# Реализация простейших алгоритмов кластеризации в задачах технического зрения

**Задание:**

На любом языке программирования выполнить приведение количества цветов изображения  к  заданному методом k-средних (k-means). Метод реализовать самостоятельно.

Теория:

Целью задачи кластеризации является разбиение множества объектов на классы (кластеры) на основе некоторой меры сходства объектов. При кластеризации цветов мерой сходства является близость яркостей пикселей, которую будем оценивать с помощью следующей функции:

*fi = 30\*(Ri-R0)2+59\*(Gi-G0)2+11\*(Bi-B0)2.*

Множители 30, 59 и 11 - отражают различную чувствительность человеческого глаза к красному, зеленому и синему цветам соответственно.

Мы будем пытаться разбить палитру на *N=k* кластеров и заменить цвета попавшие в какой-либо кластер центром этого кластера. Стандартным методом решения такой задачи является *метод k-средних*.

**Алгоритм метода k-средних**

Пусть имеем набор векторов *xi (i=1÷p)* и *k* – число кластеров, на которые нужно разбить набор *xi*. Суть алгоритма заключается в том, чтобы найти *k* средних *mj (j=1÷k)* (центров кластеров) и отнести каждый из векторов *xi* к одному из *k* кластеров.

1. Случайным образом выбрать *k* центров *mj (j=1÷k)*;
2. Для каждого *xi (i=1÷p* подсчитать расстояние до каждого из *mj (j=1÷k);*  
   Отнести (приписать) *xi* к кластеру *j* расстояние до центра которого *mj* минимально. (Расстояние между цветами предлагается вычислять по формуле различия цветов, приведенной выше).
3. Пересчитать центры кластеров *mj (j=1÷k)* по всем кластерам;
4. Повторять шаги 2, 3 пока кластеры не перестанут изменяться;

**Пример**

Пусть имеем 10 точек: 7 5 6 9 11 15 56 45 27 20. Необходимо разбить из на 3 кластера.

Выберем 3 центра кластера случайным образом: 6 9 15.

Вычисляем расстояния для каждой точки до кластеров и записываем эту точку в ближайший кластер.

|  |  |  |  |
| --- | --- | --- | --- |
| Центр кластера | 6 | 9 | 15 |
| Точки кластера | 5 6 7 | 9 11 | 15 20 27 45 56 |
| Новые центры кластеров | 6 | 10 | 33 |

Вычисляем расстояния для каждой точки до кластеров и записываем эту точку в ближайший кластер.

|  |  |  |  |
| --- | --- | --- | --- |
| Центр кластера | 6 | 10 | 33 |
| Точки кластера | 5 6 7 | 9 11 15 20 | 27 45 56 |
| Новые центры кластеров | 6 | 14 | 43 |

Вычисляем расстояния для каждой точки до кластеров и записываем эту точку в ближайший кластер.

|  |  |  |  |
| --- | --- | --- | --- |
| Центр кластера | 6 | 14 | 43 |
| Точки кластера | 5 6 7 9 | 11 15 20 27 | 45 56 |
| Новые центры кластеров | 7 | 18 | 51 |

Вычисляем расстояния для каждой точки до кластеров и записываем эту точку в ближайший кластер.

|  |  |  |  |
| --- | --- | --- | --- |
| Центр кластера | 7 | 18 | 51 |
| Точки кластера | 5 6 7 9 11 | 15 20 27 | 45 56 |
| Новые центры кластеров | 8 | 21 | 51 |

Вычисляем расстояния для каждой точки до кластеров и записываем эту точку в ближайший кластер.

|  |  |  |  |
| --- | --- | --- | --- |
| Центр кластера | 8 | 21 | 51 |
| Точки кластера | 5 6 7 9 11 | 15 20 27 | 45 56 |
| Новые центры кластеров | 8 | 21 | 51 |

Кластеры не изменились, значит можно прекратить итерации.