

УДК 004.932.2; 004.93'14; 528.854.2

ИЕРАРХИЧЕСКИЕ АЛГОРИТМЫ КЛАСТЕРИЗАЦИИ ДЛЯ СЕГМЕНТАЦИИ МУЛЬТИСПЕКТРАЛЬНЫХ ИЗОБРАЖЕНИЙ*

И. А. Пестунов¹, С. А. Рылов¹, В. Б. Бериков²

¹*Институт вычислительных технологий СО РАН,
630090, г. Новосибирск, просп. Академика Лаврентьева, 6*

²*Институт математики им. С. Л. Соболева СО РАН,
630090, г. Новосибирск, просп. Академика Коптюга, 4
E-mail: pestunov@ict.nsc.ru*

В рамках сеточного и ансамблевого подходов разработаны вычислительно эффективные иерархические алгоритмы кластеризации НСА и НЕСА для сегментации мультиспектральных изображений. Предложена специальная метрика, позволяющая выделять вложенные кластеры даже при наличии пересечений. Представлены результаты экспериментальных исследований алгоритмов на модельных и реальных данных, подтверждающие их эффективность.

Ключевые слова: иерархический ансамблевый алгоритм кластеризации, сеточный подход, сегментация мультиспектральных спутниковых изображений.

Введение. Сегментация является одним из важнейших этапов анализа цифровых изображений [1, 2]. Она заключается в разбиении изображения на непересекающиеся области на основе схожести их спектральных и/или пространственных характеристик (текстура, размер, форма и т. п.). Наиболее распространённый подход к сегментации спутниковых изображений основан на использовании алгоритмов кластеризации данных [3].

Методы кластеризации можно разделить на две большие группы: иерархические и неиерархические. Неиерархические алгоритмы позволяют получить фиксированное разбиение данных, а иерархические — систему вложенных разбиений, соответствующих различным уровням иерархии. Иерархическое представление удобно при интерпретации результатов, когда требуется информация о различных уровнях кластерной структуры, а также в ситуации, когда точное число искомых кластеров неизвестно. Традиционные методы иерархической кластеризации имеют ряд недостатков. Например, метод одиночной связи подвержен так называемому «цепочному эффекту», а методы полной и средней связи обычно хорошо работают только с кластерами сферической формы. Кроме того, эти методы не позволяют разделять пересекающиеся кластеры [4]. Другим серьёзным ограничением для таких методов является высокая вычислительная трудоёмкость, не позволяющая использовать их для обработки больших массивов данных, например мультиспектральных изображений.

В последнее время для повышения устойчивости и качества результатов кластеризации активно применяется ансамблевый подход [5–10]. Однако методам, использующим ансамбли иерархических разбиений, посвящены лишь отдельные работы [11, 12]. При этом применяемые в них алгоритмы также являются вычислительно трудоёмкими.

Цель данной работы — создание и исследование вычислительно эффективных иерархических алгоритмов кластеризации НСА и НЕСА, предназначенных для сегментации

*Работа выполнена при поддержке Российского фонда фундаментальных исследований (гранты № 14-07-31320-мол.а, № 14-07-00249-а) и Российского научного фонда (грант № 14-14-00453).

мультиспектральных спутниковых изображений. Алгоритмы обеспечивают разделение пересекающихся кластеров, а также кластеров сложной формы, разного размера и плотности. Предлагаемая работа направлена на комбинирование сеточного и ансамблевого подходов к кластеризации и является продолжением исследований авторов [13–16].

Сеточный иерархический алгоритм кластеризации НСА. Рассматриваемый иерархический алгоритм кластеризации НСА основывается на алгоритме ССА [13], который сочетает в себе преимущества сеточного [17] и плотностного подходов: высокую вычислительную эффективность и способность выделять кластеры, имеющие сложную структуру. Для описания предлагаемого алгоритма введём несколько определений.

Пусть множество классифицируемых объектов X состоит из векторов, лежащих в пространстве признаков R^d : $X = \{x_i = (x_i^1, \dots, x_i^d) \in R^d, i = \overline{1, N}\}$. Векторы x_i лежат в прямоугольном гиперпараллелепипеде $\Omega = [l^1, r^1] \times \dots \times [l^d, r^d]$, где $l^j = \min_{x_i \in X} x_i^j$, $r^j = \max_{x_i \in X} x_i^j$. Под сеточной структурой будем понимать разбиение пространства признаков гиперплоскостями: $x^j = (r^j - l^j)i/m + l^j$, $i = 0, \dots, m$, m — число разбиений Ω по каждой размерности. Минимальным элементом этой структуры является клетка (замкнутый прямоугольный гиперпараллелепипед, ограниченный гиперплоскостями). Введём общую нумерацию клеток (последовательно от одного слоя клеток к другому). Клетки B_i и B_j ($i \neq j$) являются смежными, если их пересечение не пусто. Множество смежных с B клеток обозначим через A_B . Плотностью D_B клетки B назовём отношение $D_B = N_B/V_B$, где N_B — количество элементов множества X , попавших в клетку B ; V_B — объём клетки B . Клетку B будем считать непустой, если $D_B > 0$.

Непустая клетка B_i непосредственно связана с непустой клеткой B_j ($B_i \rightarrow B_j$), если B_j — максимальная по номеру клетка, удовлетворяющая условиям $B_j = \arg \max_{B_k \in A_{B_i}} D_{B_k}$ и

$D_{B_j} \geq D_{B_i}$. Непустые смежные клетки B_i и B_j непосредственно связаны ($B_i \leftrightarrow B_j$), если $B_i \rightarrow B_j$ или $B_j \rightarrow B_i$. Непустые клетки B_i и B_j связаны ($B_i \sim B_j$), если существуют k_1, \dots, k_l такие, что $k_1 = i$, $k_l = j$, и для всех $p = 1, \dots, l - 1$ выполнено $B_{k_p} \leftrightarrow B_{k_{p+1}}$.

Введение отношения связности порождает естественное разбиение множества непустых клеток на компоненты связности $\{G_1, \dots, G_S\}$. Под компонентой связности будем понимать максимальное множество попарно связанных клеток. Представителем компоненты связности G назовём максимальную по номеру клетку $Y(G)$, удовлетворяющую условию $Y(G) = \arg \max_{B \in G} D_B$.

В предлагаемом алгоритме НСА в качестве элементов иерархии выступают не элементы исходных данных, а введённые выше компоненты связности. Благодаря использованию сеточной структуры число получаемых компонент связности относительно мало, поэтому при построении иерархии не требуется больших вычислительных затрат.

Для построения иерархического сеточного алгоритма кластеризации введём метрику между компонентами связности. Компоненты G_i и G_j будем считать смежными, если существуют смежные клетки B_i и B_j такие, что $B_i \in G_i$ и $B_j \in G_j$. Определим расстояние между смежными компонентами связности G_i и G_j по формуле

$$h_{ij} = \min_{P_{ij} \in \mathfrak{R}_{ij}} \left[1 - \min_{B_{k_t} \in P_{ij}} D_{B_{k_t}} / \min(D_{Y_i}, D_{Y_j}) \right].$$

Здесь $\mathfrak{R}_{ij} = \{P_{ij}\}$ — множество всех цепей между представителями компонент связности $P_{ij} = \langle Y_i = B_{k_1}, \dots, B_{k_t}, B_{k_{t+1}}, \dots, B_{k_l} = Y_j \rangle$ таких, что для всех $t = 1, \dots, l - 1$: 1) $B_{k_t} \in G_i \cup G_j$, 2) $B_{k_t}, B_{k_{t+1}}$ — смежные клетки.

Построим матрицу расстояний между компонентами связности на основе расстояний между смежными компонентами связности $\{h_{ij}\}$ следующим образом. Пусть $\Theta(G_i, G_j) =$

$= \Theta_{ij} = \{Q_{ij}\}$ — множество всех цепей из компонент связности $Q_{ij} = \langle G_i = G_{k_1}, \dots, \dots, G_{k_t}, G_{k_{t+1}}, \dots, G_{k_l} = G_j \rangle$ таких, что для всех $t = 1, \dots, l - 1$ компоненты $G_{k_t}, G_{k_{t+1}}$ смежные. Тогда расстояние между произвольными компонентами связности G_i и G_j будет вычисляться по формуле

$$\tilde{h}_{ij} = \min_{Q_{ij} \in \Theta_{ij}} \left[\max_t h_{k_t, k_{t+1}} \right].$$

В случае если множество Θ_{ij} пусто, то полагаем $\tilde{h}_{ij} = 1$.

Для фиксированной цепочки из компонент связности $Q_{ij} = \langle G_i = G_{k_1}, \dots, G_{k_t}, G_{k_{t+1}}, \dots, G_{k_l} = G_j \rangle$ определим длину как $d(Q_{ij}) = \max_t h_{k_t, k_{t+1}}$. Тогда расстояние между компонентами связности G_i и G_j можно переписать в виде

$$\tilde{h}_{ij} = \min_{Q_{ij} \in \Theta_{ij}} [d(Q_{ij})].$$

Введённое расстояние на основе оценки плотности распределения данных позволяет избежать присущей иерархическим методам проблемы пересекающихся классов [18].

Легко показать, что введённое отношение \tilde{h}_{ij} является ультраметрикой [19], т. е. метрикой, для которой выполняется усиленное неравенство треугольника: $\tilde{h}_{ij} \leq \max(\tilde{h}_{ik}, \tilde{h}_{kj})$, $\forall i, j, k$. Существует однозначное соответствие между матрицами расстояний со свойством ультраметрики и дендрограммами [11], поэтому их можно использовать в качестве дескрипторов для иерархических результатов.

Вычисление ультраметрики $\{h_{ij}\}$ из матрицы расстояний $\{h_{ij}\}$ является операцией минимального транзитивного замыкания [12]. Для её реализации зачастую применяются вычислительно трудоёмкие алгоритмы. Например, в работе [11] используется метод умножения матриц, вычислительная сложность которого составляет $O(n^4)$ (при матрице размера $n \times n$). В [12] применяется модификация алгоритма Флойда — Уоршелла с вычислительной сложностью $O(n^3)$. В [20] предлагается использовать рекурсивный проход всех цепочек поиском в глубину (ширину), начиная с каждого элемента. Этот подход хорошо работает на разреженных графах, но на плотных графах его сложность достигает $O(n^3)$.

В данной работе операция минимального транзитивного замыкания выполняется с помощью алгоритма построения дендрограммы методом одиночной связи (SLINK) [21], который имеет сложность $O(n^2)$. Следующее утверждение показывает, что применение метода одиночной связи к матрице расстояний $\{h_{ij}\}$ даёт в результате дендрограмму, эквивалентную искомой ультраметрике.

Утверждение. Дендрограмма, построенная по ультраметрике $\{\tilde{h}_{ij}\}$, совпадает с дендрограммой, сформированной по матрице расстояний $\{h_{ij}\}$ методом одиночной связи.

Доказательство. Покажем, что для каждого уровня дендрограммы мы получим одинаковые отношения эквивалентности (принадлежности к одному кластеру) \tilde{R}_α и R_α .

Выберем некоторое значение $\alpha \in [0; 1]$. В соответствии с методом одиночной связи элементы i и j принадлежат к одному кластеру тогда и только тогда, когда существует цепочка $Q_{ij} \in \Theta_{ij}$, в которой длины всех отрезков не превышают α , т. е. $d(Q_{ij}) \leq \alpha$.

Предположим, что по матрицам $\{h_{ij}\}$ и $\{\tilde{h}_{ij}\}$ построены две различные дендрограммы. Пусть R_α и \tilde{R}_α — отношения эквивалентности, полученные в результате среза этих дендрограмм на уровнях, соответствующих пороговому значению α .

Возможны два случая.

1. Пусть i и j принадлежат к одному кластеру по отношению R_α :

$$iR_\alpha j \Leftrightarrow \exists Q_{ij} \in \Theta_{ij} : d(Q_{ij}) \leq \alpha \Rightarrow \tilde{h}_{ij} \leq \alpha \Rightarrow i\tilde{R}_\alpha j.$$

Таким образом, i и j принадлежат к одному кластеру по отношению \tilde{R}_α .

2. Пусть i и j принадлежат к разным кластерам по отношению R_α :

$$\begin{aligned} \neg(iR_\alpha j) &\Leftrightarrow \forall Q_{ij} \in \Theta_{ij} : d(Q_{ij}) > \alpha \Rightarrow \tilde{h}_{ij} = \min_{Q_{ij} \in \Theta_{ij}} [d(Q_{ij})] > \alpha \Rightarrow \\ &\Rightarrow \forall Q_{ij} \in \Theta_{ij} : \tilde{d}(Q_{ij}) \geq \tilde{h}_{ij} > \alpha \Rightarrow \neg(i\tilde{R}_\alpha j). \end{aligned}$$

Таким образом, i и j принадлежат к разным кластерам по отношению \tilde{R}_α .

Следовательно, отношения \tilde{R}_α и R_α совпадают для каждого значения α , т. е. построенные дендрограммы полностью совпадают.

Предлагаемый иерархический сеточный алгоритм НСА(m) можно записать в виде последовательности основных шагов.

Шаг 1. Формирование клеточной структуры. На этом этапе для каждой точки $x_i \in X$ определяется содержащая её клетка и вычисляются плотности D_B всех клеток.

Шаг 2. Выделение компонент связности $\{G_1, \dots, G_S\}$ и их представителей $Y(G_1), \dots, Y(G_S)$.

Шаг 3. Вычисление матрицы расстояний $\{h_{ij}\}$ между смежными компонентами связности в соответствии с указанным выше определением.

Шаг 4. Построение дендрограммы по матрице расстояний $\{h_{ij}\}$ методом одиночной связи.

Алгоритм НСА при низких вычислительных затратах позволяет выделять сложные многомодовые кластеры, а также предоставляет пользователю информацию об иерархической структуре данных. На рис. 1 изображена модель 1 «Бананы», состоящая из 400 двумерных точек, сгруппированных в два линейно неразделимых класса. Модель построена с помощью инструментария PRTools [22] с параметром 0,7. Однако экспериментальные исследования показали, что результаты работы предложенного алгоритма существенно зависят от параметра m , который определяет масштаб элементов сеточной структуры. На рис. 2 представлен график зависимости точности кластеризации от параметра сетки m при применении алгоритма НСА к модели «Бананы».

Ансамблевый иерархический алгоритм кластеризации НЕСА. В последние годы ансамблевые методы кластеризации активно применяются для повышения устойчивости результатов [8]. С помощью ансамблевого подхода можно добиться существенного улучшения результатов работы алгоритма НСА.

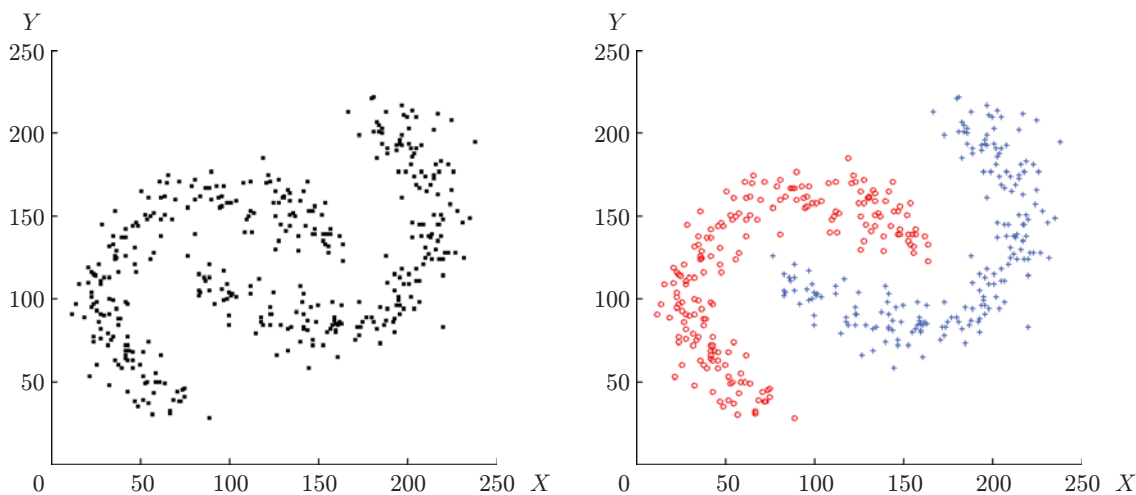


Рис. 1. Модель «Бананы» (слева) и результат эталонной кластеризации (справа)

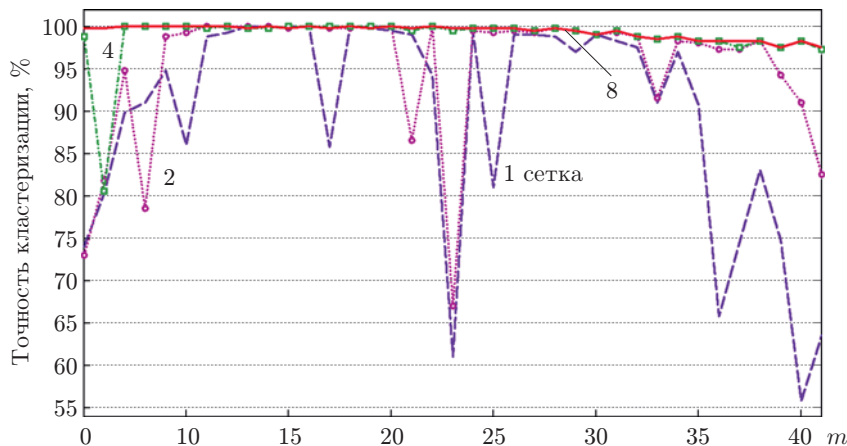


Рис. 2. График зависимости точности кластеризации от параметра начальной сетки m на модели «Бананы» для алгоритма НСА (1 сетка) и для ансамблевого алгоритма НЕСА при различном числе используемых сеток (2, 4, 8)

Проблема неустойчивости результатов алгоритма НСА к изменению масштаба сетки связана с тем, что при мелкой сетке возникает большое число компонент связности, возрастает влияние «шума» и теряются связи между близкими компонентами. С другой стороны, при крупной сетке возникают проблемы с разделением пересекающихся классов (рис. 3). Поэтому, комбинируя с помощью ансамблевого метода разномасштабную информацию, можно существенно повысить устойчивость результатов.

В данной работе предлагается ансамблевый иерархический алгоритм кластеризации НЕСА, в котором в качестве элементов ансамбля выступают результаты алгоритма НСА при фиксированных значениях параметра сетки m . Для построения коллективного решения используется согласованная матрица расстояний, полученная путём усреднения матриц расстояний по всем элементам ансамбля. Теоретическое обоснование этого способа построения коллективного решения приведено в [15].

Таким образом, для формирования ансамбля предложенный выше сеточный алгоритм НСА запускается L раз при различных значениях параметра сетки m . В результате получается L матриц расстояний между компонентами связности $\{\tilde{h}_{ij}^{(1)}\}, \dots, \{\tilde{h}_{ij}^{(L)}\}$. Заметим, что на данном этапе матрицы должны обладать свойством ультраметрики, т. е. должна быть выполнена операция транзитивного замыкания. Согласованная матрица расстояний $\{H_{ij}\}$ (размер которой совпадает с размером матрицы по самой мелкой сетке $\{\tilde{h}_{ij}^{(L)}\}$) строится следующим образом:

$$H_{ij} = \frac{1}{L} \sum_{k=1}^L \tilde{h}^{(k)}(G_i^{(k)}, G_j^{(k)}),$$

где $G_i^{(k)}$ является компонентой на сетке с номером k , в которой находится клетка-представитель компоненты $G_i^{(L)}$ из сетки с номером L . Так, с помощью представителей компонент связности решается проблема соответствия меток кластеров в ансамбле.

Полученная согласованная матрица расстояний $\{H_{ij}\}$ уже может не обладать свойством ультраметрики. К сожалению, поиск «средней» ультраметрики (наиболее близкой одновременно ко всем исходным ультраметрикам) является NP-трудной задачей [11]. Поэтому для получения окончательного иерархического результата к согласованной матрице

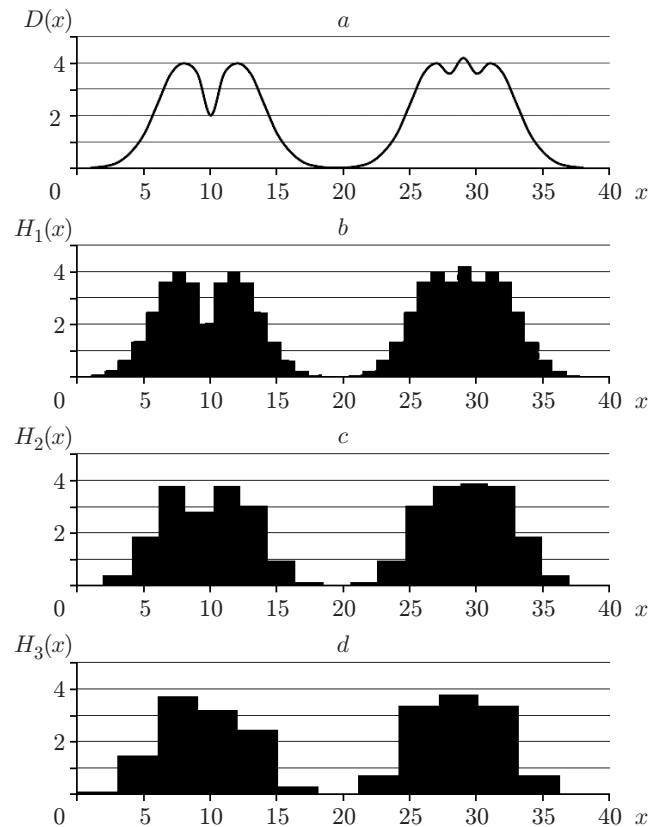


Рис. 3. Пример влияния размера сетки на разделимость пересекающихся кластеров: a — плотность распределения некоторой случайной величины, b – d — гистограммы при различных уровнях дискретизации

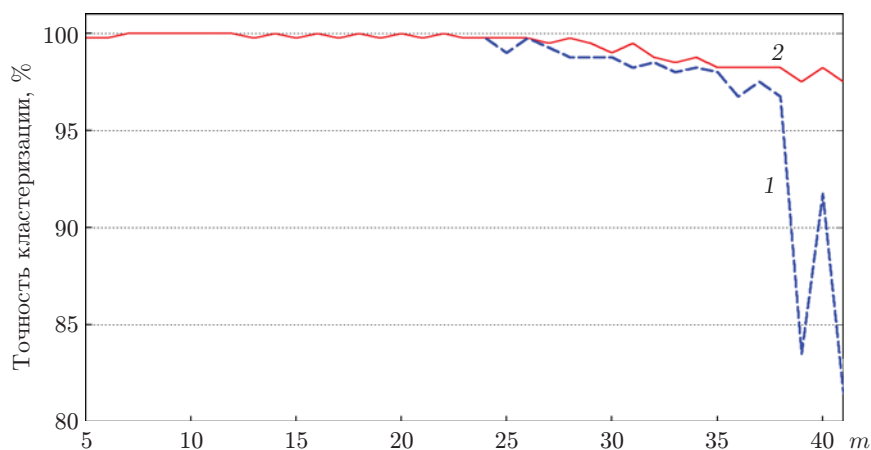


Рис. 4. График зависимости точности кластеризации от параметра сетки m для модели «Бананы» при построении дендрограммы методами одиночной (кривая 1) и средней (кривая 2) связей в ансамблевом алгоритме НЕСА при использовании восьми сеток

расстояний $\{H_{ij}\}$ применяется метод построения дендрограммы. В работе [23] отмечается, что метод невзвешенной средней связи (UPGMA) к среднему арифметическому матриц ультраметрик очень часто даёт результат, совпадающий с результатами трудоёмких оптимизационных методов по поиску ближайшей ультраметрики. Проведённые экспериментальные исследования показали превосходство метода UPGMA над методом SLINK в предложенном алгоритме. На рис. 4 приведён график зависимости точности кластеризации от параметра сетки m при использовании метода средней связи и метода одиночной связи на последнем этапе алгоритма HECA.

Экспериментальные исследования подтверждают, что применение ансамблевого подхода позволяет существенно повысить качество результатов кластеризации и их устойчивость к изменению параметра сетки. График на рис. 2 демонстрирует повышение точности результатов кластеризации при увеличении числа используемых сеток (параметр сетки m выбирался из набора $\{m, m + 2, \dots, m + 2(L - 1)\}$).

Результаты экспериментальных исследований. Многочисленные эксперименты на модельных данных показали способность алгоритма выявлять в них сложные структуры. Во всех экспериментах использовался ансамбль из восьми элементов.

Сравнение качества работы алгоритма HECA с другими алгоритмами проводилось с помощью открытого программного пакета ELKI [24], включающего в себя такие известные алгоритмы кластеризации, как k -means, EM, DBSCAN, OPTICS, DeLiClu, SLINK.

На рис. 5, *a* представлена модель 2 из восьми нормально распределённых классов. Классы сгруппированы в три изолированные группы, в одной из которых имеются значительные пересечения, что существенно затрудняет их разделение. Алгоритм HECA позволяет не только эффективно разделить все восемь классов (рис. 5, *b*), но и выявить иерархическую структуру данных (рис. 5, *b—d*).

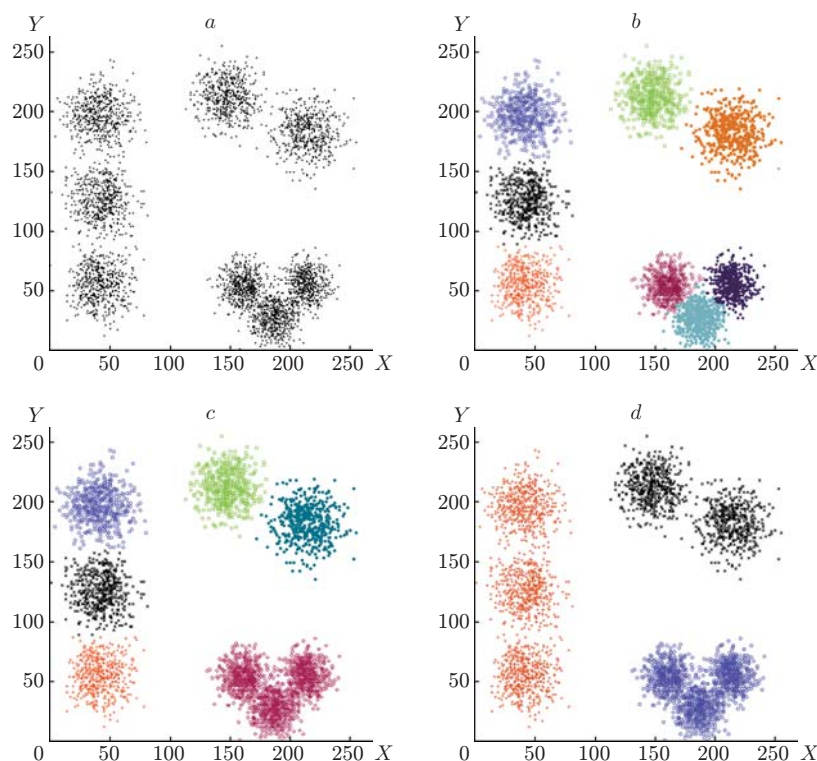


Рис. 5. Эксперимент с моделью, состоящей из нормально распределённых классов: *a* — исходные данные; *b—d* — результаты кластеризации с помощью предложенного ансамблевого алгоритма при значениях параметра объединения дендрограммы 0,4, 0,6 и 0,95 соответственно

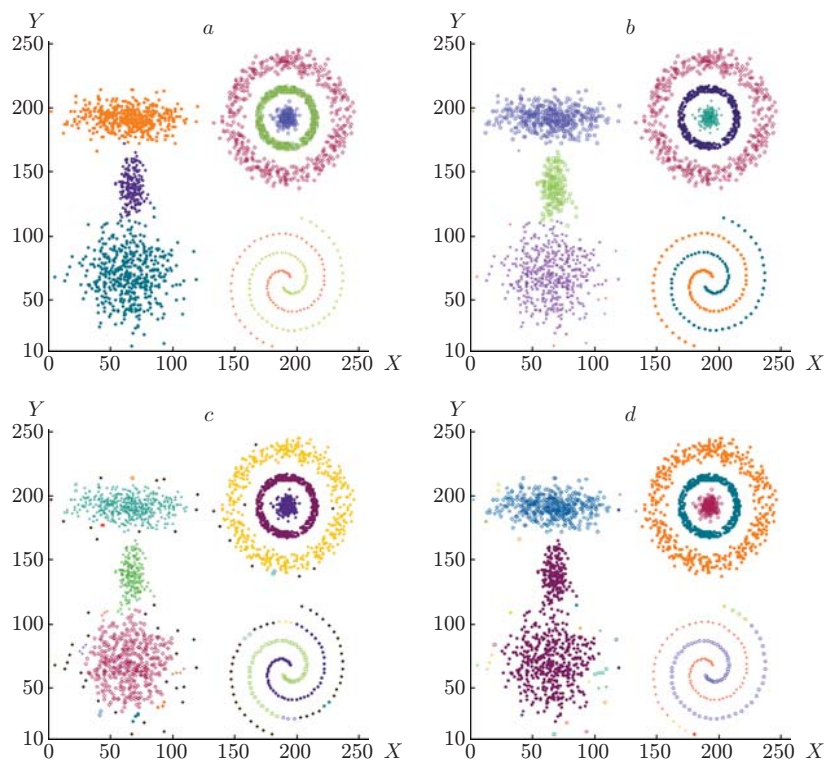


Рис. 6. Эксперимент с моделью, состоящей из восьми классов, различающихся по форме, размеру и плотности: *a* — исходные данные; *b–d* — результаты кластеризации алгоритмами HECA, OPTICS/DeLiClu и SLINK соответственно

Такие алгоритмы, как k -means и EM, предназначенные для разделения нормально распределённых классов, успешно справляются с данной моделью лишь при удачной инициализации центров. В то же время плотностные методы (DBSCAN, OPTICS и DeLiClu), а также традиционные иерархические методы (одиночной связи, дальней связи, взвешенной средней связи) не способны разделить сильно пересекающуюся группу кластеров. В результате их применения получается шесть или менее кластеров в зависимости от значений выбранных параметров (см. рис. 5, *c, d*).

Рассматриваемая модель 3, представленная на рис. 6, *a*, состоит из восьми кластеров, различающихся по форме, размеру и плотности, включая кольцевые, спиралевидные

Сравнение времени работы алгоритмов кластеризации
(использовались наиболее оптимальные параметры, время указано в секундах)

Алгоритм	Модель 1 (400 точек) (см. рис. 1)	Модель 2 (4000 точек) (см. рис. 5)	Модель 3 (9388 точек) (см. рис. 6)	Спутниковый снимок ALOS (50000 пикселей)	RGB-изображение (265776 пикселей)
HECA	0,005	0,003	0,055	0,07	0,03
k -means Lloyd	0,03	0,08	0,06	0,6	1,0
EM	0,65	18,4	1,45	43	78
DBSCAN	0,06	0,35	1,3	25	136
OPTICS	0,03	0,4	1,8	55	390
DeLiClu	0,19	0,73	2,0	35	430
SLINK	0,01	0,42	2,2	96	2752

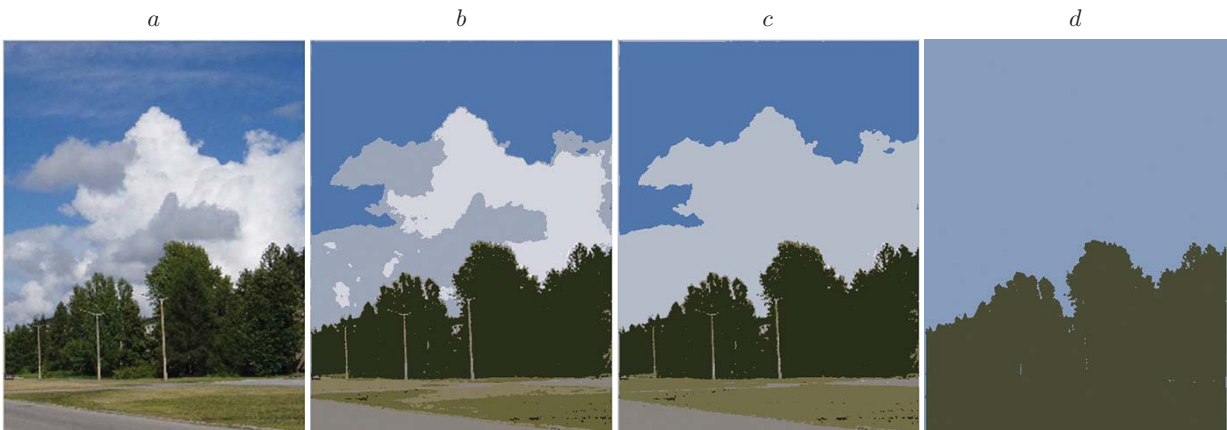


Рис. 7. Эксперимент с RGB-изображением: *a* — исходное изображение; *b–d* — результаты кластеризации алгоритмом HECA при разных уровнях иерархии

и нормально распределённые классы [25]. При использовании достаточно мелких сеток ($m \geq 60$) ансамблевый алгоритм HECA успешно выделяет все кластеры и достигает точности 99,3 % (рис. 6, *b*). Однако ни один из алгоритмов из пакета ELKI не смог качественно выделить все восемь кластеров. Из иерархических алгоритмов для этой модели наилучшие результаты продемонстрировал метод SLINK: он позволил выделить классы спиралевидной формы, но при этом возникло затруднение с разделением нормально распределённых классов (рис. 6, *d*). Лучший результат (94,6 %) показали алгоритмы OPTICS и DeLiClu. Однако они не смогли выделить кластеры спиралевидной формы (рис. 6, *c*).

Важным преимуществом предложенного алгоритма HECA является его высокая вычислительная эффективность. В таблице приведено сравнение времени работы алгоритма

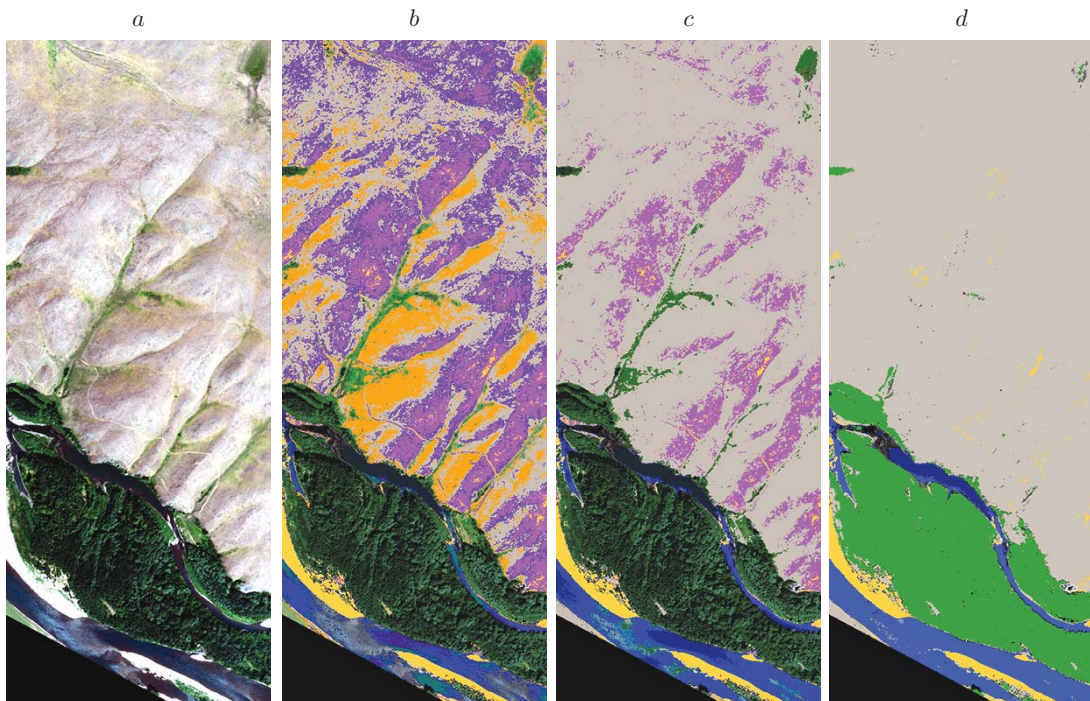


Рис. 8. Эксперимент со спутниковым изображением WorldView-2: *a* — RGB-компози́т (каналы 5, 3, 2) исходного изображения; *b–d* — результаты кластеризации иерархическим ансамблевым алгоритмом при разных уровнях иерархии

НЕСА и алгоритмов из программного пакета ELKI [24] на описанных выше моделях и реальных изображениях (цветное RGB-изображение размером 452×588 пикселей и фрагмент четырёхканального изображения размером 250×200 пикселей, полученного со спутника ALOS). Хотя этот пакет не гарантирует максимальную скорость работы включённых в него алгоритмов, он является наиболее эффективным из доступных средств, в частности, благодаря возможности использования R^* -дерева для индексации. Предложенный алгоритм НЕСА и комплекс ELKI реализованы на языке программирования Java. Обработка проводилась на четырёхъядерной ПЭВМ Intel Core i7 с тактовой частотой 3,2 ГГц и объёмом оперативной памяти 8 Гбайт.

На рис. 7 показан пример обработки цветного изображения размером 452×588 пикселей, время обработки 0,03 с. Пример обработки снимка, полученного со спутника WorldView-2, с использованием алгоритма НЕСА приведён на рис. 8. При кластеризации применялись четыре канала: 2, 3, 5, 7. Размер изображения 600×1552 пикселей, время обработки 0,7 с.

Заключение. В представленной работе предложены иерархические алгоритмы кластеризации НСА и НЕСА для сегментации спутниковых изображений. Результаты приведённых экспериментов на модельных и реальных данных подтверждают высокое качество получаемых решений и их устойчивость к изменению настраиваемых параметров. Возможность построения иерархической системы вложенных кластеров значительно облегчает процесс интерпретации результатов. Высокое быстродействие алгоритмов позволяет проводить обработку мультиспектральных изображений в диалоговом режиме.

Необходимо заметить, что использование сеточной структуры вводит ограничение на размерность обрабатываемых данных. Реализованные алгоритмы эффективно работают с мультиспектральными изображениями, содержащими до восьми спектральных каналов.

В дальнейшем планируется адаптация рассмотренных алгоритмов для реализации на графических процессорах, а также преодоление ограничений, вызванных использованием сеточной структуры.

СПИСОК ЛИТЕРАТУРЫ

1. Гонсалес Р., Вудс М. Цифровая обработка изображений. М.: Техносфера, 2006. 812 с.
2. Чочиа П. А. Сегментация изображений на основе анализа расстояний в пространстве признаков // Автометрия. 2014. **50**, № 6. С. 97–110.
3. Пестунов И. А., Синявский Ю. Н. Алгоритмы кластеризации в задачах сегментации спутниковых изображений // Вестн. КемГУ. 2012. **2**, № 4(52). С. 110–125.
4. Xu R., Wunsch D. I. Survey of clustering algorithms // IEEE Trans. Neural Networks. 2005. **16**, N 3. P. 645–678.
5. Jain A. K. Data clustering: 50 years beyond K-means // Patt. Recogn. Lett. 2010. **31**, N 8. P. 651–666.
6. Ghaemi R., Sulaiman M., Ibrahim H., Mustapha N. A survey: Clustering ensembles techniques // World Acad. Sci., Eng. and Technol. 2009. **3**, N 2. P. 535–544.
7. Hope P., Hall L., Goldgof D. A scalable framework for cluster ensembles // Patt. Recogn. 2009. **42**, N 5. P. 676–688.
8. Kashef R., Kamel M. Cooperative clustering // Patt. Recogn. 2010. **43**, N 7. P. 2315–2329.
9. Jia J., Liu B., Jiao L. Soft spectral clustering ensemble applied to image segmentation // Front. Comput. Sci. China. 2011. **5**, N 1. P. 66–78.
10. Franek L., Jiang X. Ensemble clustering by means of clustering embedding in vectorspaces // Patt. Recogn. 2014. **47**, N 2. P. 833–842.

11. **Mirzaei A., Rahmati M.** A novel hierarchical-clustering-combination scheme based on fuzzy-similarity relations // IEEE Trans. Fuzzy Syst. 2010. **18**, N 1. P. 27–39.
12. **Zheng L., Li T., Ding C.** Hierarchical ensemble clustering // Proc. of 2010 IEEE Intern. Conf. on Data Mining. IEEE, 2010. P. 1199–1204.
13. **Куликова Е. А., Пестунов И. А., Синявский Ю. Н.** Непараметрический алгоритм кластеризации для обработки больших массивов данных // Тр. 14 Всерос. конф. «Математические методы распознавание образов». М.: МАКС Пресс, 2009. С. 149–152.
14. **Пестунов И. А., Бериков В. Б., Синявский Ю. Н.** Сегментация многоспектральных изображений на основе ансамбля непараметрических алгоритмов кластеризации // Вестн. СибГАУ. 2010. № 5(31). С. 56–64.
15. **Пестунов И. А., Бериков В. Б., Куликова Е. А., Рылов С. А.** Ансамблевый алгоритм кластеризации больших массивов данных // Автометрия. 2011. **47**, № 3. С. 49–58.
16. **Пестунов И. А., Рылов С. А.** Алгоритмы спектрально-текстурной сегментации спутниковых изображений высокого пространственного разрешения // Вестн. КемГУ. 2012. **2**, № 4(52). С. 104–110.
17. **Piango M. R., Mohan V.** A survey of grid based clustering algorithms // Intern. Journ. Eng. Sci. Technol. 2010. **2**, N 8. P. 3441–3446.
18. **Yonggang L., Yi W.** PHA: A fast potential-based hierarchical agglomerative clustering method // Patt. Recogn. 2013. **46**, N 5. P. 1227–1239.
19. **Leclerc B.** Description combinatoire des ultramétries // Math. Sci. Humaines. 1981. **127**, N 73. P. 5–37.
20. **Skiena St. S.** The Algorithm Design Manual. Springer, 2008. 730 p.
21. **Olson Cl. F.** Parallel algorithms for hierarchical clustering // Parallel Comput. 1995. **21**, Is. 8. P. 1313–1325.
22. **The Matlab Toolbox for Pattern Recognition.** URL: <http://www.prtools.org> (дата обращения: 28.01.2015).
23. **Maurizio V.** Principal classifications analysis a method for generating consensus dendrograms and its application to three-way data // Comput. Stat. & Data Anal. 1998. **27**, Is. 3. P. 311–331.
24. **Achtert E., Kriegel H., Schubert E., Zimek A.** Interactive data mining with 3D-parallel-coordinate-trees // Proc. ACM Intern. Conf. on Management of Data (SIGMOD). N. Y., 2013. P. 1009–1012.
25. **Рылов С. А.** Модель двухмерных данных для кластеризации [Электронный ресурс]. Электронные данные (1 текстовый файл). 2014. URL: https://cloud.mail.ru/public/c5f33ae275a8/TestData.Rylov_2D_Labelled_2472_elements.txt (дата обращения: 23.01.2015).

Поступила в редакцию 28 января 2015 г.