных данных для кластеризации для выбранной задачи являются сырые данные фотограмметрии (создания трехмерной модели по ее фотографиям), взятые с использованием дрона. Пример получения таких данных приводится на рисунках 1, 2 [2].

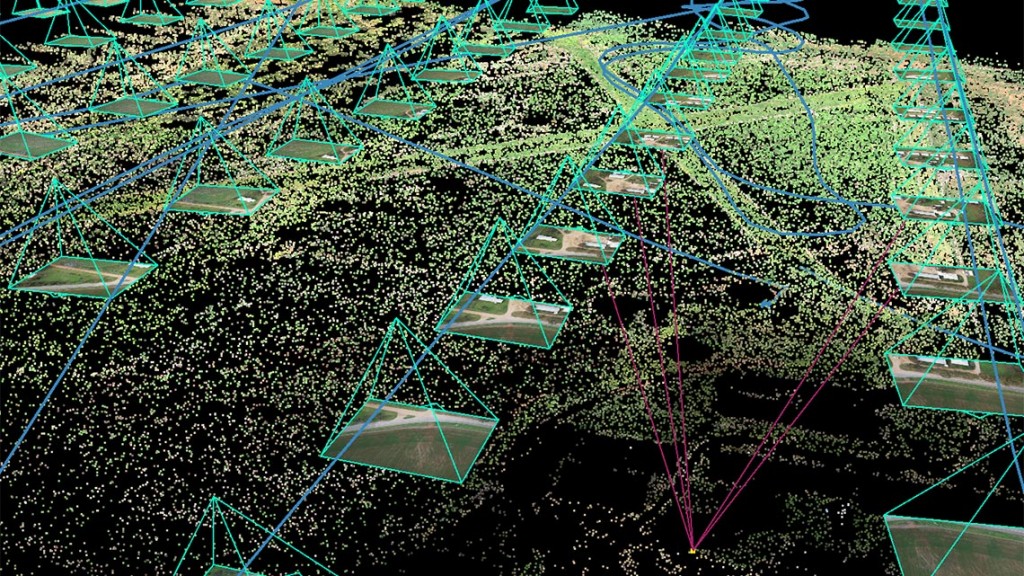


Рисунок 1 – получение облака точек

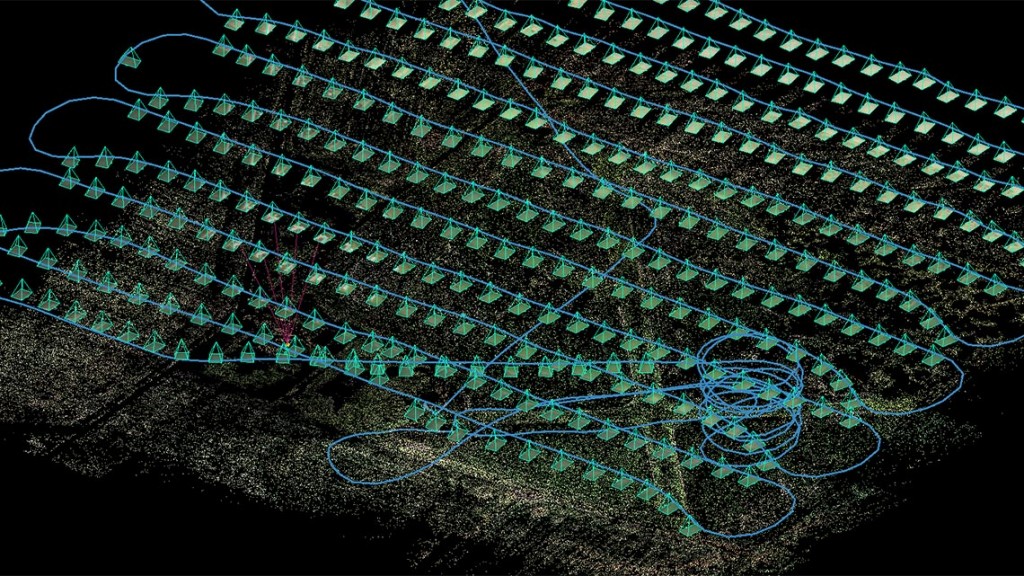


Рисунок 2 – маршрут облета

Определения, приведенные ниже, взяты из книги [3].

Вектор – упорядоченный набор чисел, называемых компонентами вектора. Так как вектор можно представить в виде строки его координат, то в дальнейшем понятия вектора и строки считаются идентичными.

Булев вектор— вектор, компоненты которого принимают значения из двухэлементного (булева) множества, например, {0,1} или {−1,1}.

Хеммингово расстояние— используется для булевых векторов и равно числу различающихся в обоих векторах компонент.

Хеммингово пространство— пространство булевых векторов, с введенным на нем расстоянием (метрикой) Хемминга.

В случае булевых векторов размерности n рассматриваемое пространство представляет собой множество вершин n-мерного гиперкуба с хемминговой метрикой. Расстояние между двумя вершинами определяется длиной кратчайшего соединяющего их пути, измеренной вдоль ребер.

Хромосома— вектор (или строка) из каких-либо чисел. Если этот вектор представлен бинарной строкой из нулей и единиц, например, 1010011, то он получен либо с использованием двоичного кодирования, либо кода Грея. Каждая позиция (бит) хромосомы называется геном.

Индивидуум (генетический код, особь) — набор хромосом (вариант решения задачи). Обычно особь состоит из одной хромосомы.

Расстояние— хеммингово расстояние между бинарными хромосомами.

Кроссинговер(кроссовер) — операция, при которой две хромосомы обмениваются своими частями. Например, 1100&1010−→1110&1000.

Мутация— случайное изменение одной или нескольких позиций в хромосоме. Например, 1010011−→1010001.

Инверсия —изменение порядка следования битов в хромосоме или в ее фрагменте. Например,1100−→0011.

Популяция— совокупность индивидуумов.

Пригодность(приспособленность) — критерий или функция, экстремум которой следует найти.

Локус— позиция гена в хромосоме.

Аллель— совокупность подряд идущих генов.

Эпистаз— влияние гена на пригодность индивидуума в зависимости от значения гена, присутствующего в другом месте. Ген считают эпистатическим, когда его присутствие подавляет влияние гена в другом локусе. Эпистатические гены из-за их влияния на другие гены иногда называют ингибирующими. Подавление проявления гена неаллельным ему геном называется гипостазом, а сам подавляемый ген — гипостатическим.

# Список использованной литературы

1. Тюрин А.Г., Зуев И.О. Кластерный анализ, методы и алгоритмы кластеризации // Вестник МГТУ МИРЭА. 2014. №2 июнь 2014 выпуск 3. С. 86-97.

2. Photogrammetry on the fly // maptek URL: https://www.maptek.com/blogs/photogrammetry-on-the-fly/ (дата обращения: 12.03.2021).

3. Панченко Т.В. Генетические алгоритмы. Астрахань: Издательский дом «Астраханский университет», 2007.