

На правах рукописи



Замятин Александр Владимирович

**Математическое и программное обеспечение
системы анализа динамики земной поверхности
по данным дистанционного зондирования Земли**

05.13.11 — Математическое и программное обеспечение
вычислительных машин, комплексов и компьютерных
сетей

А в т о р е ф е р а т
диссертации на соискание ученой степени
кандидата технических наук

Томск — 2005

Работа выполнена в Томском политехническом университете

Научный руководитель:

доктор технических наук, профессор **Н.Г. Марков**

Официальные оппоненты:

доктор технических наук, профессор **А.М. Кориков**

доктор технических наук, доцент **А.В. Скворцов**

Ведущая организация:

Институт вычислительной математики и математической геофизики СО РАН, г. Новосибирск.

Защита состоится « 16 » марта 2005 г. в 15 ч. в ауд. 214 на заседании диссертационного совета Д 212.269.06 при Томском политехническом университете по адресу: 634034, г. Томск, ул. Советская, 84, Институт «Кибернетический центр» ТПУ.

С диссертацией можно ознакомиться в библиотеке Томского политехнического университета по адресу: 634034, г. Томск, ул. Белинского, 53.

Автореферат разослан « ___ » февраля 2005 г.

Ученый секретарь
диссертационного Совета
к.т.н., доцент



М.А. Сонькин

Общая характеристика работы

Актуальность работы. Известно, что антропогенная деятельность сопровождается такими негативными процессами как обезлесение, эрозия почвы, изменение береговых линий водоемов и т.д. Все это, как правило, ведет к появлению серьезных локальных и глобальных экологических изменений, которые необходимо анализировать и оценивать, а также прогнозировать их развитие с целью комплексного исследования изучаемой территории земной поверхности и повышения эффективности использования ее природных ресурсов.

Существенная пространственная неоднородность исследуемых природно-территориальных комплексов, их динамичность и большая площадь зачастую не позволяют применять такие традиционные методы слежения за их состоянием как стационарные исследования, тематическое картирование на местности и т.п. Поэтому как альтернатива таким методам за последние десятилетия развивалась концепция экосистемного мониторинга как системы наблюдения, регистрации, контроля динамики, прогнозирования изменений и, наконец, управления и оптимизации экосистем. В рамках этой концепции широкое распространение получили аэрокосмические методы изучения экосистем, называемые *методами дистанционного зондирования Земли (ДЗЗ)*.

За последние десятилетия XX века отечественными учеными (Берлянт А.М., Виноградов Б. В., Тикунов В.С. и другие) был выполнен большой объем исследований, посвященных, в основном, теоретическим и в меньшей степени практическим аспектам решения проблем анализа динамики земной поверхности, в том числе и с использованием данных ДЗЗ. Первые практически интересные результаты при решении этих проблем получены зарубежными учеными (Агравал Ц., Бэйкер В., Белл Э., Вербург П., Хаггет П. и многими другими). Однако в их работах не уделяется должного внимания этапу автоматизированной интерпретации данных ДЗЗ, являющемуся ключевым этапом при анализе динамики земной поверхности.

На сегодняшний день наиболее распространенными среди данных ДЗЗ являются панхроматические (одноканальные) и многозональные (многоканальные) изображения, получаемые с помощью аэрофото- или космической съемки Земли. Одной из главных нерешенных проблем при использовании аэрокосмических изображений (АИ) является проблема разработки новых, более эффективных, чем существующие, методов и программных средств автоматизированной интерпретации таких данных. Аналогичная проблема существует и в области моделирования изменений земной поверхности как основной методологии решения задач анализа динамики земной поверхности. В решении этих проблем как в России, так и за рубежом сделаны только первые шаги и получены первые результаты исследований.

Учитывая все вышеизложенное, можно сделать вывод об актуальности проблемы создания высокоэффективных автоматизированных систем для решения задач анализа динамики земной поверхности, позволяющих проводить интерпретацию АИ, и на основе результатов этой интерпретации выполнять собственно анализ динамики земной поверхности.

Исследования и разработки по теме диссертационной работы проводились в соответствии с утвержденными планами НИР Института «Кибернетический центр» ТПУ в 2000-2005 гг., а также в рамках проекта Минпромнауки РФ по гранту №00-15-98478 поддержки ведущих научных школ России, проектов РФФИ по грантам №00-07-90124, №03-07-90124, № 03-07-06024 и гранту Федерального агентства по образованию №432.

Цель работы и задачи исследования. Целью диссертационной работы является создание математического и программного обеспечения системы анализа динамики земной поверхности по данным ДЗЗ.

Для достижения поставленной цели необходимо последовательное решение следующих задач:

1. Разработка концепции создания системы анализа динамики земной поверхности, включая концептуальные основы автоматизированной интерпретации АИ, как одного из основных этапов анализа динамики. Разработка на основе предложенной концепции структуры системы, которая позволит более эффективно решать задачи интерпретации АИ и моделирования изменений земной поверхности.

2. Разработка математического обеспечения (методов и алгоритмов) системы анализа динамики земной поверхности. Решение данной задачи предполагает также исследование эффективности предлагаемых методов и алгоритмов.

3. Разработка программного обеспечения системы анализа динамики земной поверхности. Результатом решения этой задачи должны явиться программные средства такой системы, созданные с учетом разработанной концепции и предложенной структуры системы и реализующие предложенные методы и алгоритмы.

4. Апробация разработанной системы для решения прикладных задач тематического картирования и анализа динамики земной поверхности с использованием АИ, полученных различными системами ДЗЗ.

Методы исследований. В работе использованы методы многомерного статистического анализа, теории распознавания образов, нейросетевого анализа, пространственной индексации данных, объектно-ориентированного программирования.

Научную новизну полученных в работе результатов определяют:

1. Методы статистической и нейросетевой классификации и реализующие их алгоритмы, предназначенные для интерпретации аэрокосмических изображений и отличающиеся от известных более высокой точностью.

2. Модифицированный алгоритм непараметрической оценки плотности распределения признаков по методу k -го ближайшего соседа, основанный на оригинальном алгоритме индексирования многомерного признакового пространства и отличающийся от подобных алгоритмов оценки плотности более высокой вычислительной эффективностью.

3. Многоэтапный подход к анализу динамики земной поверхности, отличающийся от известных совместным применением более точных методов интерпретации аэрокосмических изображений и алгоритма моделирования изме-

нений земной поверхности, основанного на использовании клеточных автоматов с вероятностным определением правил их функционирования.

4. Результаты исследований предлагаемых методов и алгоритмов, позволяющие оценить пределы их применимости и вычислительную эффективность.

Практическая ценность и реализация результатов работы. Практически значимыми являются созданные методы, алгоритмы и программное обеспечение (ПО) системы анализа динамики земной поверхности, включая ПО для интерпретации АИ, имеющее самостоятельную практическую ценность. Программные средства этой системы функционируют на компьютерах типа IBM PC под управлением ОС Windows 98/NT/2000. Объем исходного кода системы составляет более 9000 строк на языках C++ и Object Pascal.

Созданные программные средства были использованы при выполнении х/д №8-21/03 между Институтом «Кибернетический центр» Томского политехнического университета и ОАО «Востокгазпром» и внедрены в ОАО «Востокгазпром» при решении задач автоматизированного получения тематических карт территории Северо-Васюганского газоконденсатного месторождения Томской области по панхроматическим и многозональным космическим снимкам системы ДЗЗ IRS. Результаты внедрения подтверждены соответствующим актом.

Основные положения, выносимые на защиту:

1. Разработанные методы статистической и нейросетевой классификации и реализующие их алгоритмы позволяют повысить по сравнению с традиционными методами и алгоритмами классификации аэрокосмических изображений точность распознавания объектов земной поверхности.

2. Модифицированный алгоритм непараметрической оценки плотности распределения признаков по методу k -го ближайшего соседа позволяет проводить классификацию аэрокосмических изображений более эффективно, в том числе в вычислительном плане, чем подобные алгоритмы оценки плотности в случае с априори неизвестным распределением признаков.

3. Разработанные алгоритмы, реализующие моделирование изменений земной поверхности, позволяют получать более точные прогнозные карты земной поверхности, чем другие аналогичные алгоритмы.

4. Развитый человеко-машинный интерфейс, предложенная структура и созданное математическое и программное обеспечение системы анализа динамики земной поверхности позволяют более эффективно, чем с использованием подобных систем и комплексов программ, решать основные задачи анализа динамики земной поверхности.

Апробация работы. Основные результаты работы докладывались и обсуждались на следующих конгрессах, симпозиумах, конференциях и семинарах: Международные конференции Association of GeoInformation Laboratories of Europe AGILE V, VI, VII (г. Пальма, Испания, 2002 г.; г. Лион, Франция, 2003 г.; г. Ираклион, Греция, 2004 г.), Русско-американский семинар «Studies of Socio-Natural Co-Evolution from Different Parts of the World» (г. Новосибирск, 2002 г.), IX Международный симпозиум “Remote Sensing 2002” SPIE, (г. Ираклион, Греция, 2002 г.), V Международный семинар «Computer Science and Information Technologies» (г. Уфа, 2003 г.), V и VI Всероссийские научно-

технические конференции «Нейроинформатика — 2003» (г. Москва) и «Нейроинформатика — 2004» (г. Москва), XX Международный конгресс ISPRS (г. Стамбул, Турция, 2004 г.), X Юбилейная международная научно-практическая конференция студентов и молодых ученых «Современные техника и технологии» (г. Томск, 2004 г.), I Международная конференция «Земля из космоса — наиболее эффективные решения» (г. Москва, 2004 г.).

По результатам диссертационных исследований опубликовано 14 работ, в том числе 11 статей.

Личный вклад:

1. Постановка задач исследования и разработка концепции построения системы анализа динамики земной поверхности с использованием аэрокосмических изображений выполнены автором совместно с Н.Г. Марковым.

2. Математическое обеспечение системы анализа динамики земной поверхности разработано автором.

3. Постановки задач исследования эффективности предложенных методов и алгоритмов выполнены автором совместно с Н.Г. Марковым. Результаты исследования этих алгоритмов получены лично автором.

4. Разработка программного обеспечения системы анализа динамики земной поверхности выполнена автором, за исключением подсистемы формирования текстурных признаков, разработанной совместно с Напрюшкиным А.А. Программная реализация алгоритмов параметрической оценки плотности и оценки плотности распределения Розенблатта-Парзена выполнена Напрюшкиным А.А.

Объем и структура работы. Диссертация состоит из введения, четырех глав, заключения, списка использованных источников из 144 наименований и приложения. Объем основного текста диссертации составляет 148 страниц машинописного текста, иллюстрированного 46 рисунками и 11 таблицами.

Содержание работы

Во **введении** обосновывается актуальность темы диссертационной работы, формулируются цель и задачи исследования и приводится краткое содержание работы по главам.

Первая глава посвящена проблемам использования данных ДЗЗ в задачах анализа динамики земной поверхности. Приводится аналитический обзор наиболее известных способов анализа динамики земной поверхности, включающих *методы идентификации изменений*. Эти методы, как правило, позволяют выявить лишь наличие и величину произошедших изменений на исследуемом фрагменте земной поверхности. Более перспективным является подход, основанный на использовании *моделей изменения земной поверхности*. Он позволяет выявлять не только фактически произошедшие изменения, но и экстраполировать происходящие на земной поверхности процессы в будущее, прогнозируя скорость и направление ожидаемых изменений. Показано, что решение задач анализа динамики может быть сведено к решению задач моделирования изменений земной поверхности, что является наиболее перспективным методологическим подходом к решению задач анализа динамики.

Проведены анализ и классификация существующих моделей изменения земной поверхности, включая математическое обеспечение, лежащее в их основе. Показано, что подавляющее большинство этих моделей представляет собой исследовательские версии, не подвергавшиеся тщательному анализу и исследованию эффективности заложенных в них методов и алгоритмов. Поэтому достойными внимания специалистов являются только несколько наиболее законченных моделей изменения земной поверхности. Такие модели, в основном, используют детерминированный подход, основанный на методах регрессионного анализа. Однако, учитывая сложность и большую стохастическую составляющую процессов, происходящих на земной поверхности, все более широким становится использование стохастических *марковских моделей*.

Большинство подходов к моделированию пространственного развития явлений основано на *динамической интеграции*, суть которой заключается в сопоставлении последовательности АИ одной и той же территории, полученных аналогичными системами ДЗЗ в различные моменты времени. При этом необходимо решение целого ряда проблем, где одной из основных является получение серии последовательных разновременных изображений. Другие проблемы связаны с созданием новых высокоэффективных подходов, методов и алгоритмов моделирования изменений земной поверхности и автоматизированной интерпретации АИ, как ключевых этапов при решении задач анализа динамики. В решении этих проблем сделаны только первые шаги и получены исследователями только первые результаты.

На основе проведенного анализа проблем использования АИ для решения задач анализа динамики путем моделирования изменений земной поверхности сформулированы цель и задачи исследований в рамках диссертационной работы.

Во **второй главе** рассматривается концепция создания системы анализа динамики земной поверхности (далее — система). В основу концепции положен многоэтапный подход, базовые этапы которого заключаются в проведении усовершенствованной, более точной тематической интерпретации АИ, моделировании изменений земной поверхности на основе результатов этой интерпретации и в получении при моделировании прогнозных тематических карт на будущие моменты времени.

Этап усовершенствованной, более точной тематической интерпретации АИ нацелен на преодоление недостатков традиционной интерпретации и базируется на ряде сформулированных в работе принципов:

1. Использование наряду с традиционными спектральными признаками текстурных признаков АИ с различным размером анализируемой окрестности.
2. Раздельное использование спектральных и текстурных признаков при классификации, направленное на улучшение детальности интерпретации.
3. Использование для проведения классификации всей имеющейся информации без какого-либо предварительного редуцирования признакового пространства.
4. Совместное использование статистических (параметрических и непараметрических) и нейросетевых подходов к классификации АИ.

Для реализации сформулированных принципов предлагается использовать схему двухэтапной классификации. Эта схема базируется на раздельном использовании спектральных и текстурных признаков АИ, представленных в первичной компоненте (ПК) и вторичной компоненте (ВК) соответственно. При этом ПК используется на втором этапе, а ВК на первом этапе этой двухэтапной классификации. На первом этапе с использованием ВК формируются вероятностные карты для каждого типа (класса) земной поверхности, учитывающие текстурные особенности этих типов. Формирование этих карт проходит с помощью альтернативного использования предложенных *статистического* и *нейросетевого методов* классификации АИ. Второй этап двухэтапной классификации основан на использовании байесовского правила минимизации эмпирического риска, в котором для нахождения значений априорной вероятности принадлежности каждого элемента изображения (пикселя) к тому или иному классу поверхности применяются карты апостериорных вероятностей, полученные на первом этапе.

При реализации предложенных методов классификации формирование ВК может осуществляться двумя способами. В статистическом методе классификации текстурные признаки предлагается учитывать с помощью марковских случайных полей первого порядка и наиболее распространенных текстурных характеристик Харалика — *энтропия*, *энергия* и *вариация*, определяемых по гистограммам частот первого и второго порядка для анализируемого окна размером 5×5 и 9×9 . Такой способ формирования ВК позволяет учитывать пространственные особенности АИ в различных масштабах. Указанные текстурные характеристики рассчитываются для каждого из каналов исходного (панхроматического или многозонального) АИ, поэтому ВК представляет собой признаковое пространство большой размерности, требующее дополнительного преобразования.

Формирование ВК в нейросетевом методе классификации предлагается осуществлять более просто — путем формирования признакового вектора контекстно-спектральным способом. Суть этого способа заключается в том, что каждый компонент вектора представляет собой значения яркостей пикселей в некоторой выбранной окрестности для всех каналов АИ. Такой способ позволяет учитывать текстурные особенности типов земной поверхности без их специального расчета и оптимизации признакового пространства.

Моделирование изменений земной поверхности в рамках многоэтапного подхода предложено выполнять с использованием статистического аппарата *марковских цепей*. В этом случае вероятностная информация об изменениях земной поверхности содержится в *матрице вероятностей переходов* $\mathbf{P} = [p_{ij}]$, каждый элемент которой является вероятностью перехода типа земной поверхности ω_i в тип ω_j , где $i, j = 1, \dots, M$, M — число типов земной поверхности на изучаемой территории, выявленное на этапе интерпретации АИ. Количественная информация об изменениях содержится в *матрице фактических переходов территорий* $\mathbf{M}_{\text{факт}} = [m_{ij}^{\text{ф}}]$ и *матрице ожидаемых переходов территорий* $\mathbf{M}_{\text{ожид}} = [m_{ij}^{\text{ож}}]$. Матрица $\mathbf{M}_{\text{факт}}$ содержит количество элементов изображения (пикселей), перешедших из типа земной поверхности ω_i в тип ω_j , а $\mathbf{M}_{\text{ожид}}$ — ко-

личество пикселей изображения, которые должны поменять свое состояние (значение) в результате процесса моделирования изменений земной поверхности.

Одним из наиболее эффективных средств моделирования поведения сложных систем, таких как земная поверхность, на сегодняшний день являются клеточные автоматы (КА). Каждый КА представляет собой дискретную динамическую систему, состоящую из отдельных элементов (клеток), определенным образом связанных между собой. Правила, определяющие такие связи, называют *правилами функционирования КА*. Для моделирования изменений земной поверхности автором был предложен вероятностный подход к определению правил функционирования КА, основанный на двух принципах.

Первый принцип декларирует использование априорной информации о вероятности развития (угнетения) каждого из типов земной поверхности на исследуемой территории. Такую информацию удобно представить в виде M вероятностных карт, называемых *картами соответствия*, где M — число типов земной поверхности на исследуемой территории. В каждой точке i -й карты исследуемой территории содержится вероятность замещения типа поверхности ω_i на тип ω_j , обозначаемая $p_{ij}^{\text{доп}}$. Формироваться такие карты могут с использованием функций пространственного анализа геоинформационных систем (ГИС), позволяющих строить вероятностные растровые изображения.

Второй принцип декларирует использование при формировании правил функционирования КА *вероятностных и пространственных* характеристик каждого из типов земной поверхности исследуемой территории, рассчитываемых по исходной тематической карте, созданной на этапе интерпретации АИ. *Вероятностные характеристики* вычисляются на основе подхода, суть которого в том, что в некоторой анализируемой окрестности точки (пикселя) тематической карты, вероятность перехода типа поверхности ω_i в тип ω_j зависит не только от вероятности p_{ij} из матрицы \mathbf{P} , но и от количества элементов типа ω_j в этой окрестности. Таким образом, для каждого типа земной поверхности в анализируемой окрестности вычисляется вероятность $p_{ij}^{\text{всп}} = n_j \cdot p_{ij}$, $j = 1, \dots, m$, где m — общее количество типов земной поверхности в анализируемой окрестности, а n_j — количество элементов типа ω_j в этой окрестности.

Использование особенностей *пространственных характеристик* типов земной поверхности основано на применении *критерия насыщенности F* , значение которого определяется как отношение частоты встречаемости конкретного типа земной поверхности в анализируемой окрестности точки h к встречаемости этого типа в целом на изображении (тематической карте)

$$F_{h,k,d} = \frac{n_{k,d,h} / n_{d,h}}{N_k / N}, \quad (1)$$

где $F_{h,k,d}$ характеризует насыщенность анализируемой окрестности с расстоянием от центрального элемента скользящего окна d (например, при $d = 1$ размер окна 3×3) в точке h типа поверхности ω_k ; $n_{k,d,h}$ — число элементов земной поверхности типа ω_k в анализируемой окрестности, $n_{d,h}$ — общее число элементов в анализируемой окрестности, N_k — число элементов типа ω_k на всем изображении.

жении, а N — общее число элементов изображения. Расчет усредненных характеристик критерия насыщенности $\bar{F}_{h,k,d}$ для каждого типа ω_l проводится согласно выражению:

$$\bar{F}_{h,k,d} = \log \left(\frac{1}{N_l} \sum_{h \in L} F_{h,k,d} \right), \quad (2)$$

где L — множество элементов изображения, относящихся к типу ω_l , а N_l — общее число элементов типа ω_l на изображении.

Выражения (1) и (2) позволяют формировать характеристики пространственного распределения типов земной поверхности на изображении и дают информацию о вероятности $p_{ij}^{\text{пр}}$ наличия каждого типа земной поверхности в каждой точке исследуемой области в зависимости от насыщенности каждого из типов земной поверхности в анализируемой окрестности.

Учитывая проблемы, затрудняющие формирование вероятностных карт соответствия, и, следовательно, реализацию первого принципа, результирующая вероятность перехода p_{ij}^{pez} может учитывать только указанные вероятностную и пространственную составляющие и будет иметь вид $p_{ij}^{\text{pez}} = f_1(p_{ij}^{\text{вер}}, p_{ij}^{\text{пр}})$. Если удастся выявить дополнительные факторы, влияющие на развитие процессов на земной поверхности, и, следовательно, построить вероятностные карты соответствия, то результирующая вероятность в этом случае будет иметь вид $p_{ij}^{\text{pez}} = f_2(p_{ij}^{\text{вер}}, p_{ij}^{\text{пр}}, p_{ij}^{\text{доп}})$. Именно такой набор вероятностных составляющих предложено использовать для определения правил функционирования КА.

Выдвигаются требования к ПО системы анализа динамики земной поверхности, а также формулируются принципы ее построения. На основе этих требований и принципов предлагается обобщенная структура системы.

В **третьей главе** рассматриваются и исследуются методы и алгоритмы, положенные в основу системы анализа динамики земной поверхности по данным ДЗЗ, в том числе алгоритмы, реализующие методы статистической и нейросетевой классификации, предложенные в главе 2.

Предложен алгоритм, реализующий метод статистической классификации. На *первом шаге* этого алгоритма проводится формирование исходного множества текстурных признаков согласно описанному выше способу формирования ВК. На *втором шаге* с помощью критерия «хи-квадрат» осуществляется проверка каждого элемента ВК на согласованность с нормальным распределением. Это позволяет на *третьем шаге* преобразовать исходную ВК в ряд подмножеств, согласованных и не согласованных с нормальным законом распределения. В свою очередь, это дает возможность обрабатывать соответствующие подмножества параметрическим или непараметрическим байесовским классификатором и тем самым повышать вычислительную эффективность обработки ВК. Апостериорные вероятности $p_j(\omega_i | X_j^{\text{БТР}})$ ($j = 1, \dots, V'$ — количество подмножеств), полученные с помощью обработки каждого подмножества «своим» классификатором, агрегируются в общую апостериорную вероятность $p(\omega_i | X^{\text{БТР}})$, которая используется в дальнейшем. Такое агрегирование может проводиться путем усреднения всех $p_j(\omega_i | X_j^{\text{БТР}})$ или выбором максимального

$p_j(\omega_i | X_j^{\text{БТР}})$. На *четвертом шаге* полученные значения $p(\omega_i | X^{\text{БТР}})$ трансформируются в вероятностные карты для каждого типа поверхности ω_i ($i = 1, \dots, M$). На заключительных *пятом* и *шестом шагах* алгоритма, представляющих собой реализацию второго этапа двухэтапной классификации, признакововое пространство, представленное в ПК, проверяется на согласованность с нормальным распределением. Затем проводится байесовская классификация с признакововым вектором $X^{\text{БТР}}$, в которой полученная на четвертом шаге каждая i -я вероятностная карта несет информацию об априорной вероятности принадлежности к типу (классу) поверхности ω_i . Отметим, что в байесовском решающем правиле, так же как и на четвертом шаге алгоритма, используется либо традиционная параметрическая оценка условной плотности распределения (УПР) признаков, либо непараметрическая оценка УПР в зависимости от результатов проверки гипотезы о согласованности с нормальным распределением.

В настоящее время наиболее распространенными являются два способа непараметрической оценки УПР. Один способ основан на оценке УПР типа *Розенблатта-Парзена* (УПР РП)

$$\hat{p}(X | \omega_i) = \left(n_i \prod_{v=1}^P c_v^i \right)^{-1} \sum_{s=1}^{n_i} \prod_{v=1}^P \Phi \left(\frac{x_v - x_v^s}{c_v^i} \right), \quad i = 1, \dots, M, \quad (3)$$

где n_i — количество наблюдений в обучающей выборке V_i типа ω_i ; P — количество признаков каналов изображения; c_v^i — сглаживающие параметры обучающей выборки v класса ω_i ; $\Phi(u)$ — ядерная (гауссова) функция оценки плотности распределения.

Другой способ основан на УПР по методу k -го ближайшего соседа (для краткости будем приводить англоязычную аббревиатуру названия метода — k -NN), которая определяется исходя из выражения

$$\hat{p}(X | \omega_i) = \frac{1}{N} \frac{k_n - 1}{V(k_n, N, X)}, \quad i = 1, \dots, M, \quad (4)$$

где k_n — параметр близости соседа, N — величина выборки, $V(k_n, N, X)$ — объем множества всех точек обучающей выборки, расстояние которых до точки X в n -мерном пространстве меньше или равно значению R_k^n . В случае использования евклидова расстояния функция $V(k_n, N, X)$ вычисляется из выражения

$$V(k_n, N, X) = \frac{\pi^{n/2} R_k^n}{|\mathbf{A}|^{1/2} \Gamma[(n+2)/2]}, \quad (5)$$

где Γ — гамма-функция, \mathbf{A} — единичная матрица.

Основным недостатком способов оценки УПР согласно выражениям (3) и (4) является их низкая вычислительная эффективность, связанная с необходимостью анализа всех значений обучающей выборки.

Повышение вычислительной эффективности оценки УПР по методу k -NN может быть получено за счет более быстрого вычисления расстояний R_k^n в выражении (5), что достигается путем использования *методов пространственного индексирования*.

Предложены способ индексирования многомерного пространства, учитывающий несколько вариантов обхода этого пространства с помощью Z -кривой

или кривой Гильберта, а также реализующий его оригинальный алгоритм индексирования с высокой вычислительной эффективностью, достигнутой за счет использования *низкоуровневых битовых операций*. На основе этого алгоритма предложен и разработан модифицированный алгоритм непараметрической оценки УПР по методу k -NN.

Дано описание алгоритма, реализующего нейросетевой метод классификации. От алгоритма, реализующего статистический метод, его отличает более простой подход к формированию вероятностных карт на первом этапе, основанный на описанном выше контекстно-спектральном способе формирования входного вектора. Вероятностные карты для каждого типа земной поверхности формируются с помощью получения «отклика» нейросети с каждого из M выходов, значения которых интерпретируются как апостериорные вероятности соответствующих типов поверхности. В качестве нейросети используется многослойный персептрон с одним скрытым слоем, обученный с помощью традиционного алгоритма обратного распространения ошибки. Второй этап классификации, идентичный для нейросетевого и статистического методов, рассмотрен выше.

Предложен подход и реализующие его алгоритмы, позволяющие автоматизировать поиск нейросетей в базе данных (БД), обученных ранее на распознавание какого-либо класса земной поверхности.

Проведены исследования эффективности методов и алгоритмов, положенных в основу двухэтапной классификации. При этом использовались модельные изображения с различными статистическими характеристиками классов, различным количеством каналов и размером выборок, а также реальные АИ различных систем ДЗЗ. Результаты исследований, в которых в качестве критерия точности использовался *каппа индекс согласия* (КИС), позволили решить следующие задачи:

- определение рациональных параметров модифицированного алгоритма оценки УПР по методу k -NN, включая вид индексирования многомерного пространства и количество обходов признакового пространства;
- поиск пределов применимости различных подходов к непараметрической оценке УПР;
- определение рационального способа преобразования ВК большой размерности;
- определение эффективности статистического и нейросетевого метода классификации в целом.

Некоторые результаты проведенных исследований для случая признакового пространства размерности семь приведены на рис. 1 и рис. 2. Отметим, что столбчатые диаграммы точности (рис. 2а) и вычислительной эффективности (рис. 2б) для столбцов 1-3 получены с использованием байесовской одноэтапной классификации, а столбцы 4 и 5 получены с помощью предложенных методов, основанных на двухэтапной классификации.

Результаты проведенных исследований показали высокую точность предложенных и разработанных методов и алгоритмов, а также более высокую вычислительную эффективность нейросетевого метода классификации в сравне-

нии со статистическим методом. В то же время показано, что более простой способ формирования ВК в нейросетевом методе ведет к снижению точности на 3%-5%. Доверительные интервалы средней ошибки распознавания для статистического и нейросетевого метода (уровень доверия 95%, размер каждой выборки 30 наблюдений) оказались более широкими для нейросетевого метода, что также подтверждает большую надежность и точность результатов, доставляемых статистическим методом. На основании результатов исследований предложен комбинированный алгоритм непараметрической оценки УПР, основанный на модифицированном алгоритме оценки УПР по методу k -NN и известном алгоритме оценки УПР РП. Он обеспечивает более эффективную классификацию в пространстве признаков большой и малой размерности.

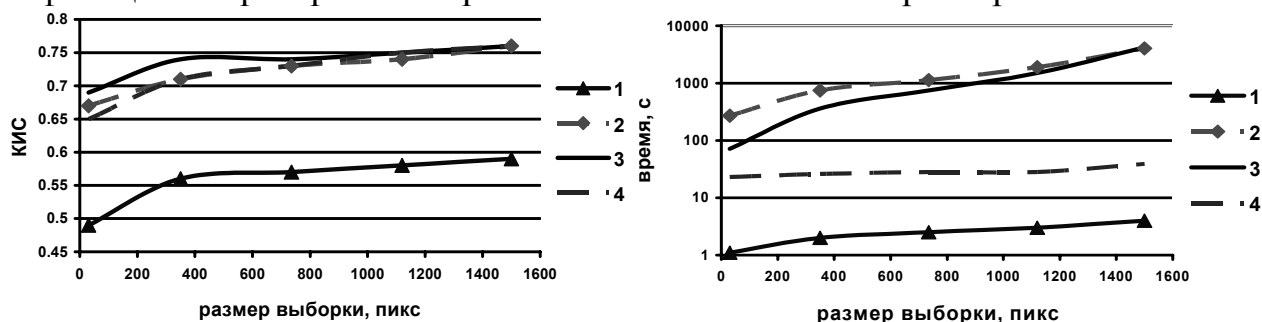


Рис. 1. Эффективность алгоритмов классификации, составляющих основу метода статистической классификации, в зависимости от размера обучающей выборки: 1 — байесовская классификация с параметрической оценкой УПР, 2 — байесовская классификация с непараметрической оценкой УПР по методу k -NN без индексирования, 3 — байесовская классификация с модифицированным алгоритмом оценки УПР РП, 4 — байесовская классификация с модифицированным алгоритмом оценки УПР по методу k -NN

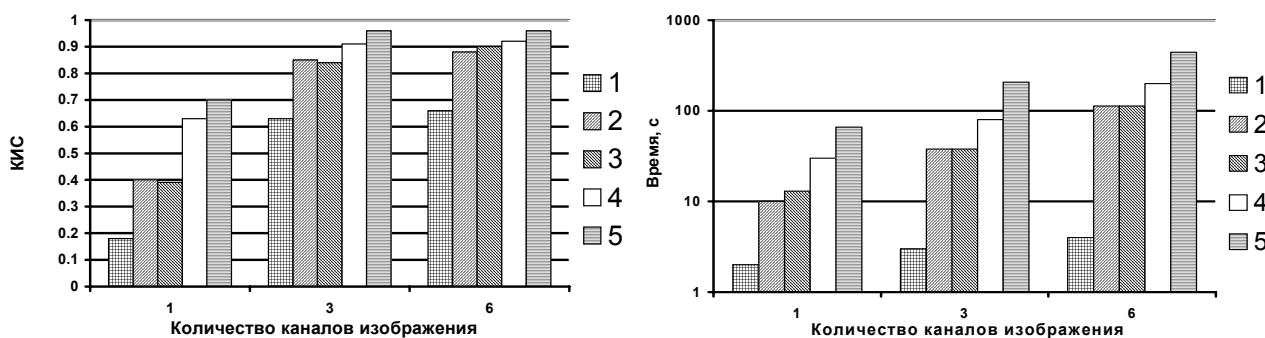


Рис. 2. Эффективность распознавания различных классификаторов в зависимости от числа каналов изображения: 1 — байесовская классификация с использованием параметрической оценки УПР, 2 — байесовская классификация с использованием УПР РП, 3 — байесовская классификация с использованием УПР по методу k -NN, 4 — классификация нейросетевым методом, 5 — классификация статистическим методом

Разработаны алгоритмы, реализующие этап моделирования многоэтапного подхода. Этот этап предполагает по двум разновременным тематическим картам, полученным с помощью интерпретации АИ, предварительный расчет матриц $M_{\text{факт}}$, P и $M_{\text{ожид}}$. Значения этих матриц необходимы для задания вероятностных и иных количественных параметров моделирования.

Моделирование изменений земной поверхности основано на использовании КА с описанным выше вероятностным подходом к формированию правил их функционирования. Каждый КА представлен в виде квадратной матрицы

$\mathbf{M}_{КА} = [c_{ij}]$ порядка d , в которой значение центрального элемента c_{kh} ($k = h = n/2$, здесь «/» — целочисленное деление) зависит определенным образом от значений всех элементов этой матрицы. Матрица $\mathbf{M}_{КА}$ в этом случае формируется путем сканирования растровой матрицы изображения \mathbf{I} скользящим окном размера $(2d+1) \times (2d+1)$ пикселей и сохранения результата обработки $c'_{kh} = f(c_{11}, c_{12}, \dots, c_{kh}, \dots, c_{dd})$ в текущей позиции центрального пикселя скользящего окна в качестве значения пикселя нового (результатирующего) изображения \mathbf{I}' .

Исследование точности прогнозных карт, полученных в результате моделирования, проводилось с использованием модельных изображений, комбинации которых имитировали разновременные АИ. При выявлении значимости предложенных вероятностных характеристик, вводимых согласно второму принципу функционирования КА, вероятность $p_{ij}^{\text{доп}}$ не учитывалась. На рис. 3 в качестве примера представлена часть результатов этих исследований.

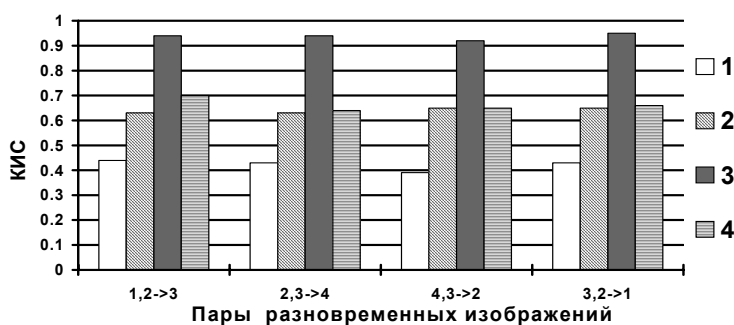


Рис. 3. Точность прогнозных карт, полученных с использованием различных составляющих в правилах функционирования КА:
1 — p_{ij} , 2 — $p_{ij}^{\text{вер}}$, 3 — $p_{ij}^{\text{вер}}$ и $p_{ij}^{\text{пр}}$, 4 — точность карты, полученной с помощью модуля *CA_Markov*

На рис. 3 в качестве примера представлена часть результатов этих исследований. На оси абсцисс рис. 3 показаны различные виды комбинаций четырех модельных изображений, имитирующих разновременные АИ. Так, запись 1,2->3 означает, что расчет параметров и построение прогнозных карт проводилось на основе модельных изображений №1 и №2, а точность полученной прогнозных карт оценивалась по изображению №3. Результаты показали значимость каждой вероятностной характеристики предлагаемых правил функционирования КА и более высокую точность полученных прогнозных карт при учете вероятностных характеристик $p_{ij}^{\text{вер}}$ и $p_{ij}^{\text{пр}}$, чем точность карт, созданных с помощью модуля *CA_Markov*, реализованного в ГИС Idrisi32.

Четвертая глава посвящена разработке ПО системы и ее апробации при решении практических задач анализа динамики земной поверхности.

Разработана структура ПО системы, укрупненная схема которой приведена на рис. 4. В состав ПО системы включено также ПО других производителей:

- программные средства предварительной обработки АИ и библиотека базовых функций работы с изображениями системы ER Mapper 5.5;
- СУБД для работы с БД₁ и БД₂, которые хранят обученные нейросети, АИ, результаты их обработки и интерпретации, а также дополнительную атрибутивную информацию;
- программные средства растровой ГИС Idrisi32 для проведения пространственного анализа.

Оригинальной программной частью созданной системы являются подсистемы, объединенные на рис. 4 в рамках *ПО для решения прикладных задач*. Человеко-машинный интерфейс системы (интерфейс пользователя) разработан в

среде Borland Delphi 6.0, а ПО оригинальных подсистем, учитывая требования к их высокой вычислительной эффективности, — в среде Microsoft Visual C++ 6.0. Взаимодействие между программными модулями, созданными с использованием разных сред и языков программирования, осуществляется посредством вызова соответствующих функций из динамически связываемых библиотек. Серверная компонента, необходимая для хранения различных данных, реализована с помощью языка Transact-SQL на основе СУБД MS SQL Server 2000, а в качестве механизма доступа к БД используется технология Microsoft ActiveX Data Objects.

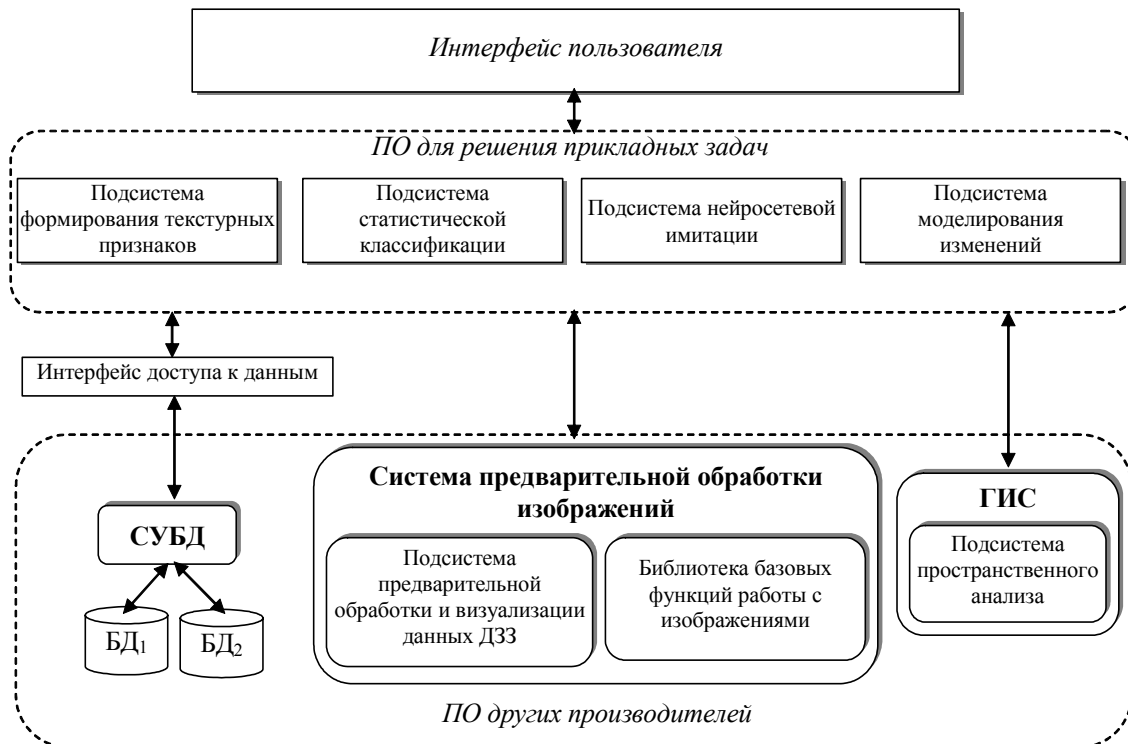


Рис. 4. Укрупненная структурная схема ПО системы

При реализации функций нейросетевой имитации, необходимых для реализации нейросетевого метода классификации, в качестве основы была использована известная библиотека классов с базовыми классами *Neuron*, *Layer* и *Net*, связанными иерархически.

Приведены результаты апробации разработанного ПО системы при решении задачи анализа динамики земной поверхности фрагмента территории Ханты-Мансийского автономного округа. Для ее решения были использованы космические снимки системы ДЗЗ Landsat ETM+ на 10 августа 1999 г., 6 сентября 2000 г. и 3 июля 2002 г. Для каждого из них были получены соответствующие тематические карты (рис. 5а для снимка 2002 г.). На основании карт 1999 и 2000 гг. были получены прогнозные карты на 2001 и 2002 гг. (рис. 5в, карта на 2002 г.), что позволило с использованием программных средств пространственного анализа оценить изменение границ и площадей и выявить динамику изменений тематических классов земной поверхности.

Так, например, на 2000 г. пойменные участки исследуемого фрагмента земной поверхности занимали общую площадь около 257.6 км^2 , а к 2002 г. —

около 205.9 км^2 , то есть общая площадь пойменных участков уменьшилась приблизительно на 51.6 км^2 . В то же время мшистые болота на 2000 г. занимали общую площадь того же фрагмента земной поверхности около 1198.5 км^2 , а к 2002 г. — около 1258.1 км^2 , то есть прослеживается увеличение площади данного типа земной поверхности на 59.6 км^2 . Водные объекты исследуемой земной поверхности, а также границы кустарников практически не претерпели никаких изменений.

Для численной оценки точности прогнозной карты на 2002 г., полученной с помощью ПО системы (рис. 5в), было проведено ее сравнение с реальной тематической картой на 2002 г. (рис. 5а). Значение КИС в этом случае составило 65% совпадения с реальной тематической картой. Для сравнения на основе аналогичных разновременных тематических карт была получена прогнозная карта этой же территории на 2002 г. с помощью программного модуля *CA_Markov* (рис. 5б). Точность этой карты оказалась значительно более низкой и КИС составила 55%. Это подтверждает результаты исследований на модельных разновременных изображениях, приведенные на рис. 3.

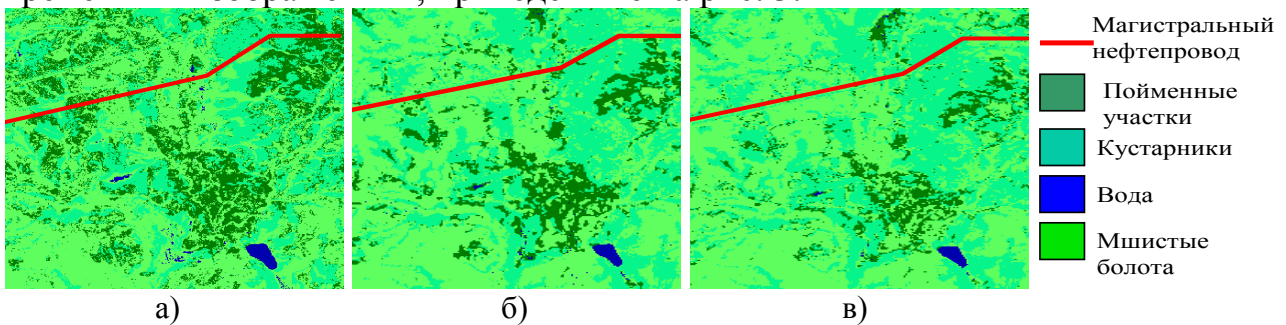


Рис. 5. Фрагменты тематических карт на 2002 г.

а) — оригинальная, б) — прогнозная, полученная с помощью модуля *CA_Markov*, в) — прогнозная, полученная с помощью системы

Описаны результаты апробации ПО системы при решении задачи тематического картирования территории Северо-Васюганского газоконденсатного месторождения ОАО «Востокгазпром» в Томской области по данным космической съемки системой ДЗЗ IRS. Из них следует, что созданное ПО системы эффективно не только при решении задач анализа динамики земной поверхности, но и при решении задачи тематического картирования. Это позволяет также сделать вывод о самостоятельной практической ценности программных модулей, обеспечивающих автоматизированную интерпретацию АИ.

Все приведенные выше результаты апробации системы с использованием данных систем ДЗЗ Landsat ETM+ и IRS подтвердили обоснованность и достоверность результатов выполненных исследований, а также показали высокую эффективность созданного математического и программного обеспечения системы при решении прикладных задач.

В заключении приведены основные выводы и результаты диссертационной работы.

В приложение вынесен акт о внедрении полученных результатов.

Основные результаты и выводы

В ходе выполнения диссертационной работы были получены следующие основные научные и практические результаты.

1. Проведен анализ проблем изучения динамики земной поверхности. Показано, что наиболее перспективным при решении подобных проблем является использование методов ДЗЗ. При этом необходима разработка нового, более эффективного математического и программного обеспечения для ведения автоматизированной интерпретации и моделирования изменений земной поверхности по данным ДЗЗ.

2. Предложена концепция создания системы анализа динамики земной поверхности. В ее основу положен многоэтапный подход, суть которого в последовательном применении более точных методов интерпретации АИ и моделирования изменений земной поверхности, основанного на использовании КА с вероятностным определением правил их функционирования. Сформулирован ряд принципов улучшенной автоматизированной интерпретации АИ и функционирования КА.

3. Разработаны методы статистической и нейросетевой классификации и реализующие их алгоритмы, предназначенные для интерпретации АИ и отличающиеся от известных более высокой точностью. Основу реализации этих методов составляет модифицированный алгоритм непараметрической оценки плотности распределения признаков по методу k -го ближайшего соседа, использующий предложенный алгоритм индексирования многомерного признакового пространства.

4. Разработаны алгоритмы моделирования изменений земной поверхности, реализующие сформулированные принципы функционирования КА.

5. Проведены исследования эффективности разработанных методов и алгоритмов, составляющих основу многоэтапного подхода. По результатам исследований предложен комбинированный алгоритм непараметрической оценки плотности.

6. Разработана структура ПО системы анализа динамики земной поверхности и в соответствии с ней на базе предложенных методов и алгоритмов разработано ПО системы. Объем исходного кода системы составляет более 9000 строк на языках C++ и Object Pascal.

7. Проведена апробация разработанного математического и программного обеспечения системы при решении задачи анализа динамики земной поверхности фрагмента территории Ханты-Мансийского автономного округа по разновременным космическим снимкам системы ДЗЗ Landsat ETM+. Проведена также апробация системы при решении задачи тематического картирования территории Северо-Васюганского газоконденсатного месторождения ОАО «Востокгазпром» (Томская область) по данным системы ДЗЗ IRS. Результаты апробации подтвердили работоспособность и эффективность созданного математического и программного обеспечения системы.

Основные публикации по теме диссертации

1. Замятин А.В. Использование искусственных нейронных сетей при классификации мультиспектральных аэрокосмических изображений // Материалы VI Всероссийской научно-технической конференции «Нейроинформатика — 2004», Ч.2. М.: Изд-во МИФИ. — 2004. — С. 239-246.

2. Замятин А.В. Классификация разновременных снимков в задачах прогнозирования изменений ландшафта земной поверхности // Сборник тезисов I Международной конференции «Земля из космоса — наиболее эффективные решения», М.: ООО «Бином-пресс». — 2004. — С. 116-117.

3. Замятин А.В., Марков Н.Г. Интеллектуальная информационная система для решения задач интерпретации объектов земной поверхности на аэрокосмических изображениях // Труды X Юбилейной Международной научно-практической конференции студентов и молодых ученых «Современные техника и технологии». Томск: Изд-во ТПУ. — 2004. — С. 140-142.

4. Замятин А.В., Марков Н.Г., Напряушкин А.А. Хранение и поиск искусственных нейронных сетей в информационной системе обработки и интерпретации данных дистанционного зондирования Земли // Материалы V Всероссийской научно-технической конференции «Нейроинформатика — 2003», Ч.2. М.: Изд-во МИФИ. — 2003. — С.178-185.

5. Замятин А.В., Марков Н.Г., Напряушкин А.А. Адаптивный алгоритм классификации с использованием текстурного анализа для автоматизированной интерпретации аэрокосмических изображений // Исследование Земли из космоса, 2004.— №2.— С. 32-40.

6. Марков Н.Г., Напряушкин А.А., Замятин А.В. Применение нейронных сетей при решении задач экологического мониторинга // Межвуз. научно-технический сборник «Кибернетика и вуз», вып. 30.— Томск: Изд-во ТПУ. — 2003.— С. 76-81.

7. Markov N.G., Napryushkin A.A., Zamyatin A.V. Adaptive system for extracting information about spatial objects from aerospace imageries // Abstracts of Russian-American Workshop «Studies of Socio-Natural Co-Evolution from Different Parts of the World», Novosibirsk, 2002, <http://iwep.ab.ru/~workshop/> (01.12.2004).

Марков Н.Г., Напряушкин А.А., Замятин А.В. Адаптивная система для извлечения информации о пространственных объектах на аэрокосмических изображениях // Тезисы Российско-Американского семинара «Изучение природно-социального развития в различных частях света».

8. Markov N.G., Napryushkin A.A., Zamyatin A.V. Application of Neural Network Methods in RS-based Thematic Mapping // Proceedings of the 5-th AGILE conference on Geographic science, Palma de Mallorca, Spain, 2002, pp. 485-490.

Марков Н.Г., Напряушкин А.А., Замятин А.В. Применение нейросетевых методов для тематического картирования на основе дистанционного зондирования // Труды 5-ой конференции по геоинформатике Ассоциации ГИС-лабораторий Европы, г. Пальма, Испания.

9. Markov N.G., Napryushkin A.A., Zamyatin A.V., Vertinskaya E. V. Adaptive Procedure of RS Images Classification with Use of Extended Feature Space // Proceedings of SPIE, 2003, Vol. 4885, pp. 489-500.

Марков Н.Г., Напрюшкин А.А., Замятин А.В., Вертинская Е.В. Адаптивная процедура классификации аэрокосмических изображений с использованием расширенного признакового пространства // Труды Международного общества оптического приборостроения.

10. N.G. Markov, A.A. Napryushkin, A.V. Zamyatin, E. V. Vertinskaya, Adaptive Procedure of RS Images Classification with Use of Extended Feature Space // Abstracts of 9th International Symposium on Remote Sensing 2002, Aghia Pelagia, Crete, Greece, 2002, p. 129.

Марков Н.Г., Напрюшкин А.А., Замятин А.В., Вертинская Е.В. Адаптивная процедура классификации аэрокосмических изображений с использованием расширенного признакового пространства // Тезисы 9-го международного симпозиума по дистанционному зондированию, г. Ираклион, Греция.

11. Markov N.G., Zamyatin A.V., Napryushkin A.A., Vertinskaya E.V., Advanced Thematic Mapping Approach for Forecasting Landscape Change Using GIS // Proceedings of the 6th AGILE Conference on Geographic Information Science, Lyon, France, AGILE, 2003, pp. 687-693.

Марков Н.Г., Напрюшкин А.А., Замятин А.В., Вертинская Е.В. Подход к улучшенному тематическому картированию для прогнозирования изменений ландшафта с использованием ГИС // Труды 6-ой конференции по геоинформатике Ассоциации ГИС-лабораторий Европы, г. Лион, Франция.

12. Zamyatin A.V., Markov N.G. Advanced GIS Tool for Assessment of Land Use Change // Proceedings of the 5th International Workshop on Computer Science and Information Technologies, Ufa, Russia, 2003, pp. 115-118.

Замятин А.В., Марков Н.Г. Улучшенный ГИС-инструмент для оценки изменений территорий землепользования // Труды 5-го международного семинара по компьютерным наукам и информационным технологиям, г. Уфа, Россия.

13. Zamyatin A.V., Markov N.G. Designing Forecast Thematic Maps Using Time Series Remotely Sensed Images // XX ISPRS Congress. Proceedings of The International Archives of The Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XXXV, part B, Istanbul, Turkey, 2004, pp. 492-497.

Замятин А.В., Марков Н.Г. Создание прогнозных карт с использованием разновременных аэрокосмических изображений // XX Международный конгресс. Труды международного архива фотограмметрии, дистанционного зондирования и наук о пространственно распределенной информации, г. Стамбул, Турция.

14. Zamyatin Alexandr V., Markov Nikolay G. Multi-stage Processing of Time Series Aerospace Images for Obtaining Enhanced Forecast Land Cover Maps // Proceedings of the 7th AGILE Conference on Geographic Information Science, Heraklion, Greece, AGILE, 2004, pp. 457-463.

Замятин Александр В., Марков Николай Г. Многоэтапная обработка разновременных аэрокосмических изображений для формирования улучшенных прогнозных карт земной поверхности // Труды 7-ой конференции по геоинформатике Ассоциации ГИС-лабораторий Европы, г. Ираклион, Греция.