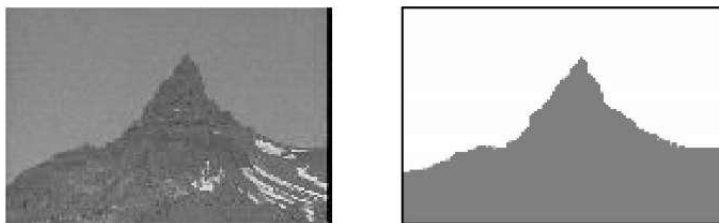


Д. Ю. Бакулин – магистрант кафедры компьютерной математики и программирования
В. А. Матъях (канд. техн. наук, доц.) – научный руководитель

ИСПОЛЬЗОВАНИЕ КЛАСТЕРИЗАЦИИ ДЛЯ ВЫДЕЛЕНИЯ ОБЪЕКТОВ НА ИЗОБРАЖЕНИЯХ

Необходимость регулярного сбора информации о поверхности Земли делает дистанционное зондирование одним из наиболее эффективных инструментов мониторинга. Космические снимки содержат большой объем информации, часть из которой является избыточной. Для упрощения обработки и анализа изображений часто используют кластеризацию.

Под кластеризацией понимается разбиение изображения на области по некоторым признакам. Предполагается, что получаемые области соответствуют реальным объектам или их частям. Часто для выделения областей одновременно используют два критерия, которые сильно связаны между собой: однородность области, определяемая уровнем сходства лежащих внутри ее пикселей, и неоднородность на границах областей, определяемая уровнем отличия граничных пикселей. При вычислении данных критериев используются значения яркости пикселей и отношение соседства на изображении [1].



Исходное изображение

Изображение после кластеризации

Рис. 1. Пример кластеризации изображения

Все методы кластеризации изображений предлагается условно можно разделить на две группы: статистические методы кластеризации и методы кластеризации, основанные на выделении перепадов яркости.

При решении проблемы кластеризации каким-либо методом из первой группы, подразумевается предварительное представление изображения в виде статистической выборки. После чего над пикселями изображения производятся операции, подобные операциям над элементами выборки.

В данной статье рассматривается алгоритм k-means (k-средних) кластеризации. Процесс кластеризации согласно этому алгоритму состоит из следующих этапов:

1. Сегментация изображения на квадратные блоки заданного размера и вычисление для каждого блока средней цветовой характеристики.

2. Произвольное распределение блоков изображения на k кластеров (начальное значение количества кластеров $k=2$).

3. Расчёт центра для каждой из характеристик каждого кластера в соответствии с формулой

$$\hat{x}_j = \frac{\sum x_i}{R_j}$$

, где R_j – количество точек j-ого региона, x_i – значения характеристик блоков, входящих в регион.

4. Перегруппировка блоков внутри кластеров - вычисление для каждого блока расстояния до

$$D(k) = \sum_{1 \leq j \leq k} [\min_i (x_i) - \hat{x}_j]$$

центра каждого из кластеров по формуле

и отнесение блока к тому из кла-

стеров, до которого расстояние по всем характеристикам блока (индекс i) минимальное.

5) После включения каждого нового блока в кластер необходимо произвести пересчет центров кластеров.

б) Далее происходит наращивание k и повторяются этапы 2 – 5.

Критерии прекращения увеличения количества кластеров:

- разброс значений $D(k)$ внутри каждого кластера становится меньше некоторого параметра θ (подобранного экспериментально);
- результат группировки при увеличении количества кластеров значительно не изменяется, т.е. $D(k) - D(k - 1) < \theta$;
- k выходит за пределы (экспериментально установлено, что не эффективно быть изображением на число кластеров больше шестнадцати), т.е. $A > 16$ [2].

Алгоритм завершается, когда на какой-то итерации не происходит изменения кластеров.

Основные достоинства алгоритма k -средних:

- простота использования;
- быстрота использования;
- понятность и прозрачность алгоритма.

Недостатки алгоритма k -средних:

- необходимо заранее знать количество кластеров;
- алгоритм очень чувствителен к выбору начальных центров кластеров. Классический вариант подразумевает случайный выбор кластеров, что очень часто являлось источником погрешности. Как вариант решения, необходимо проводить исследования объекта для более точного определения центров начальных кластеров;
- не справляется с задачей, когда объект принадлежит к разным кластерам в равной степени или не принадлежит ни одному.

На рисунке 2 показан результат практического применения алгоритма k -средних.



Исходное изображение



Изображение после кластеризации

Рис. 2. Кластеризация изображения с помощью алгоритма k -средних

Когда выполнен кластерный анализ, можно легко определить координаты целевого объекта на изображении или его границы. С помощью кластеризации выполняются задачи идентификации обнаружения; слежения за перемещениями объекта и другие. Используя процесс кластеризации можно, например, создать систему поиска изображений по визуальному содержанию или систему управления дорожным движением.

Библиографический список

1. С.А. Барталев, Т.С. Ховратович. Анализ возможностей применения методов сегментации спутниковых изображений для выявления изменений в лесах.
2. Башков Е.А., Вовк О.Л.. Кластеризация изображения методом дендрограмм.