

На правах рукописи



Синявский Юрий Николаевич

**НЕПАРАМЕТРИЧЕСКИЕ МЕТОДЫ  
И ПРОГРАММНО-АЛГОРИТМИЧЕСКИЙ ИНСТРУМЕНТАРИЙ  
ДЛЯ СЕГМЕНТАЦИИ МУЛЬТИСПЕКТРАЛЬНЫХ  
СПУТНИКОВЫХ ИЗОБРАЖЕНИЙ**

05.13.18 – математическое моделирование, численные методы  
и комплексы программ

**АВТОРЕФЕРАТ**  
диссертации на соискание учёной степени  
кандидата технических наук

Новосибирск – 2021

Работа выполнена в Федеральном государственном бюджетном научном учреждении «Федеральный исследовательский центр информационных и вычислительных технологий» (ФИЦ ИВТ), г. Новосибирск.

Научный  
руководитель: кандидат физико-математических наук, доцент  
**Пестунов Игорь Алексеевич**

Официальные  
оппоненты: **Фаворская Маргарита Николаевна,**  
доктор технических наук, профессор,  
заведующий кафедрой информатики и вычислительной техники Института информатики и телекоммуникаций Сибирского государственного университета науки и технологий имени академика М.Ф. Решетнева, г. Красноярск

**Дмитриев Егор Владимирович**  
кандидат физико-математических наук, старший научный сотрудник Института вычислительной математики имени Г. И. Марчука РАН, г. Москва

Ведущая  
организация: Федеральное государственное бюджетное учреждение науки Институт динамики систем и теории управления имени В. М. Матросова Сибирского отделения Российской академии наук, г. Иркутск

Защита состоится 1 июля 2021 г. в 10:00 на заседании диссертационного совета Д 999.141.03 на базе Федерального государственного бюджетного учреждения науки Института динамики систем и теории управления им. В. М. Матросова Сибирского отделения Российской академии наук, Федерального государственного бюджетного научного учреждения «Федеральный исследовательский центр информационных и вычислительных технологий», федерального государственного бюджетного образовательного учреждения высшего образования «Сибирский государственный университет телекоммуникаций и информатики» по адресу: 630090, г. Новосибирск, пр. Академика Лаврентьева, 6, конференц-зал ФИЦ ИВТ.

С диссертацией можно ознакомиться в библиотеке и на сайте Федерального государственного бюджетного научного учреждения «Федеральный исследовательский центр информационных и вычислительных технологий»:  
<http://www.ict.nsc.ru/ru/structure/discouncil/sinavski-yn>

Автореферат разослан 29 апреля 2021 г.

Учёный секретарь диссертационного совета  
кандидат физико-математических наук, доцент



Лебедев А. С.

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

**Актуальность.** В последние десятилетия в области создания и развития средств и технологий дистанционного зондирования Земли наблюдается стремительный прогресс. С каждым годом растёт число запускаемых космических аппаратов. Как следствие, лавинообразно растут объёмы получаемых данных, которые, зачастую, являются единственным источником оперативной и объективной информации при исследовании обширных и труднодоступных территорий (в задачах, связанных с сельским и лесным хозяйством, картированием, экологией и охраной окружающей среды, прогнозированием и ликвидацией последствий чрезвычайных ситуаций и др.).

Одним из важнейших этапов анализа цифровых изображений является сегментация. Она заключается в разбиении изображения на непересекающиеся области на основе схожести их характеристик. При обработке спутниковых изображений задачу сегментации зачастую приходится решать при отсутствии каких-либо априорных сведений о количестве классов и их вероятностных характеристиках. Это значительно затрудняет применение параметрических методов (а при отсутствии информации о виде плотностей распределения вероятности их корректное применение невозможно вовсе). Для описания реальных структур данных, представленных на спутниковых изображениях, наиболее подходящими являются скользящие локально-параметрические модели, лежащие в основе непараметрических методов кластеризации. Эффективность такого рода алгоритмов подтверждена многочисленными экспериментальными исследованиями. Однако существенным недостатком непараметрических алгоритмов является их высокая вычислительная трудоёмкость. Поэтому они не находят широкого применения при решении задач, связанных с обработкой спутниковых изображений.

В настоящее время решение практических задач, связанных с сегментацией мультиспектральных изображений, производится с помощью традиционных, но, зачастую, устаревших, методов обработки, включённых в состав распространённых программных пакетов обработки спутниковых данных (ERDAS Imagine, ITTVIS ENVI, eCognition, ESRI ArcGIS, QGIS, SAGA GIS и др.). Эти методы реализуют лишь некоторые из стандартных подходов к кластеризации и не учитывают специфику спутниковых данных. В то же время более эффективные алгоритмы обработки остаются невостребованными из-за сложностей, связанных с их внедрением (закрытость распространённых программных пакетов и/или необходимость реализации модулей расширения на внутреннем языке программирования), а также высокой стоимостью коммерческих и ограниченной функциональностью открытых пакетов обработки.

Таким образом, наряду с созданием новых методов и алгоритмов для анализа и обработки спутниковых изображений, необходимо уделить внимание разработке удобного механизма обеспечения доступа к ним.

**Цель работы** – разработка эффективных непараметрических алгоритмов сегментации спутниковых изображений и современной платформы для стандартизованного доступа к ним.

Для достижения этой цели потребовалось решить следующие **основные задачи**:

- провести анализ алгоритмов кластеризации данных применительно к задаче сегментации мультиспектральных спутниковых изображений;
- разработать, теоретически обосновать и программно реализовать вычислительно эффективные непараметрические алгоритмы кластеризации для сегментации спутниковых изображений;
- выполнить экспериментальное сравнение разработанных и описанных в литературе алгоритмов на модельных данных и мультиспектральных изображениях;
- создать современную платформу для обеспечения стандартизованного доступа к алгоритмам сегментации мультиспектральных изображений, удобную как разработчикам, так и потенциальным пользователям.

**Объектом исследования** являются данные, которые представлены в виде мультиспектрального изображения, полученного при помощи многозональной оптико-электронной спутниковой системы.

**Методы исследования** опираются на современные информационно-вычислительные технологии, предусматривающие использование адекватных математических моделей изучаемого объекта и эффективных вычислительных алгоритмов. Для анализа мультиспектральных изображений используются методы теории вероятности и математической статистики. Для исследования разработанных алгоритмов применяются методы Монте – Карло. Программная реализация алгоритмов выполняется в рамках парадигмы объектно-ориентированного программирования с использованием стандартных паттернов проектирования Interface, Bridge, Prototype, Iterator и Template Method. Экспериментальные исследования выполнялись с использованием пакета программ «Image Processing Toolkit», разработанного автором.

**На защиту выносятся следующие положения**, соответствующие пунктам 3 («разработка, обоснование и тестирование эффективных вычислительных методов с применением современных компьютерных технологий»), 4 («реализация эффективных численных методов и алгоритмов в виде комплексов проблемно-ориентированных программ для проведения вычислительного эксперимента») и 8 («разработка систем компьютерного и имитационного моделирования») паспорта специальности 05.13.18 – «Математическое моделирование, численные методы и комплексы программ» (технические науки).

1. Вычислительно эффективный алгоритм кластеризации MeanSC для сегментации мультиспектральных изображений, разработанный на основе непараметрических оценок Розенблатта – Парзена с учётом характерных особенностей спутниковых снимков.
2. Непараметрический ансамблевый алгоритм кластеризации EMeanSC для сегментации мультиспектральных изображений, построенный на основе согласованной матрицы различий.
3. Пакет программ «Image Processing Toolkit», включающий оригинальный программно-алгоритмический инструментарий для сегментации мультиспектральных изображений.
4. Набор стандартизованных веб-сервисов на основе предложенных алгоритмов и программного обеспечения с открытым исходным кодом.

**Научная новизна** диссертационной работы заключается в следующем.

1. Предложен вычислительно эффективный непараметрический алгоритм кластеризации MeanSC на основе оценок плотности Розенблатта – Парзена для сегментации мультиспектральных спутниковых изображений. Эффективность достигается за счёт введения сеточной структуры в пространстве признаков и переходу к рабочей выборке значительно меньшего объёма, в которой гарантированно содержатся представители всех классов, присутствующих на изображении. Сеточная структура впервые использована для повышения вычислительной эффективности поэлементного алгоритма кластеризации.
2. Предложен подход к построению ансамбля непараметрических алгоритмов кластеризации, основанных на оценках плотности Розенблатта – Парзена, с помощью согласованной матрицы различий. В рамках этого подхода на основе непараметрического алгоритма MeanSC создан ансамблевый алгоритм кластеризации EMeanSC, позволяющий обеспечить простоту настройки параметров и обработку мультиспектральных спутниковых изображений в диалоговом режиме.
3. На основе предложенных алгоритмов кластеризации разработаны методы разделения формаций лесной растительности с близкими спектрально-яркостными характеристиками и обнаружения усыхающих древостоев по мультиспектральным изображениям. Эти методы позволяют обеспечить качественное выделение мелких и сильно пересекающихся классов, которые не обнаруживаются при использовании традиционных методов автоматизированной обработки.

**Практическая значимость** полученных результатов обусловлена следующим.

Разработанные алгоритмы превосходят описанные в литературе по качеству классификации и/или вычислительной эффективности, что позволяет повысить эффективность сегментации спутниковых изображений в условиях малой априорной информации при решении задач, связанных с исследованием и оценкой состояния территорий по данным дистанционного зондирования Земли. Алгоритмы внедрены в геоинформационную систему с открытым исходным кодом GRASS GIS, а также оформлены в виде стандартизованных веб-сервисов, что позволяет обеспечить доступ к ним по протоколу WPS.

Основные результаты работы были использованы при выполнении междисциплинарных интеграционных проектов СО РАН №№ 3 (2003-2005 гг.), 86 (2006-2008 гг.), 50 (2009-2011 гг.), проектов РФФИ (№№ 09-07-12087-офи\_м, 11-07-12083-офи\_м, 11-07-00202, 11-07-00346, 14-07-31320-мол\_а, 18-37-00492-мол\_а), международного гранта фонда «Научный потенциал» («Human Capital Foundation», 2006 г.) № 66, а также гранта мэрии г. Новосибирска № 09-09 (2009-2010 гг.).

Результаты диссертационного исследования используются в Институте почвоведения и агрохимии СО РАН при крупномасштабном моделировании пространственной организации степной растительности с целью выявления важных закономерностей формирования растительного покрова, что подтверждено актом об использовании.

**Достоверность результатов** обеспечивается корректным применением используемых методов и подтверждается проведенными экспериментальными исследованиями на модельных и прикладных задачах.

**Представление работы.** Результаты работы обсуждались на следующих научных мероприятиях: Международная конференция «Вычислительные и информационные технологии в науке, технике и образовании» (Павлодар, Казахстан, 2006); Всероссийская конференция «Современные методы математического моделирования природных и антропогенных катастроф» (Барнаул, 2007); Всероссийская конференция «Обработка пространственных данных и дистанционный мониторинг природной среды и масштабных антропогенных процессов» (Барнаул, 2013); Всероссийская конференция молодых ученых по математическому моделированию и информационным технологиям (Красноярск, 2006; Томск, 2013); Международная конференция «Automation, Control and Information Technology» (Новосибирск, 2005; Новосибирск, 2010); Всероссийская конференция «Математические методы распознавания образов» (Суздаль, 2009); Международная выставка и научный конгресс «ГЕО-Сибирь» (Новосибирск, 2010; Новосибирск, 2018); Всероссийская конференция с международным участием «Обработка пространственных данных в задачах мониторинга природных и антропогенных процессов» (Бердск, 2019); Международная научная конференция «Региональные проблемы дистанционного зондирования Земли» (Красноярск, 2014; Красноярск, 2018); объединенный семинар «Информационно-вычислительные технологии» в ИВТ СО РАН (Новосибирск, 2005-2015).

**Публикации.** По теме диссертации опубликовано 29 печатных работ, в том числе 7 статей в изданиях из перечня ВАК, 3 – в изданиях, индексируемых в WoS и Scopus, 8 – в других рецензируемых журналах, 4 – в трудах и 7 – в тезисах международных и всероссийских конференций; получено 5 свидетельств о государственной регистрации программы для ЭВМ.

**Личный вклад автора.** Автор принимал активное участие в разработке методов и алгоритмов, а также в интерпретации результатов. Алгоритмы сегментации разработаны автором совместно с Пестуновым И. А. и Бериковым В. Б. Проектирование и программная реализация разработанных алгоритмов, создание пакета программ «Image Processing Toolkit» и проведение численных экспериментов, а также реализация алгоритмов в виде модулей для GRASS GIS и в виде стандартизованных веб-сервисов, выполнены автором лично.

**Структура и объем диссертации.** Работа состоит из введения, пяти глав, заключения, списка цитируемой литературы из 154 наименований и трёх приложений. Каждая глава завершена краткими выводами. Полный объем диссертации составляет 151 страницу, включая 49 рисунков и 7 таблиц.

**Благодарность.** Автор выражает благодарность всем сотрудникам Федерального исследовательского центра информационных и вычислительных технологий за знания, полученные на лекциях, семинарах и практических занятиях, а также за полезные научные дискуссии в ходе выполнения работы.

## СОДЕРЖАНИЕ ДИССЕРТАЦИИ

**Во введении** обоснована актуальность работы, сформулирована цель и поставлены задачи исследования. Показаны научная новизна и практическая значимость полученных результатов, представлены основные положения, выносимые на защиту.

В задачах сегментации спутниковых данных каждому пикселю мультиспектрального изображения можно сопоставить вектор спектральных характеристик (признаков), полученных съёмочной аппаратурой. Эти признаки по своей природе являются случайными, поэтому в данной работе изображение рассматривается, как набор реализаций многомерной случайной величины. В этом случае наиболее подходящим для решения задачи сегментации является вероятностно-статистический подход с использованием методов кластеризации данных.

В **первой главе** приведена содержательная постановка задачи кластеризации данных. Сформулированы характерные особенности задачи сегментации мультиспектральных спутниковых изображений, для учёта которых алгоритмы кластеризации должны обеспечивать: 1) возможность выделения кластеры разной структуры (формы, размера, плотности) в условиях малой априорной информации (в том числе неизвестного числа кластеров); 2) низкую вычислительную сложность, позволяющую обрабатывать спутниковые изображения в диалоговом режиме; 3) возможность обработки данных, содержащих «шум» и выбросы; 4) простоту настройки параметров.

Выполнен анализ известных алгоритмов кластеризации применительно к задаче сегментации мультиспектральных изображений. Выделены наиболее эффективные алгоритмы, характеристики которых приведены в таблице 1. Перечислены наиболее перспективные пути для разработки алгоритмов сегментации мультиспектральных изображений в условиях малой априорной информации.

**Вторая глава** посвящена описанию и исследованию непараметрического алгоритма кластеризации MeanSC, предложенного в ходе диссертационной работы. Вводятся следующие определения.

**Определение 1.** Пусть в точке  $x^* \in R^k$  достигается локальный максимум (мода) функции плотности  $f(x)$ . Точка  $x \in R^k$  считается *связной с  $x^*$* , если существует отрезок кривой  $p_x(t) \in R^k$ ,  $t \in [0,1]$ :

$$p_x(0) = x, \quad p_x(1) = x^*, \quad (t_1 - t_2) \left( f(p_x(t_1)) - f(p_x(t_2)) \right) \geq 0 \quad \forall t_1, t_2 \in [0,1].$$

**Определение 2.** *Кластером*, определяемым модой плотности  $x^*$ , называется непустое множество  $Q(x^*) \subseteq R^k$ , каждая точка которого связна с  $x^*$ . Если  $f(x^*) < \varepsilon$ , то кластер  $Q(x^*)$  относится к «шуму». Здесь  $\varepsilon \geq 0$  – порог «шума».

Сформулирована *формальная постановка задачи кластеризации* в рамках вероятностно-статистического подхода: на основе анализа набора векторов-признаков  $X$ , разбить пространство  $R^k$  на заранее неизвестное число кластеров, определяемых модами плотности, и «шум».

Таблица 1 – Характеристики алгоритмов кластеризации

Название алгоритма	Число параметров	Вычислительная сложность	Обработка большого объёма данных	Результат не зависит от порядка ввода данных	Не требуется задание количества классов	Выделение кластеров сложной структуры	Выполнение в присутствии «шума» и выбросов	Обнаружение иерархической структуры кластеров	Число итераций определено заранее
BIRCH	1	$O(N)$	-	-	-	-	-	+	+
CURE	4	$O(N^2 \log N)$	-	+	-	+	+	+	+
$k$ -средних	1	$O(dMN)^{(1)}$	+	+	-	-	-	-	-
ISODATA	6	$O(dMN)^{(1)}$	+	+	± <sup>(2)</sup>	-	+	-	+
$k$ -представителей	1	$O(MN^2)$	-	+	-	-	-	-	-
DBSCAN	2	$O(N \log N)$	+	+	+	-	+	-	+
OPTICS	2	$O(N \log N)$	+	+	+	+	+	+	+
DeLiClu	нет	$O(N \log N)$	+	+	+	+	+	+	+
SUBCLU	2	$O(N \log N)$	+	+	+	+	+	-	+
BRIDGE	1	$O(N \log N)$	+	+	-	-	+	-	-
GDILC	2	$O(N)$	+	+	+	+	+	-	+
AGRID	2	$O(N)$	+	+	+	+	+	-	+
AGRID+	2	$O(dN)$	+	+	+	+	+	-	+
DBCLASD	нет	1.5-3 × DBSCAN	+	-	+	+	+	-	+
DENCLUE	2	$O(N \log N)$	+	-	+	+	+	-	+
GCOD	1	$O(dNK)^{(3)}$	+	-	+	+	+	-	+
Алгоритм*	2	$O(\tilde{N}^2)^{(4)}$	+	+	+	+	+	-	+
STING	нет	$O(N)$	+	+	+	+	+	-	+
WaveCluster	нет	$O(N)$ при малом количестве кластеров	+	+	+	+	+	-	+
FC	2	$O(N)$	+	-	+	+	+	-	+
ASGC	3	$O(N + K)^{(3)}$	+	+	+	+	+	-	+
Алгоритм**	6	$O(N \log N)^{(1)}$	-	+	+	+	+	-	+
CLIQUE	2	$O(dN + M^d)$	+	+	+	+	+	-	+
ENCLUS	3	$O(dN)$	+	+	+	+	+	-	+
GRIDCLUS	3	$O(N)$	+	+	+	+	-	+	+
GCHL	2	$O(dN \log N)$	+	-	+	+	+	-	+
MAFIA	2	$O(dN + \text{const}^d)$	+	+	+	+	+	-	+

<sup>(1)</sup> Вычислительная сложность каждой итерации.

<sup>(2)</sup> Необходимо задать лишь примерное число кластеров.

<sup>(3)</sup>  $K$  – число ячеек сеточной структуры.

<sup>(4)</sup>  $\tilde{N}$  – объём рабочей выборки

Приведено описание непараметрической оценки плотности распределения Розенблатта – Парзена  $\hat{f}_N(x)$ , которая для произвольной точки  $x \in R^k$  по выборке  $X = \{x^{(1)}, \dots, x^{(N)}\}$  вычисляется следующим образом

$$\hat{f}_N(x) = \frac{1}{Nh^k} \sum_{i=1}^N \Phi\left(\frac{x - x^{(i)}}{h}\right).$$

Здесь  $h$  – параметр сглаживания;  $\Phi(x)$  – ограниченная радиально симметричная функция (ядро), удовлетворяющая следующим условиям сходимости:

$$\sup_{x \in R^k} |\Phi(x)| < \infty, \quad \int_{R^k} |\Phi(x)| dx < \infty, \quad \lim_{\|x\| \rightarrow \infty} \|x\| \Phi(x) = 0, \quad \int_{R^k} \Phi(x) dx = 1.$$

\* Freedman D., Kisilev P. Fast mean shift by compact density representation // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2009. P. 1818-1825.

\*\* Shi Y., Song Y., Zhang A. A shrinking-based clustering approach for multidimensional data // IEEE Transactions on Knowledge and Data Engineering. 2005. Vol. 17. No. 10. P. 1389-1403.

Для уменьшения вычислительной сложности применяются финитные радиально-симметричные ядра. В работе используется оптимальное в среднеквадратичном смысле ядро Епанечникова:

$$\Phi_E(\|x\|) = \begin{cases} \frac{1}{2} V_k^{-1} (k+2) (1 - \|x\|^2), & \text{если } \|x\| < 1, \\ 0 & \text{иначе.} \end{cases}$$

Здесь  $V_k$  – объём единичного  $k$ -мерного шара.

Описана итерационная процедура «среднего сдвига». Она заключается в переходе от  $x \in R^k$  к  $x_1 = x + m_h(x, \Psi)$ , затем от  $x_1$  к  $x_2 = x_1 + m_h(x_1, \Psi)$  и т.д. до точки  $x^*$  для которой  $m_h(x^*, \Psi) = 0$  (здесь и далее верхним индексом  $*$  обозначается мода плотности, к которой сходится процедура «среднего сдвига», стартовавшая из точки пространства). Здесь  $\Psi(x) = -c_\Psi \phi'(\|x\|^2)$ , а вектор «среднего сдвига»  $m_h(x, \Psi)$  вычисляется по формуле

$$m_h(x, \Psi) = \frac{\sum_{i=1}^N x^{(i)} \psi\left(\left\|\frac{x - x^{(i)}}{h}\right\|^2\right)}{\sum_{i=1}^N \psi\left(\left\|\frac{x - x^{(i)}}{h}\right\|^2\right)} - x.$$

Последовательность  $y = \{y_i \in R^k \mid y_0 = x, y_{i+1} = y_i + m_h(y_i, \Psi_E)\}$  называется *траекторией «среднего сдвига»*. В дальнейшем она используется для построения выборочной оценки кривой  $p_x(t)$  из определения 1. Это позволяет использовать выборку  $X$  для оценки локальных мод плотности и выделения кластеров в соответствии с определением 2.

Процедуру «среднего сдвига» предлагается применять не к исходной выборке, а к выборке гораздо меньшего объёма, которая названа рабочей. При этом оценка плотности вычисляется по исходной выборке, а рабочая используется только в качестве набора векторов для запуска процедуры «среднего сдвига». Это позволяет значительно снизить вычислительную сложность без потери точности при вычислении локальных мод плотности.

Для формирования рабочей выборки предлагается следующий метод. В пространстве признаков вводится сеточная структура с ячейками размером  $2h$ , где  $h$  – параметр сглаживания ядра. Затем из каждой ячейки, содержащей достаточно большое число точек исходной выборки, выбирается представитель (центр ячейки). Это позволяет, с одной стороны, сформировать рабочую выборку, гарантированно содержащую представителей из всех значимых классов, а с другой – эффективно выделить «шум» на начальном этапе обработки.

Использование финитного ядра и выбранный размер ячеек позволяют при вычислении  $\hat{f}_N(x)$  исключить перебор элементов выборки, гарантированно не влияющих на плотность, что приводит к значительному уменьшению времени обработки без потери точности. Экспериментально показано, что даже при уменьшении выборки в 1024 раза качество результатов остаётся достаточно высоким при многократном уменьшении времени обработки.

Для распределения точек исходной выборки по сформированным кластерам предлагается использовать не только точки рабочей выборки, но и все точки, в которых вычислялась плотность при применении процедуры «среднего сдвига» (все

точки всех построенных траекторий «среднего сдвига»). Элемент исходной выборки относится к тому же кластеру, что и точка с наибольшей плотностью среди всех точек на расстоянии не более  $2h$ , для которых в ходе работы вычислялось значение плотности. Это позволяет повысить точность выделения границ кластеров.

Далее приведён анализ основных стратегий выбора параметра сглаживания  $h$  и обоснован выбор фиксированного значения, задаваемого пользователем в качестве параметра алгоритма. Перечисляются следующие идеи, позволившие значительно снизить трудоёмкость разработанного алгоритма и повысить качество результата.

- Введение в пространстве признаков сеточной структуры с размером ячейки  $2h$ .
- Формирование представительной рабочей выборки малого объёма.
- Формирование таблицы весов векторов исходного множества (благодаря высокой повторяемости значений, это позволяет значительно сжать исходную выборку без потери информации).

Использование выборочной оценки плотности  $\hat{f}_N(x)$  позволяет модифицировать определения 1 и 2 следующим образом.

**Определение 3.** Пусть в точке  $x^* \in R^k$  достигается локальный максимум оценки плотности  $\hat{f}_N(x)$ . Тогда точка  $x \in R^k$  *связна* с  $x^*$ , если процедура «среднего сдвига», стартовавшая из  $x$ , сходится к  $x^*$ .

**Определение 4.** *Кластером*, определяемым локальным максимумом  $x^*$  ( $\hat{f}_N(x^*) \geq \varepsilon$ ), называется непустое подмножество точек  $Q(x^*) \subseteq X$  связных с  $x^*$ . Если  $\hat{f}_N(x^*) < \varepsilon$ , то компонента  $Q(x^*)$  считается «шумом». Здесь  $\varepsilon \geq 0$  – порог «шума».

Разбиение исходной выборки на кластеры в соответствии с определением 4 приводит к чрезмерно раздробленным результатам. В работе предлагается оригинальный критерий, позволяющий выделять многомодовые кластеры.

**Определение 5.** *Многомодовым кластером* называется непустое подмножество  $C \subseteq X$ , удовлетворяющее условиям:

- 1)  $\forall x \in C$  выполнено  $Q(x^*) \in C$ ;
- 2)  $\forall x_1, x_2 \in C$  существует кривая  $P \subset R^k$ , соединяющая  $x_1^*$  и  $x_2^*$ , вдоль которой

$$\left( 1 - \frac{\hat{f}_N(x)}{\min(\hat{f}_N(x_1^*), \hat{f}_N(x_2^*))} \right) \leq T. \text{ Здесь } T \in [0, 1] \text{ – параметр, задаваемый пользователем}$$

и отвечающий за уровень детализации результата.

Учитывая определения 3–5, разработанный алгоритм MeanSC( $m, \varepsilon, T$ ) можно представить в виде следующей последовательности шагов.

**Шаг 1.** Сформировать клеточную структуру данных в пространстве признаков (разбить пространство признаков  $[0, K_1 - 1]_1 \times \dots \times [0, K_k - 1]_k$  на гиперкубы со стороной  $2h$ ,  $h = \min_{1 \leq i \leq k} K_i/m$ ).

**Шаг 2.** Ввести общую нумерацию клеток (последовательно от одного слоя к другому). С каждой клеткой связать набор попавших в неё векторов из  $X$ .

**Шаг 3.** Сформировать таблицу «весов» векторов множества  $X$  (здесь под «весом» вектора  $x$  понимается число вхождений  $x$  в множество  $X$ ).

**Шаг 4.** Сформировать множество начальных векторов  $S$  для запуска процедуры «среднего сдвига» (если плотность в представителе клетки менее  $\varepsilon$ , такую клетку отнести к «шуму»).

Шаг 5. Применить процедуру «среднего сдвига» к векторам множества  $S$ .

Шаг 6. Распределить исходную выборку по кластерам в соответствии с определением 3.

Шаг 7. Выделить многомодовые кластеры в соответствии с определением 4.

Глава завершается результатами экспериментального исследования разработанного алгоритма методом статистического моделирования на модельных данных и реальных изображениях. Для оценки точности кластеризации вводится следующее определение.

**Определение 6.** Пусть для набора данных  $X = \{x_1, \dots, x_N\}$  известно эталонное разбиение  $g^*: X \rightarrow \{G_0^*, \dots, G_M^*\}$ . Тогда для произвольного разбиения  $g: X \rightarrow \{G_0, \dots, G_K\}$  установим соответствие  $\gamma(G): \{G_0^*, \dots, G_M^*\} \rightarrow \{G_0, \dots, G_K, \emptyset\}$ , при котором выполняется  $\forall (G \neq \bar{G}): \gamma(G) = \gamma(\bar{G}) \Leftrightarrow \gamma(G) = \emptyset$  и достигается наибольшее значение выражения  $n_\gamma = \sum_{i=1}^M [ |C_i^* \cap \gamma(C_i^*)| \cdot I(\gamma(C_i^*) \neq \emptyset) ]$ , где  $I(\cdot)$  – характеристическая функция. Тогда *точность кластеризации (acc)* определяется по формуле  $acc = 100\% * n_\gamma / N$ .

Показано, что алгоритм вычислительно эффективен и способен выделять кластеры сложной формы, размера и плотности. На модельных данных продемонстрирована зависимость количества кластеров и времени обработки от параметра  $m$  (рисунок 1).

В **третьей главе** описывается непараметрический ансамблевый алгоритм EMeanSC, разработанный в ходе выполнения диссертационной работы.

Приведена формальная постановка задачи формирования коллективного решения, которая заключается в следующем.

Пусть с помощью набора группированных функций  $g = \{g^{(1)}, \dots, g^{(L)}\}$  получено  $L$  частных решений  $G = \{G^{(1)}, \dots, G^{(L)}\} \subset \mathbb{G}$ , где  $\mathbb{G}$  – множество всех возможных разбиений  $X$  на произвольное число кластеров. Тогда *итоговое коллективное решение (согласующая функция)* представляет собой отображение  $m: \mathbb{G}^L \rightarrow \mathbb{G}$ .

Выполнен анализ известных стратегий выбора согласующей функции, которые позволяют объединять результаты кластеризации. Показано, что быстроедействие, достаточное для обработки мультиспектральных спутниковых изображений, позволяет обеспечить только подход на основе согласованной матрицы различий, который заключается в следующем. По каждому разбиению  $G \in \mathbb{G}$  строится матрица различий  $H(G) = \{H^{(i,j)}(G)\}$ , где

$$H^{(i,j)}(G) = \begin{cases} 0, & \text{если } x^{(i)} \text{ и } x^{(j)} \text{ отнесены в один кластер в } G, \\ 1, & \text{иначе.} \end{cases}$$

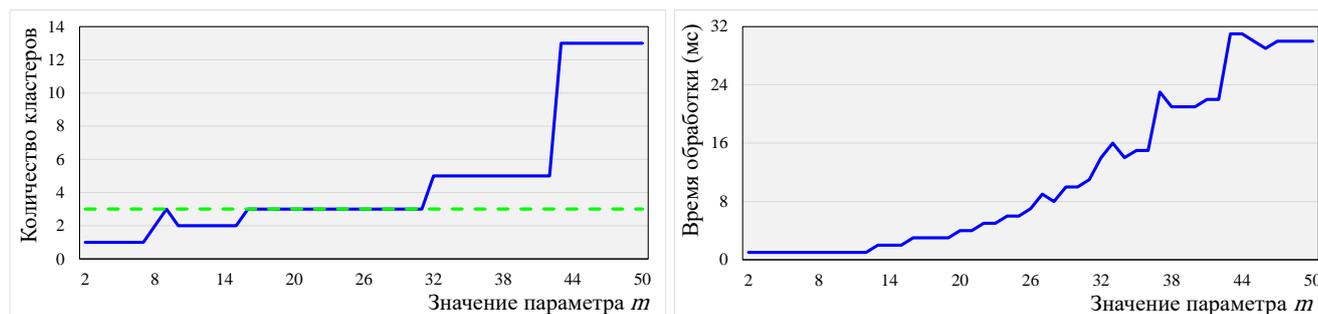


Рисунок 1 – Зависимость количества кластеров (слева) и времени обработки (в мс) (справа) от значения параметра  $m$  для модельного набора данных. Зелёная штриховая линия обозначает число классов, присутствующих в данных

Тогда согласованная матрица различий для  $\mathbf{G}$  вычисляется по формуле

$$H(\mathbf{G}) = \sum_{G \in \mathcal{G}} H(G).$$

Приведено исследование статистических свойств ансамбля, построенного на основе согласованной матрицы различий. Показано, что с ростом мощности ансамбля ожидаемая вероятность ошибочной классификации уменьшается, а дисперсия вероятности ошибки стремится к нулю.

Предлагается непараметрический ансамблевый алгоритм EMeanSC, в соответствии с которым согласованная матрица различий строится по результатам выполнения алгоритма MeanSC с разными значениями параметра  $m$ . Затем для формирования итогового разбиения используется классический агломеративный метод построения дендрограммы. Расстояние между кластерами определяется по принципу «средней связи»

Разработанный алгоритм EMeanSC( $\bar{m} = \{m^{(1)}, \dots, m^{(L)}\}, \varepsilon, T, T_d$ ) можно представить в виде следующей последовательности шагов.

Шаг 1. Для каждого значения параметра сглаживания  $m \in \bar{m}$  построить соответствующее разбиение MeanSC( $m, \varepsilon, T$ ).

Шаг 2. Вычислить согласованную матрицу различий.

Шаг 3. По согласованной матрице различий построить дендрограмму (построение продолжать до тех пор, пока расстояние между ближайшими группами не превысит пороговое значение  $T_d$ ).

Для уменьшения вычислительных затрат предлагается при нахождении итогового решения элементы, метки которых не отличаются ни в одном из разбиений, объединять до начала формирования согласованной матрицы различий.

Глава заканчивается экспериментальным исследованием алгоритма EMeanSC. На модельных данных демонстрируется, что использование ансамблевого подхода позволяет значительно упростить процедуру поиска оптимальных значений параметров алгоритма (пример обработки приведён на рисунке 2). В таблице 2 приводится сравнение результатов обработки данных по ирисам, собранных Э. Андерсоном, алгоритмом EMeanSC с результатами выполнения сеточного алгоритма GCOD (данные для сравнения взяты из статьи<sup>\*</sup>).

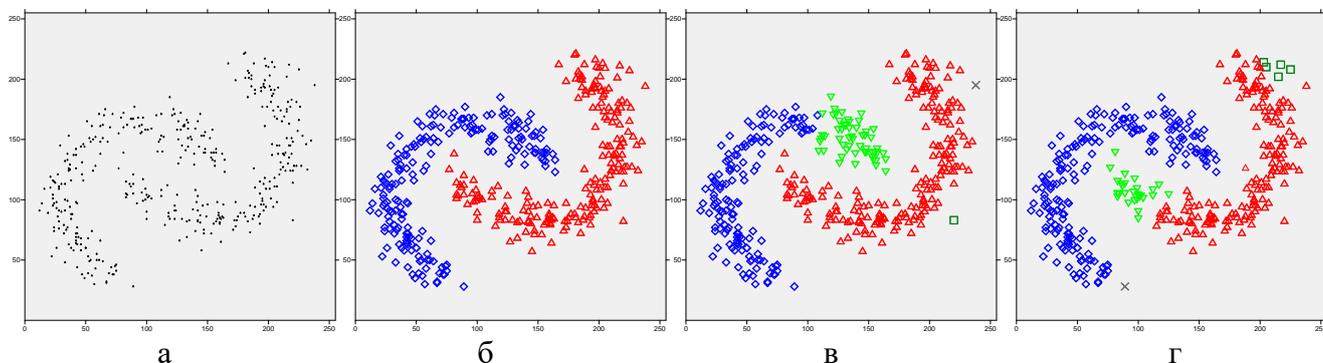


Рисунок 2 – Исходное множество (а), результаты выполнения ансамблевого алгоритма EMeanSC (б) и элементы ансамбля (в и г)

<sup>\*</sup> Qiu B.-Z., Li X.-L., Shen J.-Y. Grid-based clustering algorithm based on intersecting partition and density estimation // Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining. Springer, Berlin, Heidelberg, 2007. P. 368-377.

Таблица 2 – Сравнение алгоритмов EMeanSC и GCOD на данных по ирисам

	$ C_i^0 $	Алгоритм EMeanSC				Алгоритм GCOD			
		$ C_i^S $	$ C_i^0 \cap C_i^S $	Точность, %	Покрытие, %	$ C_i^S $	$ C_i^0 \cap C_i^S $	Точность, %	Покрытие, %
$i = 1$	50	50	50	100	100	50	50	100	100
$i = 2$	50	46	46	100	92	19	18	94.7	36
$i = 3$	50	54	50	92.6	100	81	49	60.5	98

$|C_i^0|$  – число точек, принадлежащих  $i$ -му классу.

$|C_i^S|$  – число точек, отнесенных в соответствующий кластер, выделенный алгоритмом кластеризации.

$|C_i^0 \cap C_i^S|/|C_i^S|$  и  $|C_i^0 \cap C_i^S|/|C_i^0|$  – точность алгоритма и покрытие соответственно.

В четвёртой главе приводится сравнительный анализ предложенных алгоритмов с алгоритмами, включёнными в широко распространённого пакет для обработки спутниковых данных ENVI и из пакетов для анализа данных ELKI (<https://elki-project.github.io>) и Smile (<https://haifengl.github.io/smile>). На семи модельных наборах данных (рисунок 3) показано, что разработанные алгоритмы позволяют получить результаты, не уступающие по точности и/или времени обработки известным непараметрическим алгоритмам (таблица 3). Продемонстрировано, что известные алгоритмы часто не позволяют корректно разбить множество данных на кластеры (см., например, рисунок 4).

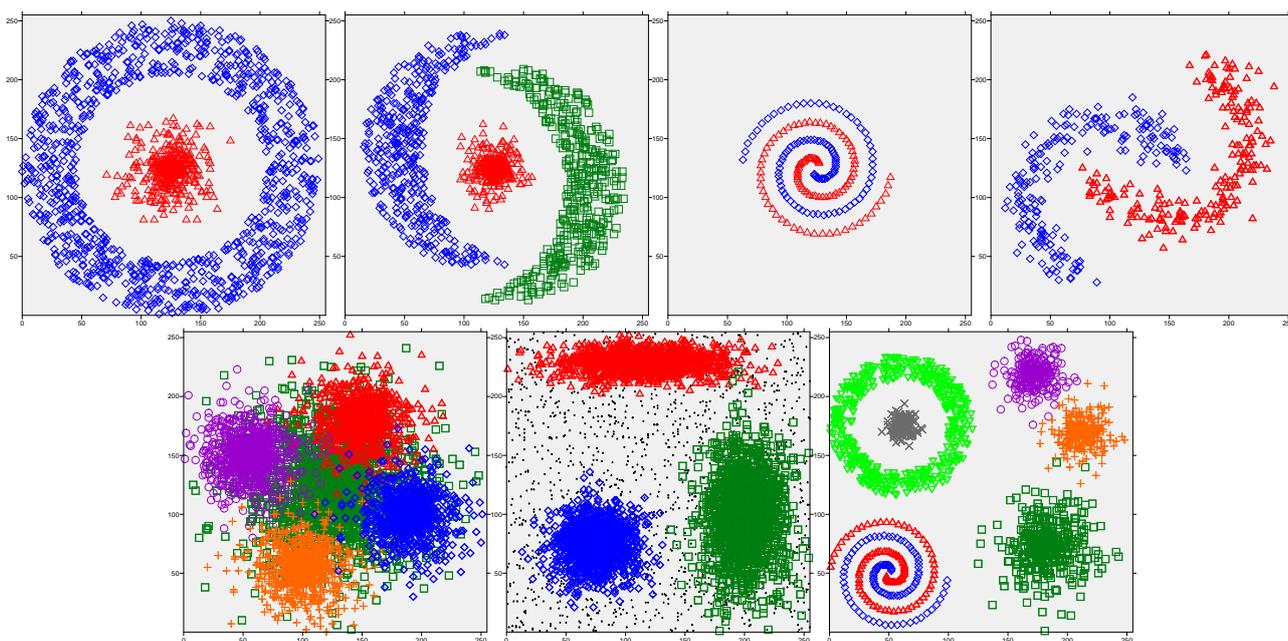


Рисунок 3 – Эталонные разбиения модельных наборов данных 1–7 (чёрные точки – «шум»)

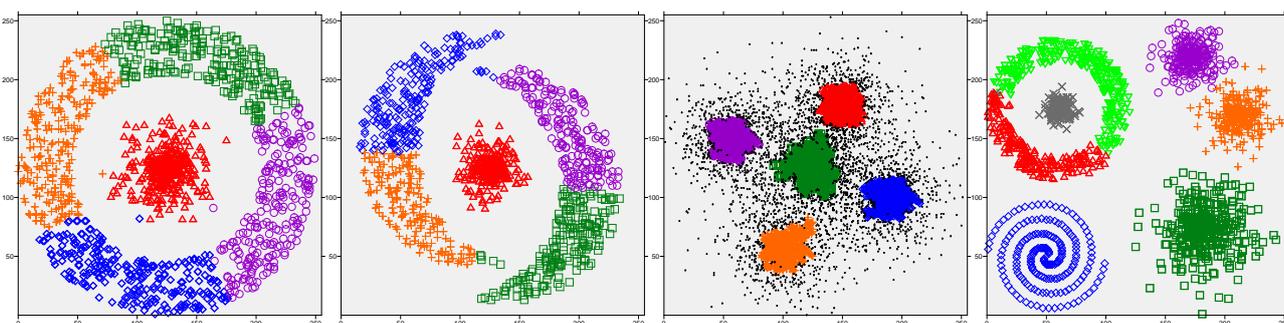


Рисунок 4 – Примеры неправильных разбиений модельных данных (точки – «шум»)

Таблица 3 – Точность кластеризации и время обработки модельных наборов данных

Алгоритм	Модельный набор данных						
	1	2	3	4	5	6	7
MeanSC	100% 0.002 с	100% 0.005 с	100% 0.004 с	99.75% 0.002 с	86.7% 0.01 с	89.18% 0.014 с	98.6% 0.017 с
EMeanSC	100% 0.018 с	100% 0.017 с	100% 0.007 с	99.75% 0.007 с	86.7 0.13 с	89.16% 0.19 с	98.7% 0.17 с
DBSCAN	100% 0.031 с	100% 0.016 с	100% 0.016 с	99% 0.016 с	65.2% 0.374 с	84.56% 1.17 с	90.63% 0.031 с
DENCLUE	65.14% 0.178 с	99.36% 0.09 с	50% 0.06 с	50.5% 0.14 с	85.9% 0.184 с	79.92% 3.26 с	83.8% 0.721 с
<i>k</i> -средних	47.79% 0.016 с	64.73% 0.015 с	53% 0.015 с	47.75% 0.016 с	84.67% 0.016 с	79.92% 0.085 с	78.99% 0.017 с
SLINK	100% 0.031 с	100% 0.016 с	100% 0.016 с	99% 0.016 с	54.08% 0.266 с	78.92% 0.415 с	91.08% 0.031 с

На реальных изображениях (рисунок 5) продемонстрировано, что разработанные алгоритмы обладают достаточной производительностью для обработки изображений размером до 14 млн пикселей в диалоговом режиме (таблица 4).



Рисунок 5 – Тестовые изображения 1-5 (цифровые фотографии) и 6-8 (фрагменты спутниковых снимков WorldView-2)

Таблица 4 – Время обработки тестовых изображений (в секундах), прочерки соответствуют времени обработки более 18 часов

Изображение	1	2	3	4	5	6	7	8
Размер изображения (млн пикселей)	0.3	1	2.2	5	13.8	4.2	9	12
Число каналов	3	3	3	3	3	4	4	4
MeanSC	0.09	0.51	0.86	1.44	8.99	1.44	8.16	4.2
EMeanSC	0.39	2.25	3.16	5.21	31.31	4.97	28.74	10.47
MeanShift	2.91	52	102	67	388	4138	217	62388
<i>k</i> -средних	0.5	36	5	17	1196	75	302	588
ISODATA	1	15	5	9	1178	68	332	337
DBSCAN	194	2731	13098	–	–	39965	–	–
OPTICS	638	5244	40013	–	–	–	–	–
DENCLUE	6934	39849	–	–	–	–	–	–

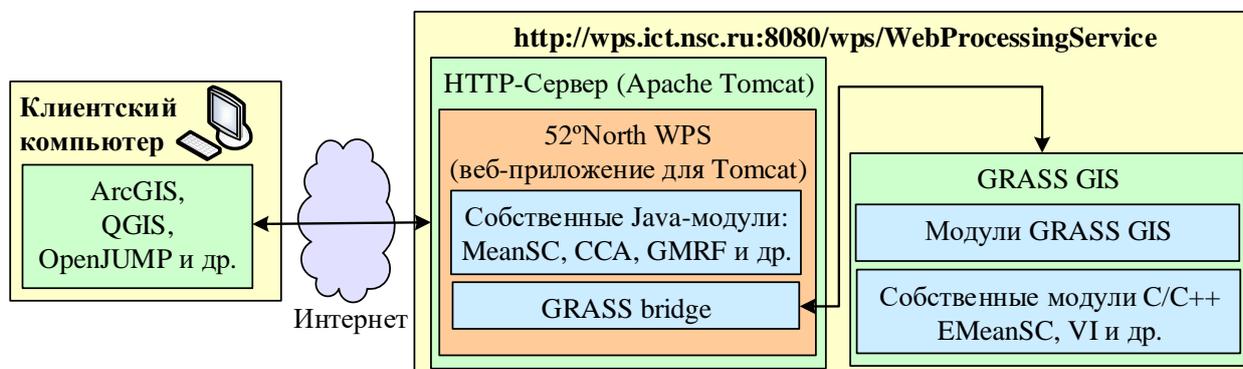


Рисунок 6 – Архитектура разработанной системы сервисов

**Пятая глава** посвящена описанию разработанного программного обеспечения и его использованию для решения практических задач. Представлена технология публикации алгоритмов обработки пространственных данных в виде стандартизованных веб-сервисов по протоколу WPS (Web Processing Service) и их внедрения в распределённую сервис-ориентированную геоинформационную систему Института вычислительных технологий СО РАН. Приведён обзор пакетов программ с открытым исходным кодом, поддерживающих протокол WPS.

Описана разработанная система веб-сервисов (рисунок 6) на основе WPS-сервера 52°North (<http://52north.org>), которая доступна по адресу <http://wps.ict.nsc.ru:8080/wps/WebProcessingService> (протокол доступа – WPS). С её помощью выполнена публикация алгоритма MeanSC.

Далее описан механизм внедрения алгоритмов в виде модулей в геоинформационную систему с открытым исходным кодом GRASS GIS (<http://grass.osgeo.org>). Приводится описание модуля *i.ict.emean*, реализующего алгоритм EMeanSC.

Представлен созданный пакет программ «Image Processing Toolkit», который включает набор эффективных непараметрических алгоритмов (в том числе, MeanSC и EMeanSC). Приведено описание структуры и основных функций пакета, а также включённых в него алгоритмов обработки и сегментации мультиспектральных спутниковых изображений.

Глава заканчивается описанием технологий решения следующих практических задач на основе разработанных алгоритмов.

1. Разделение формаций лесной растительности с близкими спектрально-яркостными характеристиками по данным Landsat-7. Алгоритм MeanSC обеспечивает качественное выделение мелких кластеров, которые зачастую не обнаруживаются при использовании традиционных алгоритмов кластеризации.
2. Обнаружение и классификация усыхающих кедровых древостоев по мультиспектральным изображениям высокого пространственного разрешения по данным Pleiades. Алгоритм EMeanSC позволяет с высокой степенью достоверности выявлять очаги повреждений древостоев (в том числе на ранних стадиях).

В **Заключении** сформулированы следующие результаты работы.

1. Разработан и исследован вычислительно эффективный непараметрический алгоритм кластеризации MeanSC на основе оценок плотности Розенблатта – Парзена для сегментации мультиспектральных спутниковых изображений. Эффективность достигается благодаря введению сеточной структуры в пространстве

признаков и переходу к рабочей выборке значительно меньшего объёма, в которой гарантированно содержатся представители всех классов, присутствующих на изображении. Сеточная структура в пространстве признаков впервые использована для повышения вычислительной эффективности поэлементного алгоритма кластеризации.

2. Предложен подход к построению ансамбля непараметрических алгоритмов кластеризации, основанных на оценках плотности Розенблатта – Парзена, с помощью согласованной матрицы различий. В рамках этого подхода на основе алгоритма MeanSC разработан ансамблевый алгоритм кластеризации EMeanSC, позволяющий обеспечить простоту настройки параметров и обработку мультиспектральных спутниковых изображений в диалоговом режиме.
3. На основе предложенных алгоритмов кластеризации разработаны методы разделения формаций лесной растительности с близкими спектрально-яркостными характеристиками и обнаружения усыхающих древостоев по мультиспектральным изображениям. Эти методы позволяют обеспечить качественное выделение мелких и сильно пересекающихся классов, которые не обнаруживаются при использовании традиционных методов автоматизированной обработки.
4. Выполнен сравнительный анализ предложенных алгоритмов MeanSC и EMeanSC с алгоритмами, включёнными в широко распространённый пакет для обработки спутниковых данных ENVI и в пакеты для анализа данных ELKI и Smile. На модельных данных показано, что алгоритмы MeanSC и EMeanSC превосходят известные непараметрические алгоритмы по точности и/или быстродействию. На реальных изображениях продемонстрировано, что разработанные непараметрические алгоритмы позволяют обрабатывать мультиспектральные изображения в диалоговом режиме.
5. На основе программных продуктов с открытым исходным кодом создана платформа для публикации алгоритмов обработки пространственных данных в виде стандартизованных веб-сервисов (WPS-процессов). В виде WPS-процессов опубликовано пять алгоритмов классификации и кластеризации, в том числе MeanSC и EMeanSC. Платформа доступна по адресу <http://wps.ict.nsc.ru:8080/wps/WebProcessingService> (протокол доступа – WPS). Разработан модуль для геоинформационной системы с открытым исходным кодом GRASS GIS, реализующий алгоритм EMeanSC.
6. Создан пакет программ для обработки и анализа мультиспектральных изображений «Image Processing Toolkit», который включает алгоритмы MeanSC и EMeanSC, а также непараметрический алгоритм сегментации изображений SPMeanSC, учитывающий пространственную информацию, и алгоритм наращивания обучающей выборки, разработанные автором. Пакет «Image Processing Toolkit» передан в Институт почвоведения и агрохимии СО РАН, где используется при крупномасштабном картографическом моделировании структурной организации растительности и почвенного покрова. Результаты его апробации показали эффективность разработанного программно-алгоритмического инструментария при решении практических задач, связанных с анализом мультиспектральных спутниковых изображений.

В качестве **приложений** приведены описание графических пользовательских интерфейсов пакета «Image Processing Toolkit», свидетельства о государственной регистрации пяти программ и акт об использовании результатов исследования в Институте почвоведения и агрохимии СО РАН.

## СПИСОК ОСНОВНЫХ ПУБЛИКАЦИЙ ПО ТЕМЕ ДИССЕРТАЦИИ

### В рецензируемых журналах, рекомендованных ВАК:

1. Пестунов, И. А. Непараметрический алгоритм кластеризации данных дистанционного зондирования на основе grid-подхода / И. А. Пестунов, Ю. Н. Синявский // *Автометрия*. – 2006. – Т. 42. – № 2. – С. 90-99.
2. Пестунов, И. А. Сегментация многоспектральных изображений на основе ансамбля непараметрических алгоритмов кластеризации / И. А. Пестунов, В. Б. Бериков, Ю. Н. Синявский // *Вестник Сибирского государственного аэрокосмического университета*. – 2010. – Вып. 5 (31). – С. 56-64.
3. Пестунов, И. А. Алгоритмы кластеризации в задачах сегментации спутниковых изображений / И. А. Пестунов, Ю. Н. Синявский // *Вестник Кемеровского государственного университета*. – 2012. – Т. 52. – № 4/2. – С. 110-125.
4. Шокин, Ю. И. Распределенная информационная система сбора, хранения и обработки спутниковых данных для мониторинга территорий Сибири и Дальнего Востока / Ю. И. Шокин, И. А. Пестунов, В. В. Смирнов, Ю. Н. Синявский и др. // *Журнал Сибирского федерального университета. Техника и технологии*. – 2008. – Т. 1. – Выпуск 4. – С. 291-314.
5. Синявский, Ю. Н. Методы и технология сегментации мультиспектральных изображений высокого разрешения для исследования природных и антропогенных объектов / Ю. Н. Синявский, И. А. Пестунов, О. А. Дубровская и др. // *Вычислительные технологии*. – 2016. – Т. 21. – № 1. – С. 127-140.
6. Смирнов, В. В. Корпоративные картографические сервисы Сибирского отделения РАН / В. В. Смирнов, И. А. Пестунов, Д. И. Добротворский, Ю. Н. Синявский // *Горный информационно-аналитический бюллетень*. – 2009. – № S18. – С. 130-135.
7. Шокин, Ю. И. Корпоративная информационная система СО РАН для сбора, хранения и обработки спутниковых и наземных данных / Ю. И. Шокин, И. А. Пестунов, В. В. Смирнов, Ю. Н. Синявский и др. // *Горный информационно-аналитический бюллетень*. – 2009. – № S17. – С. 9-15.

### В изданиях, индексируемых в WoS и Scopus:

8. Sinyavskiy, Y. N. Extension of training set using mean shift procedure for aerospace images classification / Y. N. Sinyavskiy, P. V. Melnikov, I. A. Pestunov // *E3S Web of Conferences*. – 2019. – Vol. 75. – No 14. – Article no. 01010. – 2018 Regional Problems of Earth Remote Sensing, RPERs 2018. Krasnoyarsk, Russian Federation. 11 Sept. 2018 – 14 Sept. 2018. **(WoS, Scopus)**
9. Pestunov, I. A. Non-parametric grid-based clustering algorithm for remote sensing data / I. A. Pestunov, Yu. N. Sinyavsky // *Proceedings of the Second IASTED International Multi-Conference on Automation, Control, and Information Technology*. Novosibirsk, Russia, 2005. P. 5-9. **(WoS, Scopus)**

10. Pestunov, I. A. Computationally efficient methods of clustering ensemble construction for satellite image segmentation / I. A. Pestunov, S. A. Rylov, Yu. N. Sinyavskiy, V. B. Berikov // CEUR Workshop Proceedings of the International Conference Information Technology and Nanotechnology. Session Image Processing, Geoinformation Technology and Information Security. – 2017. – P. 194-200. (Scopus)

#### **В других рецензируемых журналах:**

11. Гопп, Н. В. Разделение формаций растительности с близкими спектрально-яркостными характеристиками по данным съемки со спутника Landsat 7 ETM+ / Н. В. Гопп, Е. А. Куликова, И. А. Пестунов, Ю. Н. Сиявский, В. В. Смирнов // Вычислительные технологии. – 2007. – Т. 12. – Спецвыпуск 2. – С. 194-201.
12. Пестунов, И. А. Непараметрический алгоритм кластеризации многоспектральных аэрокосмических данных, основанный на процедуре «среднего сдвига» / И. А. Пестунов, Ю. Н. Сиявский // Вычислительные технологии. – 2004. – Т. 9. – Спецвыпуск. – Труды Совещания российско-казахстанской рабочей группы по вычислительным и информационным технологиям. – С. 125-132.
13. Пестунов, И. А. Обнаружение и картирование повреждений кедровых древостоев по изображениям со спутника Pleiades / И. А. Пестунов, П. В. Мельников, О. А. Дубровская, Ю. Н. Сиявский, В. И. Харук // Интерэкспо Гео-Сибирь. – 2014. – Т. 3. – № 2. – С. 400-408.
14. Сиявский, Ю. Н. Экспериментальное сравнение непараметрических алгоритмов кластеризации для сегментации мультиспектральных изображений / Ю. Н. Сиявский, С. А. Рылов // Интерэкспо Гео-Сибирь. – 2018. – Т. 1. – № 4. – С. 109-114.
15. Добротворский, Д. И. Веб-сервисы для непараметрической классификации спутниковых данных / Д. И. Добротворский, Е. А. Куликова, И. А. Пестунов, Ю. Н. Сиявский // Гео-Сибирь. – 2010. – Т. 1. – № 2. – С. 171-175.
16. Сиявский, Ю. Н. Совместная обработка разнородных данных при сегментации спутниковых изображений высокого разрешения / Ю. Н. Сиявский, И. А. Пестунов, С. А. Рылов, П. В. Мельников // Интерэкспо Гео-Сибирь. – 2015. – Т. 4. – № 2. – С. 57-61.
17. Пестунов, И. А. Технология и программный инструментарий для сегментации спутниковых изображений высокого пространственного разрешения / И. А. Пестунов, С. А. Рылов, П. В. Мельников, Ю. Н. Сиявский // Интерэкспо Гео-Сибирь. – 2013. – Т. 4. – № 1. – С. 202-208.
18. Шокин, Ю. И. Система сбора, хранения и обработки данных дистанционного зондирования для исследования территорий Западной и Восточной Сибири / Ю. И. Шокин, И. А. Пестунов, В. В. Смирнов, Ю. Н. Сиявский, А. П. Скачкова, И. С. Дубров // Гео-Сибирь. – 2009. – Т. 4. – № 1. – С. 165-170.

#### **В трудах международных и всероссийских конференций:**

19. Сиявский, Ю. Н. Использование разнородных данных при сегментации спутниковых изображений высокого разрешения / Ю. Н. Сиявский, И. А. Пестунов, О. А. Дубровская и др. // Материалы Международной конференции «Вычислительные и информационные технологии в науке, технике и образовании (CITech-2015)». – Алмата, Казахстан, 2015. – С. 316-323.

20. Синявский, Ю. Н. Методы и технология сегментации мультиспектральных изображений высокого разрешения для исследования природных и антропогенных объектов / Ю. Н. Синявский, И. А. Пестунов, С. А. Рылов, П. В. Мельников // Сборник трудов Всероссийской конференции «Обработка пространственных данных в задачах мониторинга природных и антропогенных процессов». – 24 – 28 августа 2015 г., с. Усть-Сема, Республика Алтай. – Новосибирск: ИВТ СО РАН, 2015. – С. 107-114. [Электронный ресурс]. – URL: <http://conf.nsc.ru/files/conferences/SDM-2015/294652/SDM-2015%20Thesis.pdf>
21. Пестунов, И. А. Подход к построению ансамбля непараметрических алгоритмов кластеризации для сегментации спутниковых изображений / И. А. Пестунов, С. А. Рылов, Ю. Н. Синявский, В. Б. Бериков // Сборник трудов III Международной конференции и молодежной школы «Информационные технологии и нанотехнологии (ИТНТ-2017)». – Самара: Самарский национальный исследовательский университет имени академика С. П. Королева, 2017. – С. 775-780.
22. Синявский, Ю. Н. Нарращивание обучающей выборки с помощью процедуры «среднего сдвига» в задачах классификации спутниковых изображений / Ю. Н. Синявский, П. В. Мельников, И. А. Пестунов // Материалы Международной научной конференции «Региональные проблемы дистанционного зондирования Земли». – Сибирский федеральный университет, Институт космических и информационных технологий, 2018. – С. 211-215.

#### **Зарегистрированные программы для ЭВМ:**

1. Синявский, Ю. Н. Программа для сегментации изображений «IP\_EMeanSC» / Ю. Н. Синявский, И. А. Пестунов // Свидетельство о государственной регистрации программы для ЭВМ № 2019664225 от 01.11.2019 г. (Федеральная служба по интеллектуальной собственности, патентам и товарным знакам).
2. Синявский, Ю. Н. MeanSC / Ю. Н. Синявский, И. А. Пестунов // Свидетельство о государственной регистрации программы для ЭВМ № 2016617547 от 07.07.2016 г. (Федеральная служба по интеллектуальной собственности, патентам и товарным знакам).
3. Синявский, Ю. Н. Программа для обработки спутниковых изображений «Image Processing Toolkit» / Ю. Н. Синявский // Свидетельство о государственной регистрации программы для ЭВМ № 2019615674 от 06.05.2019 г. (Федеральная служба по интеллектуальной собственности, патентам и товарным знакам).
4. Синявский, Ю. Н. Программа «Image Processing Toolkit GUI» / Ю. Н. Синявский // Свидетельство о государственной регистрации программы для ЭВМ № 2019615205 от 22.04.2019 г. (Федеральная служба по интеллектуальной собственности, патентам и товарным знакам).
5. Синявский, Ю. Н. Программа для наращивания обучающей выборки в задачах классификации спутниковых изображений «IP\_SPMSROIExtension» / Ю. Н. Синявский, И. А. Пестунов // Свидетельство о государственной регистрации программы для ЭВМ № 2019664226 от 01.11.2019 г. (Федеральная служба по интеллектуальной собственности, патентам и товарным знакам).