

УДК 004.932

## КЛАСТЕРИЗАЦИЯ ГИПЕРСПЕКТРАЛЬНЫХ СПУТНИКОВЫХ СНИМКОВ ЗЕМНОЙ ПОВЕРХНОСТИ НА ОСНОВЕ МЕТОДА ПЛОТНОСТИ БЛИЖАЙШИХ СОСЕДЕЙ

*Ларионов С.М., м.н.с. НИИ «Фотон» РГРТУ, e-mail: foton@rsreu.ru;*

*Макаренко А.А., к.т.н., с.н.с. НИИ «Фотон» РГРТУ, e-mail: foton@rsreu.ru.*

### CLUSTERING OF HYPERSPECTRAL SATILLITE IMAGES OF THE EARTH'S SURFACE BASED ON THE NEAREST NEIGHBORS DENSITY METHOD

*Larionov S.M., Makarenkov A.A.*

*The paper describes modified approach of clustering of Earth remote sensing hyperspectral data based on nearest neighbors density method. Features of this approach, its advantages and limitations are considered. The results of experimental studies are presented.*

**Key words:** Earth remote sensing, hyperspectral images, image clustering, nearest neighbors method, nearest neighbors density.

**Ключевые слова:** дистанционное зондирование Земли, гиперспектральные снимки, кластеризация изображений, метод ближайших соседей, плотность ближайших соседей.

*Описывается модификация подхода к кластеризации гиперспектральных данных дистанционного зондирования Земли на основе метода плотности ближайших соседей. Рассматриваются особенности данного подхода, его преимущества и ограничения. Приводятся результаты экспериментальных исследований.*

#### Введение

На сегодняшний день дистанционное зондирование земной поверхности из космоса позволяет решать ряд важных тематических задач в картографии, метеорологии, изучении растительности и водных ресурсов, исследовании состава атмосферы и почвы, добыче полезных ископаемых, сельском хозяйстве, предупреждении и ликвидации последствий чрезвычайных ситуаций, наблюдении за антропогенными объектами и в других отраслях экономики и научных областях [1-6]. В последнее время на отечественных и зарубежных космических аппаратах устанавливается гиперспектральная аппаратура [7]. Каждая точка гиперспектрального изображения представляет собой спектральную характеристику, определяющую физико-химический состав малого участка подстилающей поверхности [8], что позволяет решать задачу классификации и оценки параметров объектов сцены с более высокой точностью по сравнению с обработкой традиционной мультиспектральной и панхроматической информации.

#### Подходы к классификации изображений

Методы классификации изображений можно разделить на две группы: на обучаемые и необучаемые подходы [8, 9]. Обучаемые подходы, например, такие как нейросети [10, 11], требуют значительного объема дополнительной информации в виде обучающих выборок. Подготовка «обучающих» изображений является сложным, время затратным процессом, не имеющим единственного решения. Результаты такой разметки во многом зависят от навыков и интуитивных предположений эксперта о принадлежности каждой конкретной точки изображения к той или иной смысловой группе. Таким образом получение обучающей выборки обычно требу-

ет существенных финансовых затрат [12]. При этом эффективность обучаемых методов кластеризации напрямую зависит от объема и качества обучающей выборки [8]. Поэтому большое внимание со стороны научного сообщества обращено в сторону необучаемых методов классификации, то есть кластеризации изображений. Под кластеризацией [13] понимается разбиение изображения на хорошо разделяемые в выбранном пространстве признаков смысловые области точек, обладающих некоторой степенью общности. Известен ряд методов кластеризации изображений, некоторые из которых рассмотрены далее.

**Метод k-средних.** На первом шаге метода [14] произвольным образом выбираются начальные центры кластеров. Затем каждая точка изображения относится к тому кластеру, центр которого находится к ней (точке) ближе всего в пространстве признаков. После чего центры кластеров пересчитываются как центры масс кластеров. Эта процедура перерасчета точек и перерасчета центров кластеров повторяется до тех пор, пока положение центров кластеров не стабилизируется. Одна из проблем данного метода в том, что существует вероятность схождения к локальному, а не глобальному оптимуму. К тому же необходимы априорные данные о числе классов на изображении, а результаты кластеризации зависят от начального выбора центров кластеров.

**Метод нечеткой кластеризации c-средних.** Согласно данному подходу [15] вначале задаются центры кластеров. Затем рассчитывается расстояние в выбранном пространстве признаков от каждой точки до центра каждого кластера и степень принадлежности каждой точки каждому из кластеров. На основе этих значений вычисляется некоторая целевая функция. За искомое принимается разбиение, обеспечивающее минимум вы-

бранной целевой функции. Основной проблемой метода является сильная зависимость результатов от выбора первоначальных центров кластеров, а также существенные вычислительные затраты.

**Метод DBSCAN.** Согласно данному методу [16], вначале необходимо задать некоторую  $\epsilon$ -окрестность (расстояние  $\epsilon$ ) в выбранном пространстве признаков, в пределах которой определяются точки кластера, а также задать минимальное число  $mPtrs$  точек, образующих кластер. Если в  $\epsilon$ -окрестности некоторой точки находится как минимум  $mPtrs$  точек, то анализируемая точка считается *плотной*.

Алгоритм состоит из следующих шагов.

**Шаг 1.** Выбирается непомеченная точка изображения.

**Шаг 2.1.** Находятся все точки, лежащие в  $\epsilon$ -окрестности выбранной точки.

**Шаг 2.2** Если количество найденных точек меньше  $mPtrs$ , то данная точка помечается как *шум* и осуществляется переход к **Шагу 1**.

**Шаг 2.3.** В противном случае (то есть когда данная точка является *плотной*) данная точка и точки, лежащие в её  $\epsilon$ -окрестности, объединяются в кластер.

**Шаг 3.1.** Среди найденных на **Шаге 2.1** точек, лежащих в  $\epsilon$ -окрестности рассматриваемой точки, также определяются *плотные* точки.

**Шаг 3.2.** Точки, лежащие в  $\epsilon$ -окрестности каждой *плотной* точки, найденной на **Шаге 3.1**, добавляются в данный кластер.

**Шаг 3.3** Процедура **Шаг 2.1-Шаг 3.2** повторяется для каждой из добавленных точек до тех пор, пока среди них не останется *плотных*. Это означает, что кластер сформирован и необходимо перейти к следующей непомеченной точке изображения, повторив **Шаги 1-3.3**.

Минусом подхода является невозможность корректной кластеризации при условии разной плотности реально присутствующих на изображении классов объектов, так как в данном случае сложно подобрать подходящие  $\epsilon$ -окрестность и число  $mPtrs$ .

**Метод OPTICS** [17] является развитием метода DBSCAN. На первом шаге также задается расстояние  $\epsilon$  в выбранном пространстве признаков и число  $mPtrs$ . Для каждой точки  $v$  находится основное расстояние  $e_c(v)$  – наименьшее расстояние, при котором анализируемая точка  $v$  является *плотной*, причем рассматриваются только  $e_c(v) \leq \epsilon$ . Если такого  $e_c(v)$  не находится, то  $v$  помечается как *шум*. Затем для каждой точки определяется расстояние достижимости  $e_r(v)$  в соответствии с выражением:

$$e_r(v) = \begin{cases} \text{не определено, если } w\text{-помечена как шум} \\ \max[e_c(w), \text{dist}(v, w)], \text{ иначе,} \end{cases} \quad (1)$$

где  $\max[\bullet]$  – оператор определения максимального значения;  $\text{dist}(v, w)$  – расстояние между точками  $v$  и  $w$ .

Согласно OPTICS создается упорядоченная база данных, в которой для каждой точки сохраняются основное расстояние и расстояние достижимости.

Такой подход позволяет избежать проблемы, связанной с различной дисперсией физических классов

изображения и выделить все классы, которые занимают в выбранном пространстве признаков окрестность, меньше или равную заранее заданному  $\epsilon$ . Однако этот подход требует больших вычислительных затрат, в особенности применительно к гиперспектральным изображениям, содержащим миллионы многомерных элементов.

**EM-алгоритм.** Данный подход [13] требует априорного знания функции плотности вероятности для каждого физического класса с соответствующим значением математического ожидания и дисперсией. Результаты кластеризации сильно зависят от первоначальной оценки вида распределения.

В публикациях [18-20] рассматривается подход к кластеризации гиперспектральных данных на основе вычисления плотности ближайших соседей в выбранном пространстве признаков (*Graph Watershed using Nearest Neighbors – метод GWENN*). Главное преимущество этой группы алгоритмов в том, что для их применения не требуется каких-либо предварительных знаний о структуре данных, таких как фактическое количество физических классов на изображении, статистическое распределение исходных данных, экземпляры классов или обучающие выборки.

#### Метод на основе плотности ближайших соседей

Идея подхода (*метод GWENN*) [18-20] заключается в определении метки каждой точки изображения исходя из меток ближайших соседей и плотностей этих ближайших соседей в выбранном пространстве признаков. На этапе предобработки выбирается число  $k$  ближайших соседей, пространство признаков, мера расстояния  $\Delta$  и функция плотности точки  $\varphi$ .

В качестве пространства признаков могут быть выбраны значения спектральной плотности энергетической яркости (СПЭЯ) точек гиперспектрального изображения [8].

За меру расстояния  $\Delta$  может приниматься, например, Евклидова метрика:

$$\Delta(x, y; i, j) = \sqrt{\sum_h (I_h(x, y) - I_h(i, j))^2}, \quad (2)$$

где  $(x, y)$ ,  $(i, j)$  – координаты точек на изображении  $I$ ;  $\Delta(x, y; i, j)$  – расстояние между точкой  $(x, y)$  и точкой  $(i, j)$ ;  $I_h(x, y)$ ,  $I_h(i, j)$  – значения СПЭЯ в  $h$ -м канале точки  $(x, y)$  и точки  $(i, j)$ .

Плотность  $\varphi(x, y)$  может быть выбрана как величина, обратно пропорциональная суммарному расстоянию от точки  $(x, y)$  до каждой из её  $k$  ближайших соседей:

$$\varphi(x, y) = \frac{1}{\sum_{(i, j) \in kNN(x, y)} \Delta(x, y; i, j)}, \quad (3)$$

где  $kNN(x, y)$  – множество  $k$  ближайших соседей точки  $(x, y)$ .

В ряде источников функция  $\varphi(x, y)$  выбирается таким образом, чтобы вес каждого слагаемого в знаменателе выражения (3) убывал в зависимости от «удаленности» в пространстве признаков соответствующего «соседа» от данной точки.

Согласно методу *GWENN*, вначале для каждой точки изображения вычисляется массив расстояний от анализируемой точки до каждой другой точки изображения. Элементы массива сортируются в порядке возрастания. Для каждой точки с координатами  $(x, y)$  определяются её  $k$  ближайших соседей и её плотность  $\varphi(x, y)$ . Затем точки ранжируются в порядке убывания значения плотности. Предполагается, что наиболее плотные точки являются типичными экземплярами физических классов, поэтому процесс кластеризации начинается именно с них (они являются первыми элементами формируемых кластеров). Принадлежность к какому-либо физическому классу точек, менее плотно окруженных своими соседями в пространстве признаков, определить сложнее, поэтому их кластеризация производится, в последнюю очередь, когда большинство их ближайших соседей уже имеют метки. Также точки с наименьшей плотностью могут быть помечены как шумовые.

Этап кластеризации точек, согласно методу *GWENN*, состоит из следующих шагов.

**Шаг 1.** Выбирается наиболее плотная точка  $(x, y)$ , не имеющая метки.

**Шаг 2.1.** Если среди множества  $kNN(x, y)$  нет точек, имеющих метку, точке  $(x, y)$  присваивается новая метка (число кластеров увеличивается на единицу) и осуществляется переход к **Шагу 1**. Метки принимают целые положительные значения, начиная с 1. Метка каждого нового кластера равняется текущему числу кластеров, увеличенному на единицу.

**Шаг 2.2.** Если среди множества  $kNN(x, y)$  есть точки, имеющие метку, то точке  $(x, y)$  присваивается метка, соответствующая наибольшей суммарной плотности среди  $kNN(x, y)$ :

$$Lbl(x, y) = \arg \max_{l \in [1, L]} \left[ \sum_{\substack{(i, j) \in kNN(x, y) \\ Lbl(i, j) = l}} \varphi(i, j) \right], \quad (4)$$

где  $Lbl(x, y)$  – метка точки  $(x, y)$ ;  $l$  – значение метки;  $L$  – текущее число кластеров;  $\varphi(i, j)$  – плотность точки

$(i, j)$ ;  $\sum_{\substack{(i, j) \in kNN(x, y) \\ Lbl(i, j) = l}} \varphi(i, j)$  – сумма плотностей точек, имеющих метку  $l$ , из множества  $kNN(x, y)$  ближайших соседей точки  $(x, y)$ .

После того, как точка  $(x, y)$  получила метку, осуществляется переход к **Шагу 1**.

Основным недостатком подхода является неоднозначность выбора числа  $k$ , непосредственно влияющего на точность кластеризации и на требуемые вычислительные затраты. С одной стороны, чтобы добиться высокого быстродействия алгоритма,  $k$  должен принимать как можно меньшее значение. Однако, например, проведенные экспериментальные исследования на изображении, полученном датчиком AVIRIS над долиной Салинас (изображение *corrected Salinas*), размером 512 строк на 217 столбцов и включающим 16 физических классов объектов, показали следующие результаты. При выборе  $k$ , равном 0,072 % от общего числа то-

чек, в результате кластеризации получены 214 кластеров. То есть физические классы, реально присутствующие на изображении, оказались разбиты на большое число малых кластеров, что в свою очередь затрудняет идентификацию объектов. Для повышения точности кластеризации, необходимо увеличить значение  $k$ . Однако при этом растет время обработки (или требуемая оперативная память), к тому же кластеры становятся более крупными и включают в себя уже несколько физических классов объектов. Например, для изображения *corrected Salinas* при выборе  $k$ , равным 3,6 % от общего числа точек получено 7 кластеров. То есть более мелкие физические классы оказались объединены с более крупными, что недопустимо. Рассмотренный метод при кластеризации натурной информации (маршрутов размером 1000x3000 пикселей) предъявляет высокие требования к аппаратным средствам (при оптимизации по времени выполнения требуемый объем оперативной памяти составляет ~43 Гб).

Известна реализация описываемого алгоритма [19] на основе пирамидальной обработки, состоящая из следующих шагов.

**Шаг 1.** Получение пирамиды изображений. Каждое изображение  $s$ -го уровня формируется путем вейвлет-преобразования изображения предыдущего  $(s-1)$ -го уровня. Изображение  $s$ -го уровня является более грубым по сравнению с изображением  $(s-1)$ -го уровня. Каждой точке  $s$ -го уровня соответствует 4 точки изображения  $(s-1)$ -го уровня.

**Шаг 2.** Кластеризация нижнего уровня пирамиды при помощи метода *GWENN*. Выделение наиболее плотной точки для каждого полученного кластера – т.е. нахождение экземпляров кластеров.

**Шаг 3.1.** Кластеризация изображения следующего  $s$ -го уровня, в ходе которой при нахождении ближайших соседей используются точки изображения  $(s-1)$ -го уровня, соответствующие экземплярам кластеров изображения  $s$ -го уровня.

**Шаг 3.2.** Выделение экземпляров кластеров на изображении  $s$ -го уровня. Повторение процедуры **Шаг 3.1** – **Шаг 3.2** до уровня исходного изображения.

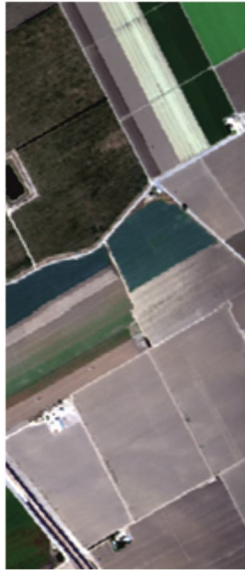
Пирамидальная реализация требует меньших вычислительных затрат, однако приводит к повышенной, по сравнению с непиримидальным методом *GWENN*, потере точек малых классов.

Для решения описанных выше проблем предлагается модификация метода *GWENN*, состоящая из следующих шагов.

**Шаг 1.** Кластеризация исходного изображения с малым  $k$  (в экспериментах использовалось  $k = 0,01$  % от числа  $N$  точек изображения). В результате получается большое число первичных кластеров, каждый из которых содержит малое количество элементов. Физические классы объектов на изображении оказываются разделены на несколько первичных кластеров.

**Шаг 2.** Нахождение «средних спектров» для каждого первичного кластера:

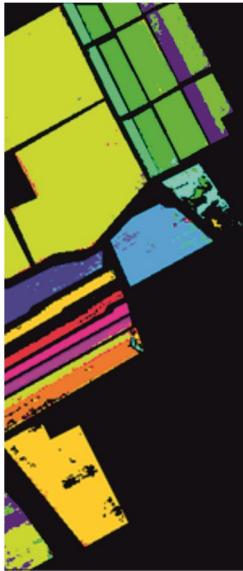
$$\bar{I}_h(l) = \frac{1}{N_l} \sum_{(x, y)} I_h^l(x, y), \quad l = \overline{1, L}, \quad (5)$$



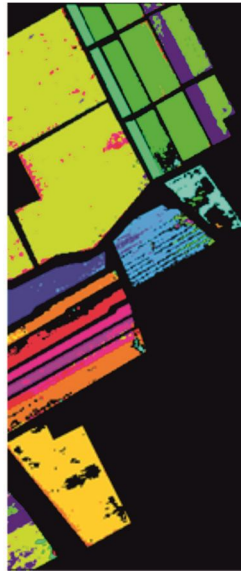
а) Исходное изображение



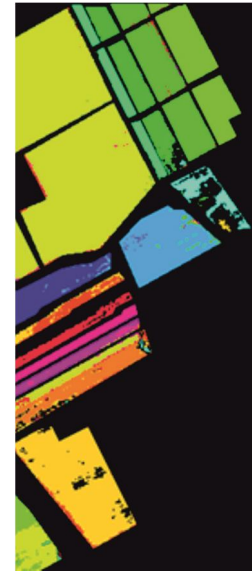
б) Эталон кластеризации



в) Результат кластеризации методом GWENN



г) Результат кластеризации пирамидальным (двухуровневым) GWENN



д) Результат кластеризации предлагаемой модификацией GWENN

Рис. 1. Кластеризация изображения *corrected Salinas*

где  $\bar{I}_h(l)$  – среднее значение СПЭЯ точек  $l$ -го первичного кластера в  $h$ -м канале;  $I_h^l(x, y)$  – значение СПЭЯ в  $h$ -м канале точки с координатами  $(x, y)$ , имеющей первичную метку  $l$ ;  $N_l$  – число точек, имеющих первичную метку  $l$ ;  $L$  – число первичных кластеров.

**Шаг 3.** Определение оптимального  $k'_{opt}$  в смысле выбранной меры разделимости  $R(k', way)$  для объединения «средних спектров»:

$$k'_{opt} = \max_{\substack{k' \in [1, L] \\ way \in \{ascend, descend\}}} (R(k', way)), \quad (6)$$

где  $k'$  – число ближайших соседей на этапе объединения первичных кластеров;  $way$  – параметр, определяющий направление кластеризации: *ascend* – от «менее плотных» точек к «более плотным», *descend* – от «более плотных» точек к «менее плотным».

**Шаг 4.** Кластеризация «средних спектров» с числом  $k'_{opt}$  ближайших соседей.

**Шаг 5.** Получение итоговых кластеров изображения – переразметки малых первичных кластеров в соответствии с результатами кластеризации их «средних спектров».

Для оптимизации числа  $k'$  ближайших соседей находится максимум функции следующего вида:

$$R(k', way) = \frac{\sigma_{inter}^2(k', way)}{\sigma_{intra}^2(k', way) \cdot NC^t(k', way)}, \quad (7)$$

где  $\sigma_{inter}^2(k', way)$  – среднее расстояние между кластерами в выбранном пространстве признаков;  $\sigma_{intra}^2(k', way)$  – среднее расстояние внутри кластеров в выбранном пространстве признаков;  $NC^t(k', way)$  – число полученных кластеров;  $t$  – показатель степени числа кластеров.

Мера разделимости  $R(k', way)$  учитывает внутрикластерное и межкластерное расстояние и число кластеров и обеспечивает компактность и отделимость результирующих кластеров в выбранном пространстве признаков. При кластеризации «средних спектров» использу-

ется проход как от «среднего спектра» с наибольшей плотностью, так и в обратном направлении (параметр *way*). Предполагается, что «средний спектр», соответствующий некоторому первичному кластеру и отстоящий от других «средних спектров» в пространстве признаков, не является шумовым, а соответствует малому классу объектов на изображении. Данный подход позволяет назначить такому малому физическому классу уникальную метку и таким образом сохранить его в итоговых кластерах, не объединяя с более крупными. Для снижения вычислительных затрат можно при кластеризации «средних спектров» перебирать не все значения  $k' \in [1, L]$ . Если при некотором значении  $k'$  все «средние спектры» объединяются в один кластер (то есть все точки изображения относятся к одному итоговому кластеру), дальнейшее увеличение значения  $k'$  не имеет смысла, потому как при этом «средние спектры» так же будут объединяться в один кластер.

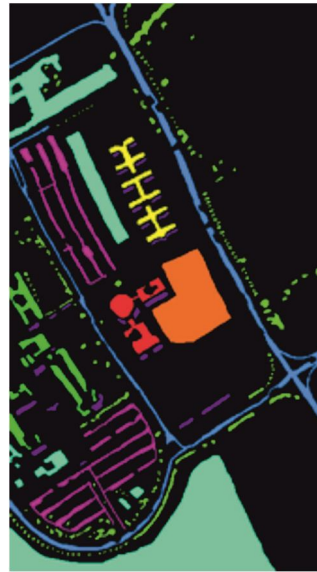
В табл. 1 и на рис. 1 и рис. 2 представлены результаты экспериментальных исследований.

### Заключение

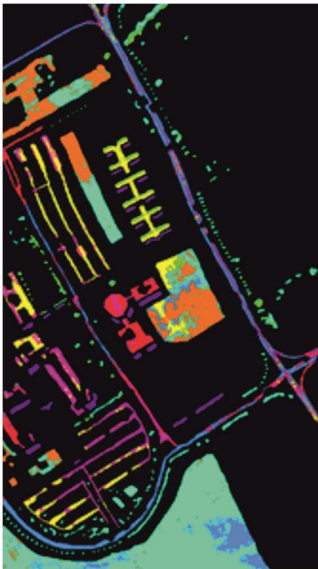
Проведенные экспериментальные исследования предлагаемой модификации метода *GWENN* показали повышение точности кластеризации и уменьшение затрат памяти. Применение предлагаемого алгоритма позволяет корректно кластеризовать присутствующие классы объектов, занимающие малую часть сцены (таким образом избежать потери информации). Дальнейшие исследования авторов направлены на подбор меры разделимости  $R$ , обеспечивающей необходимые точностные параметры и низкие вычислительные затраты при работе с гиперспектральными изображениями с различными соотношением размеров объектов наблюдаемой сцены и степенью однородности их спектральных характеристик.



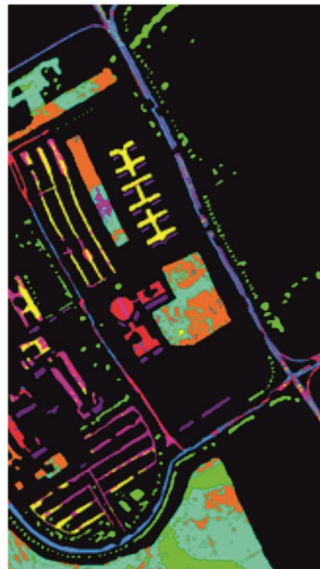
а) Исходное изображение



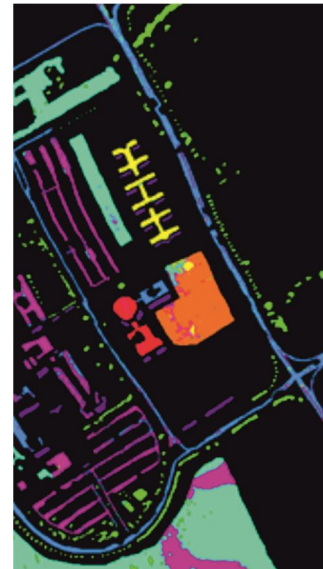
б) Эталон кластеризации



в) Результат кластеризации методом *GWENN*



г) Результат кластеризации пирамидальным (двухуровневым) *GWENN*



д) Результат кластеризации предлагаемой модификацией *GWENN*

Рис. 2. Кластеризация изображения *Pavia University*

Таблица 1. Результаты экспериментальных исследований

	Метод GWENN	Пирамидальный (двухуровневый) GWENN	Предлагаемая модификация GWENN
Изображение <i>corrected Salinas</i> (512 строк, 217 столбцов, 214 каналов, 16 физических классов) [21]			
Доля верно кластеризованных точек	70,1251 % ( $k = 1200$ )	67,7142 % ( $k = 290$ )	72,4750 % ( $t = 0,73$ )
Требуемый объем памяти	1,6 GB	98 MB	66 MB
Время обработки	8 мин. 53 сек.	41 сек.	10 мин. 16 сек.
Изображение <i>Pavia University</i> (610 строк, 340 столбцов, 103 каналов, 13 физических классов) [21]			
Доля верно кластеризованных точек	65,43 % ( $k = 1500$ )	63,17 % ( $k = 350$ )	68,01 % ( $t = 0,87$ )
Требуемый объем памяти	3 GB	178 MB	102 MB
Время обработки	15 мин. 8 сек.	1 мин. 38 сек.	18 мин. 27 сек.

**Примечание.** Экспериментальные исследования проводились на персональном компьютере со следующими характеристиками:

- процессор – Intel Core i7, 3.6 ГГц;
- объем оперативного запоминающего устройства – 32 Гб
- операционная система – Windows 7 Профессиональная (64-разрядная).

### Литература

1. Загородная А.А. Опыт применения материалов ДЗЗ в работах экологической тематики / Н.Б. Пыстина, А.А. Загородная // Защита окружающей среды в нефтегазовом комплексе. – 2005. № 9. – С. 42-48.
2. Сай С.В. Алгоритмы морфологической кластеризации растительности на базе космических снимков применительно к территории г. Хабаровска / С.В. Сай, Г.Я. Маркелов, С.В. Плесовских // Вестник Тихоокеанского государственного университета. 2017 № 3 (46). – С. 13-22.
3. Клейн Н.А. Анализ подтопленных сельскохозяйственных территорий в Западной Сибири с использованием ДЗЗ / Н.А. Клейн, Л.В. Березин, М.С. Балуков // Инновации в природообустройстве и защите в чрезвычайных ситуациях: Материалы IV международной научно-практической конференции. – 2018. – С. 88-91.
4. Кениг А.В. Применение технологии ГИС и методов ДЗЗ в системе выявления, мониторинга и охраны объектов археологического наследия / А.В. Кениг, Е.А. Зайцева // ИНТЕРЭКСПО ГЕО-Сибирь. – 2013. – Т. 8. – С. 48-51.
5. Уфимцев А.Е. Изучение структуры природных и антропогенных ландшафтов с применением ГИС и данных ДЗЗ / А.Е. Уфимцев, О.Ю. Вавер // Бассейновые территории: проблемы и пути их решения: Материалы международной научно-практической конференции. редактор-составитель Г.С. Кощеева. 2013. – 2016. – С. 56-60.
6. Абросимов А.В. Перспективы применения данных ДЗЗ из космоса для повышения эффективности сельского хозяйства в России / А.В. Абросимов, Б.А. Дворкин // Геоматика. – 2009. – № 4. – С. 46-49.
7. Третьяков В.А., Основные тенденции развития гиперспектральной аппаратуры в мире / В.А. Третьяков // Космонавтика и Ракетостроение. 2013. № 4 (73). С. 36-40.
8. Advances in Hyperspectral Image and Signal Processing: A Comprehensive Overview of the State of the Art. / P. Ghamisi, N. Yokoya, J. Li, W. Liao, S. Liu, J. Plaza, B. Rasti, A. Plaza // IEEE Geoscience and Remote Sensing Magazine. – 2017. – December. – Vol. 5, no. 4 – Pp. 37-78. DOI:10.1109/mgrs.2017.2762087.
9. Hyperspectral Remote Sensing Classifications: A Perspective Survey / D. Chutia, D.K. Bhattacharyya, K.K. Sarma, R. Kalita, S. Sudhakar // Transactions in GIS. – 2016. – August. – Vol. 20, no. 4. – Pp. 463-490.
10. Zhang S.L. A Study of Image Classification of Remote Sensing Based on Back-Propagation Neural Network with Extended Delta Bar Delta / S.L. Zhang, T.C. Chang // Mathematical Problems in Engineering. – 2015. – October. – DOI: 10.1155/2015/178598.
11. Convolutional Neural Networks Based Hyperspectral Image Classification Method with Adaptive Kernels / C. Ding, Y. Li, Y. Xia, W. Wei, L. Zhang, Y. Zhang // – 2017. – June. – Vol. 9, no. 6.
12. Towards a completely blind classifier for hyperspectral images / P. Halle; S. Le Moan, C. Cariou // International Conference on Image and Vision Computing New Zealand (IVCNZ). – 2017.
13. Черезов Д.С. Обзор основных методов классификации и кластеризации данных / Д.С.Черезов, Н.А. Тюкачев // Вестник ВГУ, серия: системный анализ и информационные технологии. – 2009. – № 2. С. 25-29.
14. Arthur D. How slow is the k-means method? / D. Art-hur, S. Vassilivskii // Proceedings of the 2006 Symposium on Computational Geometry. – 2006. – Pp. 144-153.
15. Fuzzy c-means algorithm for segmentation of aerial photography data obtained using unmanned aerial vehicle / M.V. Akinin, N.V. Akinina, A.Y. Klochkov, M.B. Nikiforov, A.V. Sokolova // International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. – 2015. – Vol. XL-5/W6. – Pp. 113-115.
16. DBSCAN revisited, revisited: why and how you should (still) use DBSCAN / E. Schubert, J. Sander, M. Ester, H.P. Kriegel, X. Xu // ACM transactions on database systems. – 2017. – Vol. 42 № 3 – Pp. 1-21.
17. OPTICS: ordering points to identify the clustering structure / M. Ankerst, M. M. Breunig, H.-P. Kriegel, J. Sander // Proc. 1999 ACM SIGMOD Intern. Conf. on Management of data. –1999. – Pp. 49-60.
18. Cariou C. A new k-nearest neighbor density-based clustering method and its application to hyperspectral images / C. Cariou, K. Chehdi // International Geoscience and Remote Sensing Symposium (IGARSS). – 2016.
19. Cariou C. Nearest neighbor-density-based clustering methods for large hyperspectral images / C. Cariou, K. Chehdi // Proc. SPIE 10427, Image and Signal Processing for Remote Sensing XXIII. – 2017. – October. – DOI: 10.1117/12.2278221.
20. Cariou C. Application of unsupervised nearest neighbor density-based approaches to sequential dimensionality reduction and clustering of hyperspectral images / C. Cariou, K. Chehdi // Proc. SPIE 10789, Image and Signal Processing for Remote Sensing. – 2018. – October.
21. [http://www.ehu.es/ccwintco/index.php?title=Hyper-spectral Remote Sensing Scenes.](http://www.ehu.es/ccwintco/index.php?title=Hyper-spectral%20Remote%20Sensing%20Scenes)