

# Методы автоматизированного дешиффрирования многозональных снимков

***Неконтролируемая классификация:*** алгоритм быстрого выделения кластеров, итерационный алгоритм последовательной кластеризации.

***Контролируемая классификация:*** формирование обучающей выборки, метод классификации по минимальному расстоянию, метод параллелепипеда, метод максимального правдоподобия.

# 1. Неконтролируемая классификация

- Смысл **неконтролируемой классификации** заключается в разделении всех пикселов изображения на группы (кластеры), название, спектральные характеристики и даже само существование которых предварительно неизвестны.
- **Критерием** отнесения пикселов к тому или другому кластеру служит **схожесть спектральных характеристик**.
- В задачу дешифровщика входит последующее определение соответствия выделенных кластеров классам земной поверхности, которое выполняется с использованием дополнительной информации — материалов наземных наблюдений, карт и т.д.

# 1.1. Алгоритм быстрого выделения кластеров

Анализ цифрового снимка осуществляется построчно, предварительно задаются значения **пороговой величины**, т.е. предельного расстояния от пиксела до центра кластера, и среднего квадратического отклонения.

**В качестве центра первого кластера** выбирается **первый элемент строки**. Далее последовательно **вычисляется евклидово расстояние от этого центра до следующего пиксела в строке**. **Это расстояние сравнивается с выбранным порогом**. Если расстояние от анализируемого пиксела до центра первого кластера меньше порогового значения, то пиксел относится к сформированному кластеру, центр которого перевычисляется, т.е. вычисляется среднее из двух значений. В противном случае формируется новый кластер, а второй пиксел становится его центром. Аналогично рассматривается каждый следующий пиксел в строке.

После просмотра всей первой строки выполняется анализ следующей. При анализе каждой очередной строки учитываются результаты просмотра предыдущих строк и в качестве центров кластеров задаются полученные ранее координаты. Так последовательно просматриваются все пикселы снимка и формируются **кластеры, именами которых служат порядковые номера**.

## 1.2. Итерационный алгоритм последовательной кластеризации

Часто неконтролируемая классификация выполняется в 2 этапа: на первом используется алгоритм быстрого выделения кластеров, а на втором — итерационный алгоритм последовательной кластеризации.

*В отличие от первого алгоритма он требует задания числа кластеров, значений их центров или числа итераций.*

- *Первый шаг* работы алгоритма заключается в отнесении пикселов изображения к одному из кластеров. В качестве критерия может быть взято минимальное расстояние до центра кластера.
- *Вторым шагом* является вычисление нового положения центра кластера с учетом отнесенных к нему пикселов. Задача считается решенной, если вычисленные координаты центров совпадут с заданными).

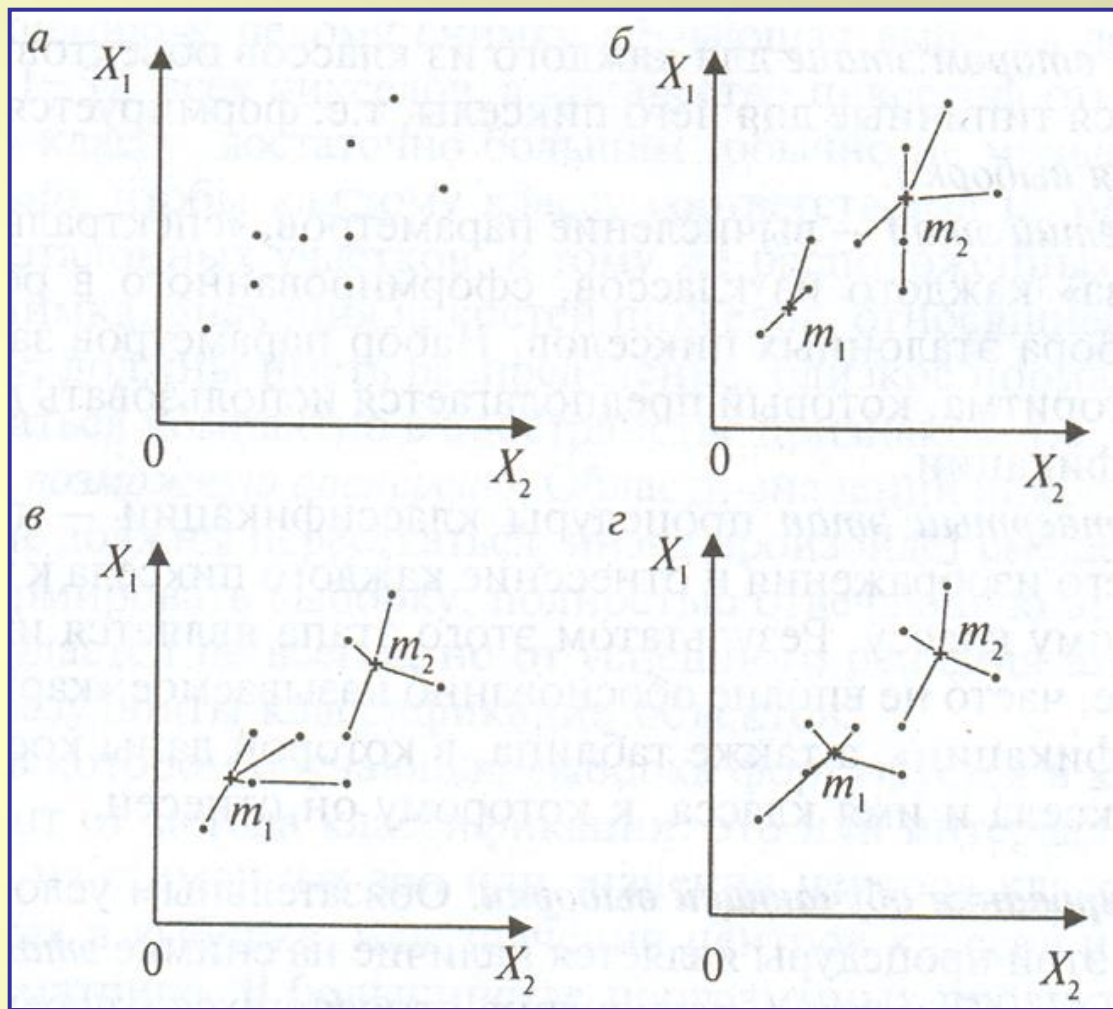
**Итерация** - это многократное повторение некоторой матем. операции при организации обработки данных в программировании, при котором используется накопленный предыдущий результат такой же операции или действия.

**Критериями качества кластеризации** служат

значения  
среднеквадратических  
отклонений расстояния от  
центра кластера до каждой  
включенной в него точки и  
сумма этих отклонений для  
всех кластеров.

Процесс продолжается до  
достижения заданных  
значений числа кластеров  
или среднеквадратического  
отклонения или до  
выполнения заданного числа  
итераций.

Обычно, для того чтобы  
добиться хороших результатов,  
необходимо выполнить  
кластеризацию несколько раз  
при разных значениях  
параметров.



**Графическая иллюстрация действия  
итерационного алгоритма последовательной  
кластеризации:**

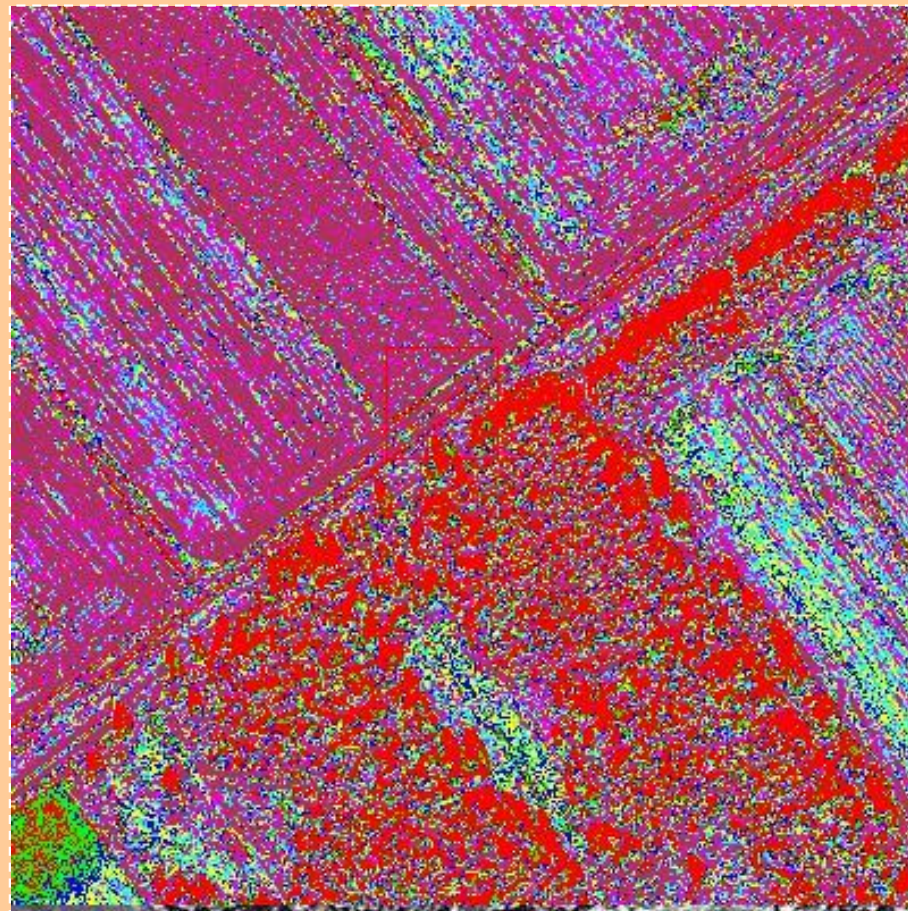
а — положение точек в двумерном пространстве признаков;  
б — первая итерация равна 20; в — вторая итерация равна 19;  
г — третья, последняя итерация равна 16

# Популярные алгоритмы

- Наиболее популярными среди алгоритмов *неконтролируемой классификации (кластеризации)* являются алгоритмы: **K-Means (K-средних)** и **ISODATA**.

Фрагмент мультиспектрального изображения с КА IKONOS

Применение алгоритма ISODATA с количеством классов 5-10 и числом итераций равным 1 (7 классов)



# Алгоритм ISODATA

## 2. Контролируемая классификация

Смысл *контролируемой классификации* заключается в отнесении каждого из пикселей снимка к определенному классу объектов на местности, которому соответствует некоторая область в пространстве признаков.

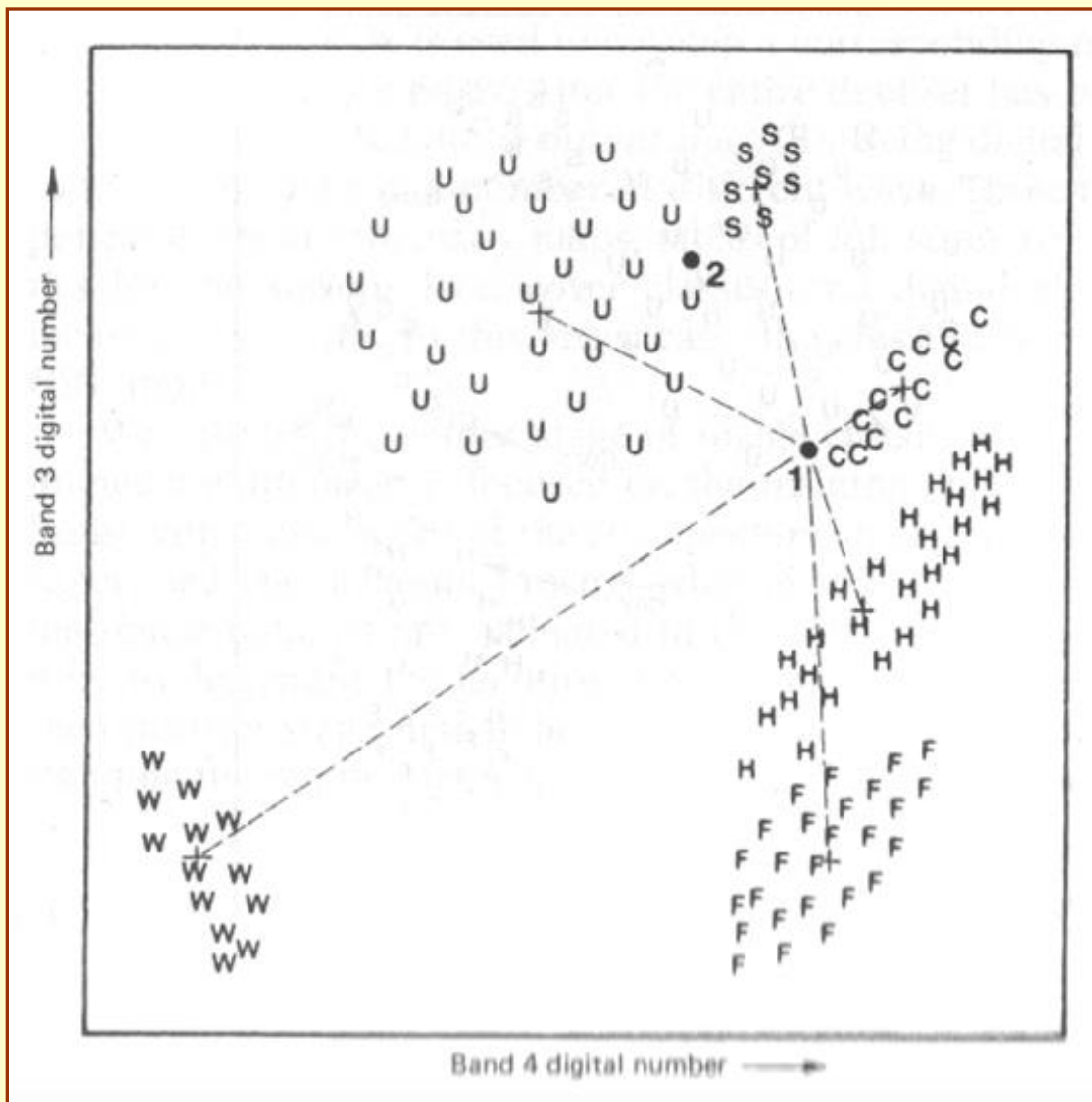
*Контролируемая классификация включает несколько этапов.*



- 1. Первый этап** заключается **в определении, какие классы объектов будут выделены в результате выполнения всей процедуры.** Это могут быть виды растительности, с/х-ные культуры, породы леса, гидрографические объекты и т.д.
- 2. На втором этапе** для каждого из классов объектов выбираются типичные для него пикселы, т.е. формируется **обучающая выборка.**
- 3. Третий этап** — **вычисление параметров, «спектрального образа» каждого из классов,** сформированного в результате набора эталонных пикселов. Набор параметров зависит от алгоритма, который предполагается использовать для классификации.
- 4. Четвертый этап** процедуры классификации — **просмотр всего изображения и отнесение каждого пиксела к тому или иному классу.** Результатом этого этапа является изображение (**«карта классификации»**), а также **таблица,** в которой даны координаты пиксела и имя класса, к которому он отнесен.

## 2.1. Метод классификации по минимальному расстоянию

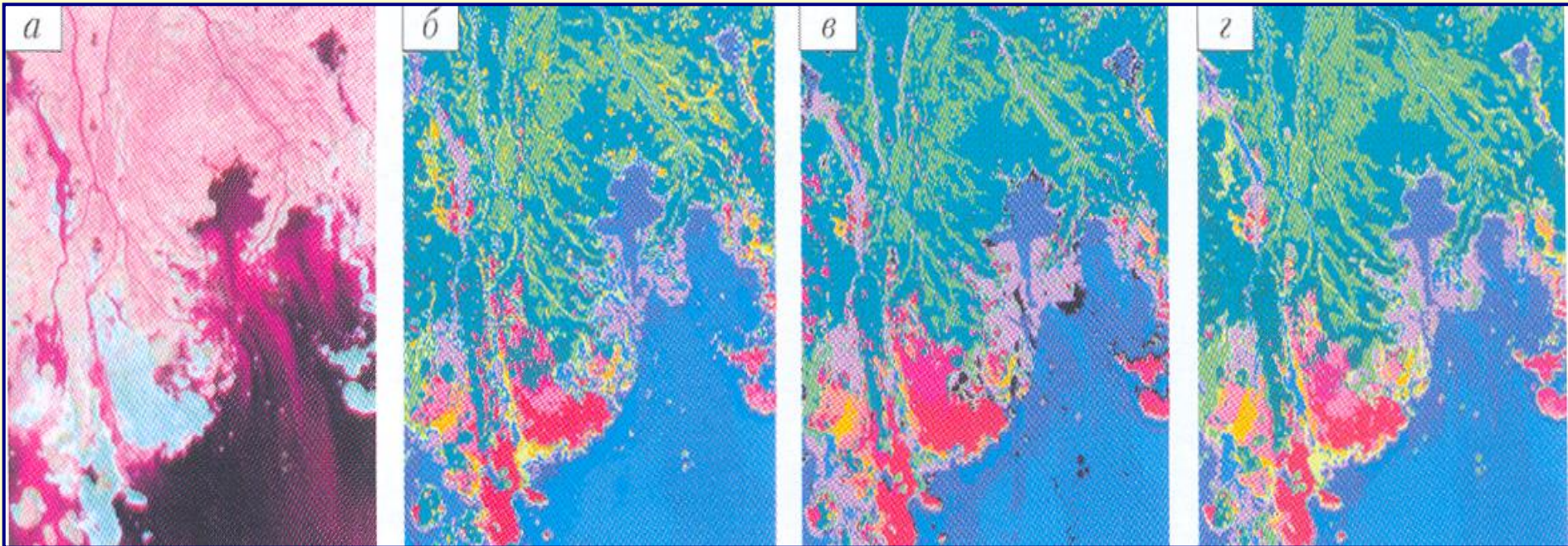
- Это один из методов классификации, основанный на **детерминированном подходе**. Он прост в вычислительном отношении. **Смысл метода** заключается в отнесении пиксела к тому эталонному классу, евклидово расстояние до центра которого в пространстве признаков **минимально**.
- **Обучающая выборка используется для вычисления средних значений яркости, характеризующих каждый класс**. Для каждого пиксела вычисляются расстояния до центров (средних значений яркости) классов, затем ему присваивается название класса, расстояние до которого минимально.
- Этот способ целесообразно использовать при ограниченном числе классов в обучающей выборке.



**Для успешной классификации** необходимо, чтобы значения спектральной яркости объектов компактно группировались вокруг средних значений соответствующих классов.

**Преимущества:** математически прост

**Недостаток:** не учитывается распределение (дисперсия) спектральных характеристик внутри эталонов



Фрагмент цветного синтезированного космического снимка (а) и результаты классификации:

б - по методу минимального расстояния;

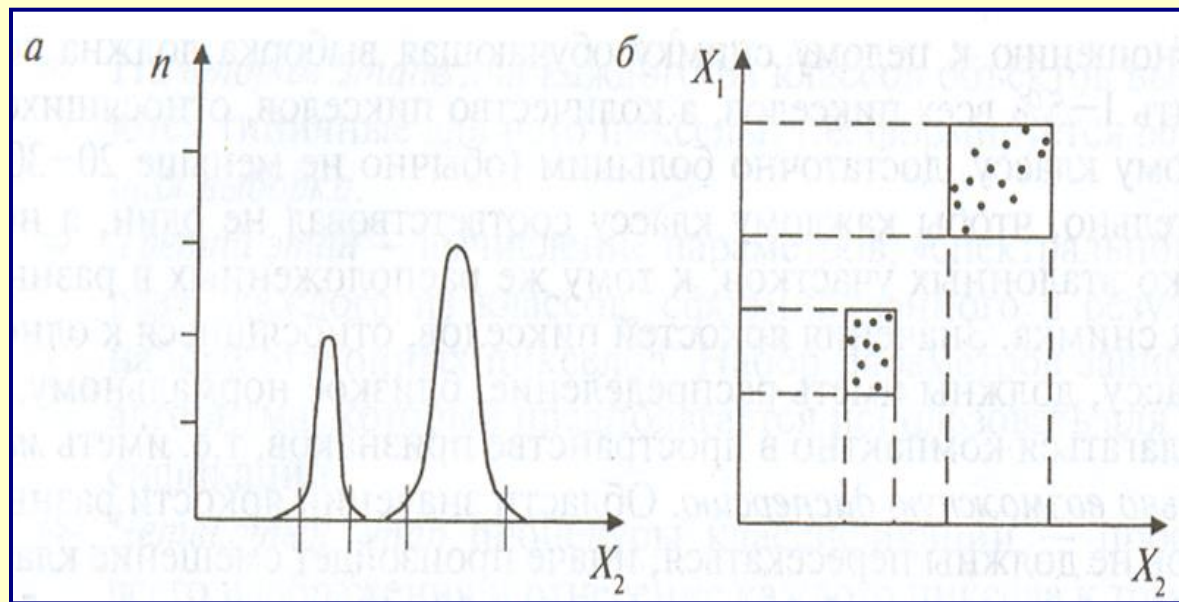
в – по методу параллелипипеда;

г – по методу максимального правдоподобия

## 2.2. Метод параллелепипедов

- Реализует наиболее простой алгоритм, основанный **на статистическом подходе**. Он заключается в отнесении пикселов изображения к эталонным классам.
- В качестве эталонных характеристик классов задаются **некоторые интервалы значений яркости**. Они выбираются *в результате анализа гистограммы* распределения яркости на изображении. В пространстве признаков интервалы значений яркости определяют **замкнутую область**, которая в зависимости от размерности поля признаков может иметь форму параллелограмма, параллелепипеда или гиперпараллелепипеда для четырех и более зон.

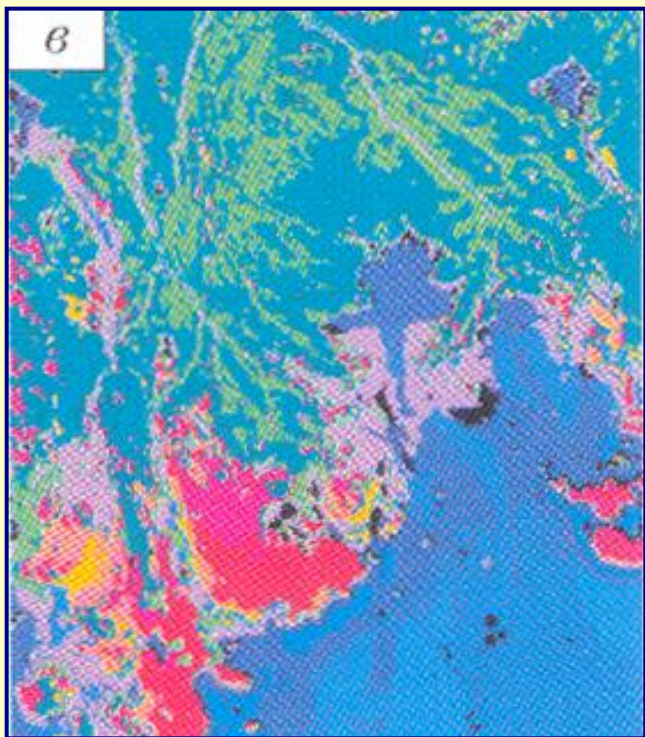
- На рис. показан принцип выбора интервалов значения яркости для двух объектов. Классификация всех пикселов снимка, состоящего, например, из трех съемочных зон, заключается в отнесении каждого из них к тому классу объектов, в параллелепипед которого он попадает по своим значениям яркости.



### Принцип классификации по методу параллелепипедов:

**а** — выбор спектральных интервалов эталонных классов; **б** — построение параллелограммов в двумерном поле признаков

- Поскольку область распределения значений спектральных признаков каждого из классов имеет четкие рамки, на снимке могут оказаться **пикселы, не попадающие ни в одну из областей**. Такие пикселы называют **неклассифицированными**, и на результирующем изображении им обычно присваивается или **белый**, или **черный** цвет.



# Метод параллелепипедов

**Преимущества:** простота обучения и использования

**Недостатки:**

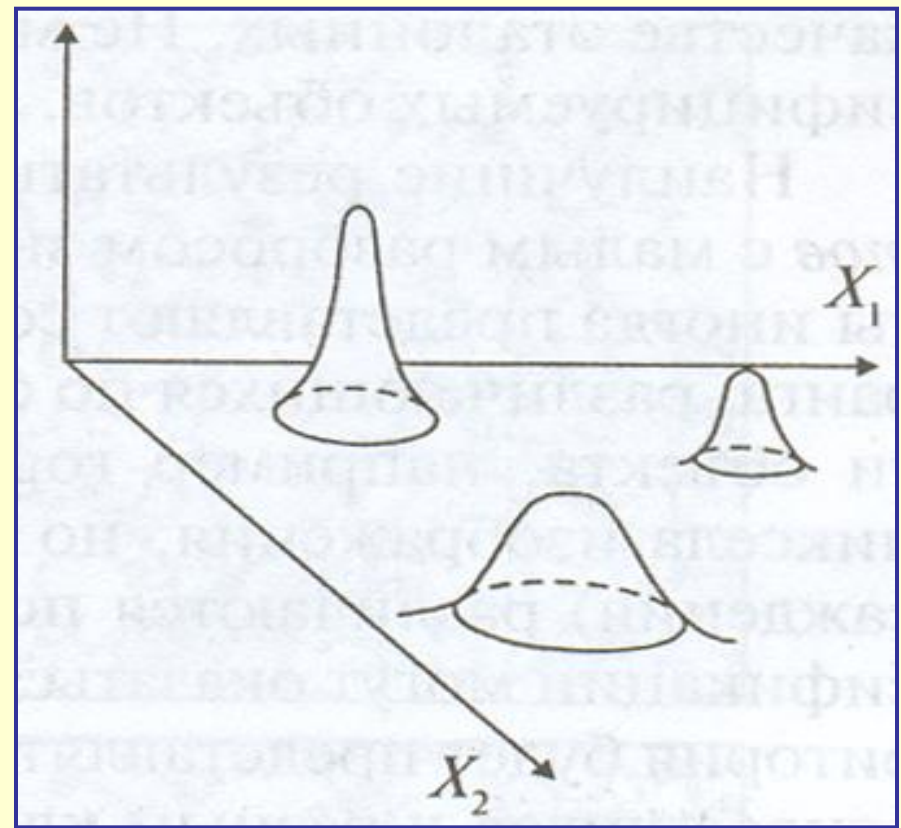
- пикселы в областях между параллелепипедами не классифицируются;
- пикселы в областях перекрывающихся параллелепипедов не могут быть классифицированы

## 2.3. Метод максимального правдоподобия

- Также основан *на статистическом подходе*. Смысл этого метода сводится к определению вероятности попадания пиксела в тот или иной класс.
- Данный способ классификации использует условие **максимальной вероятности отнесения пиксела к тому или иному классу** (*максимальное правдоподобие*)



- На рис. показано двумерное множество признаков обучающей выборки с тремя спектральными классами. Каждый из них имеет нормальное распределение значений яркости, и *вертикальная ось характеризует вероятность попадания пиксела в данный класс.*



**Двумерное пространство признаков с тремя спектральными классами**

**Отличительная особенность** классификации по методу максимального правдоподобия заключается в том, что все пикселы относятся к тому или иному классу; **отсутствуют неклассифицированные пикселы**, что характерно для этого алгоритма, как и для алгоритма минимального расстояния.