

РАЗРАБОТКА МЕТОДОВ И АЛГОРИТМОВ ТЕМАТИЧЕСКОЙ ОБРАБОТКИ СПУТНИКОВЫХ СНИМКОВ НА ОСНОВЕ СТРУКТУРНОГО МОДЕЛИРОВАНИЯ

Абрамова Л.В., Алешко Р.А., Батраков Н.М., Гурьев А.Т., Шошина К.В., Щеников В.С.
САФУ им. М.В. Ломоносова (163002, г. Архангельск, наб. Сев. Двины, 17),
r.aleshko@gmail.com

В статье представлены исследования по разработке методов и алгоритмов, предназначенных для автоматизированной обработки данных дистанционного зондирования. Авторами проведено подробное исследование предметной области – лесных ресурсов, рассмотрены основные параметры лесных насаждений. В работе осуществлен выбор наиболее информативных дешифровочных признаков лесных насаждений по обзорным спутниковым снимкам. В статье приведены структурные схемы связей параметров лесных ресурсов и свойств снимков, разработаны производственные правила определения параметров лесов по отражательной способности объектов в различных спектральных диапазонах. Авторами проанализирована достоверность получаемой информации о природных объектах, путем расчета коэффициента «каппа».

Приведены примеры практического применения разработанных алгоритмов. По данным спутниковой съемки построены тематические карты лесных ресурсов с указанием породно-возрастного состава.

Ключевые слова: автоматизация дешифрирования, данные дистанционного зондирования, методы, алгоритмы.

DEVELOPMENT OF METHODS AND ALGORITHMS FOR THEMATIC PROCESSING OF SATELLITE IMAGES BASED ON STRUCTURAL MODELING

Abramova L.V., Aleshko R.A., Batrakov N.M., Gurev A.T., Shoshina K.V., Schenikov V.S.
NArFU Lomonosov University (163002, Arkhangelsk, emb. Northern Dvina, 17),
r.aleshko@gmail.com

The paper presents the research on the development of methods and algorithms designed for the automated processing of remote sensing data. The authors held a detailed study of the subject area - of forest resources, described the main parameters of forest plantations. The work carried out to select the most informative interpretive signs of forest plantations on reviewing satellite images. The paper presents the structural parameters of the circuit connections of forest resources and properties of images, designed production rules determining the parameters of forests on the reflectivity of the objects in different spectral ranges. The authors analyzed the reliability of the information on natural objects, by calculating the coefficient of "kappa".

Examples of the practical application of the developed algorithms. According to satellite imagery built thematic maps of forest resources with an indication of the breed and age composition.

Keywords: automation decoding, remote sensing data, methods, algorithms.

В настоящее время в сфере лесного хозяйства и лесопользования потребность в информации о природных ресурсах уже не удовлетворяется использованием только специализированных цифровых карт и баз атрибутивных данных. Недостаточная актуальность имеющейся информации приводит к замедлению принятия управленческих решений. На сегодняшний день существует потребность в регулярно обновляемых данных о лесных ресурсах. В большинстве случаев для правильной оценки выбранного варианта решения поставленной задачи необходимо иметь достоверную и актуальную информацию о лесах. Такую информацию могут предоставить материалы дистанционного зондирования Земли. Анализ существующих методик и алгоритмов автоматизированного дешифрирования аэрокосмических снимков лесных территорий показал, что, несмотря на значительные результаты исследований в данной области, применение методик автоматизированной тематической обработки спутниковых снимков в лесопромышленном комплексе весьма ограничено. Это, как правило, связано с недостаточно развитым математическим аппаратом в данной области и сложностью формализации процесса дешифрирования лесов.

В связи с вышеизложенным, актуальным является исследование, направленное на теоретическое обобщение подходов к тематической обработке спутниковых снимков лесных территорий и решение научно-технической задачи, связанной с созданием методик и алгоритмов тематической обработки спутниковых снимков лесных территорий с использованием структурного моделирования.

Пусть даны: обзорные и детальные спутниковые снимки, данные наземного обследования территории, набор характеристик таежных лесных насаждений, набор характеристик спутниковых снимков. Спутниковая съемка проводилась в середине летнего периода времени, полученные снимки прошли необходимую радиометрическую, геометрическую коррекцию, была выполнена привязка снимков к географическим координатам.

Требуется найти наиболее информативные дешифровочные признаки лесных территорий; состав таксационных показателей лесных насаждений, определяемых по снимку; взаимосвязи свойств объектов спутникового снимка и свойств объектов лесных насаждений; разработать методику тематической обработки спутниковых снимков таёжных лесов.

Современная практика лесоустройства основывается на количественной и качественной характеристике древостоя, которая включает следующие таксационные показатели (свойства): преобладающая порода, форма, состав, возраст, класс возраста, средняя высота, средний диаметр, класс бонитета, тип леса, полнота, запас на 1га, класс товарности. При общей характеристике насаждения определяют его основные параметры: преобладающая порода, класс возраста, класс бонитета, тип леса.

С применением многоспектральной съемочной системы ТМ (Landsat 5) определим наиболее информативные каналы для дешифрирования класса возраста насаждений. Для этого, на спутниковый снимок наложим векторную карту полевого лесоустройства с указанием классов возраста.

Произведем предварительную обработку данных спутниковой съемки. Пересчет данных из значений яркости пикселей DN (в диапазоне от 0 до 255) в реальные значения приходящего излучения L_λ производится помощью формулы:

$$L_\lambda = \frac{L_{max\lambda} - L_{min\lambda}}{Q_{cal\max}} \cdot Q_{cal} + L_{min\lambda}$$

где $L_{min\lambda}$ — минимальное количество приходящего излучения, $L_{max\lambda}$ — максимальное количество приходящего излучения, $Q_{cal\max}$ — максимальное калиброванное значение DN (255), Q_{cal} — калиброванное значение (DN)

Следующим шагом после пересчета DN в количество приходящего на сенсор излучения является преобразование полученных значений в коэффициент отражения объекта. Такие

данные удобны тем, что они не зависят от времени и сезона съемки, влияния атмосферы и зависят только от самого объекта.

$$\rho_p = \frac{\pi \cdot L_\lambda \cdot d^2}{ESUN_\lambda \cdot \cos\theta_s},$$

где ρ_p – коэффициент отражения объекта; d – расстояние между Землей и Солнцем (а.е.), $ESUN_\lambda$ – внеатмосферная интенсивность солнечной радиации; θ_s – угол зенита Солнца.

Средние коэффициенты отражения в каждом из спектральных каналов сенсора ТМ объектов насаждений с различными классами возраста приведены в таб. 1.

Таблица 1- Средние коэффициенты отражения (класс возраста)

	I	II	III	IV
Канал 1	0,077	0,077	0,075	0,075
Канал 2	0,076	0,069	0,060	0,049
Канал 3	0,035	0,030	0,028	0,021
Канал 4	0,101	0,097	0,091	0,085
Канал 5	0,012	0,011	0,010	0,010
Канал 6	0,020	0,020	0,020	0,020
Канал 7	0,002	0,002	0,001	0,001

Проведем анализ информативности различных спектральных каналов для определения таксационных показателей. Результат этого анализа позволит отбросить некоторые из дешифровочных признаков перед классификацией изображения.

В настоящее время существует большое количество методов определения информативности признаков объектов. Были рассмотрены следующие методы расчета показателей информативности дешифровочных признаков: метод определения дивергенции, метод главных компонент; метод определения расстояния Джеффриса-Мацуситы (JM). Применительно к задаче дешифрирования лесных территорий наиболее подходящим является метод Джеффриса-Мацуситы. Расстояния JM для различных пар признаков рассчитываются с использованием следующих формул:

$$\alpha = \frac{1}{8} (\mu_i - \mu_j)^T \left(\frac{C_i + C_j}{2} \right)^{-1} (\mu_i - \mu_j) + \frac{1}{2} \ln \left(\frac{|(C_i + C_j)/2|}{\sqrt{|C_i| \times |C_j|}} \right),$$

$JM_{ij} = \sqrt{2(1 - e^{-\alpha})}$, где i и j – два тематических класса; C_i – матрица ковариаций; μ_i – вектор средних для соответствующего тематического класса.

Расстояние JM рассчитывалось для всех спектральных каналов и различных сочетаний тематических классов. Например, для класса возраста значения JM приведены в таблице 2.

На основе рассчитанных сумм расстояний JM определим спектральные каналы, использование которых целесообразно для тематического дешифрирования возраста

насаждений. Наиболее пригодными для классификации являются спектральные каналы сенсора ТМ с номерами 2, 3, 1, 4. Аналогичным образом можно оценить информативность спектральных каналов для всех определяемых таксационных показателей. Таким образом, основные таксационные параметры наиболее достоверно дешифрируются по четырем спектральным каналам: ближнему инфракрасному, и трем видимым: красному, зеленому и синему.

Таблица 2 – Расстояния JM для классов возраста

	I-II	I-III	I-IV	II-III	II-IV	III-IV	Сумма
Канал 1	0,241	0,491	0,548	0,571	0,562	0,263	2,676
Канал 2	1,273	1,431	1,414	1,367	1,492	1,402	8,379
Канал 3	1,075	1,292	1,413	0,548	1,367	1,273	6,968
Канал 4	0,928	1,390	1,414	1,190	1,409	1,310	7,641
Канал 5	0,388	0,548	0,609	0,349	0,411	0,239	2,544
Канал 6	0,212	0,259	0,198	0,271	0,241	0,278	1,459
Канал 7	0,241	0,349	0,349	0,501	0,530	0,216	2,186

Для формализованного описания процесса тематического дешифрирования представим структуру лесных насаждений и структуру элементов спутникового снимка в виде множеств с указанием свойств их элементов.

Для формализации *структуры лесных насаждений* введем обозначения объектов леса: Q^{-IV} -лесничество, Q^{-III} -участковое лесничество, Q^{-II} -участок (урочище, дача), Q^{-I} -квартал, Q^0 -выдел, Q^1 -часть выдела, Q^2 -ярус, Q^3 -порода, q^4 -дерево. Набор свойств, характеризующих лесные насаждения, имеет, соответственно, следующий состав: $F(Q^{-IV})$ – свойства, характеризующие лесничество; $F(Q^{-III})$ – свойства, характеризующие участковое лесничество; $F(Q^{-II})$ – свойства, характеризующие лесной участок; $F(Q^{-I})$ – свойства, характеризующие квартал; $F(Q^0)$ – свойства, характеризующие выдел; $F(Q^1)$ – свойства, характеризующие часть выдела; $F(Q^2)$ – свойства, характеризующие ярус; $F(Q^3)$ – свойства, характеризующие породу; $F(q^4)$ – свойства, характеризующие дерево.

Каждая группа в свою очередь задается своим подмножеством свойств. Рассмотрим такое подмножество на примере свойств выдела. Определенные ранее четыре базовых таксационных параметра выдела выступают в качестве подмножества его свойств:

$F(Q^0) = (F^0_{Л1}, F^0_{Л2}, F^0_{Л3}, F^0_{Л4})$, где $F^0_{Л1}$ – преобладающая порода; $F^0_{Л2}$ – класс бонитета; $F^0_{Л3}$ – тип леса; $F^0_{Л4}$ – класс возраста (свойства, связанные с лесными насаждениями имеют подстрочные индексы вида L_n , где n – порядковый номер свойства).

Для формализации *структуры спутникового снимка* введем следующие обозначения элементов множества: P^{-II} - сцена, P^{-I} - фрагмент сцены, P^0 – объект, p^1 - пиксель изображения.

Набор свойств, характеризующих спутниковый снимок, имеет следующий состав: $F(P^{-II})$ – свойства, характеризующие сцену, $F(P^{-I})$ – свойства, характеризующие фрагмент сцены,

$F(P^0)$ – свойства, характеризующие объект, $F(p^1)$ – свойства, характеризующие пиксель изображения. Состав свойств объектов многоспектрального *обзорного снимка* характеризуется совокупностью яркостных характеристик в различных диапазонах съемки. На примере объекта снимка рассмотрим подмножество характеристик свойства $F(P^0)$:

$F(P^0) = (F^0_{C1}, F^0_{C2}, F^0_{C3}, F^0_{C4})$, где F^0_{C1} – яркость в видимом синем спектре; F^0_{C2} – яркость в видимом зеленом спектре; F^0_{C3} – яркость в видимом красном спектре; F^0_{C4} – яркость в ближнем инфракрасном спектре (свойства, связанные с обзорным снимком, имеют подстрочные индексы вида C_n , где n – порядковый номер свойства).

Ранее были определены спектральные каналы, использование которых наиболее целесообразно для определения того или иного таксационного параметра. Руководствуясь этими данными и данными множества свойств выдела и объекта спутникового снимка, построим булеву матрицу $[F(Q^0) \times F(P^0)]$ для обзорных спутниковых снимков (Таб. 3).

Разработанная система связей свойств лесных насаждений и дешифровочных признаков спутниковых снимков позволяет перейти к разработке структурных моделей процесса тематической обработки изображений.

Таблица 3 – Булева матрица $[F(Q^0) \times F(P^0)]$ для обзорных снимков

	$F^0_{Л1}$	$F^0_{Л2}$	$F^0_{Л3}$	$F^0_{Л4}$
F^0_{C1}				•
F^0_{C2}	•	•	•	•
F^0_{C3}	•	•	•	•
F^0_{C4}	•	•	•	

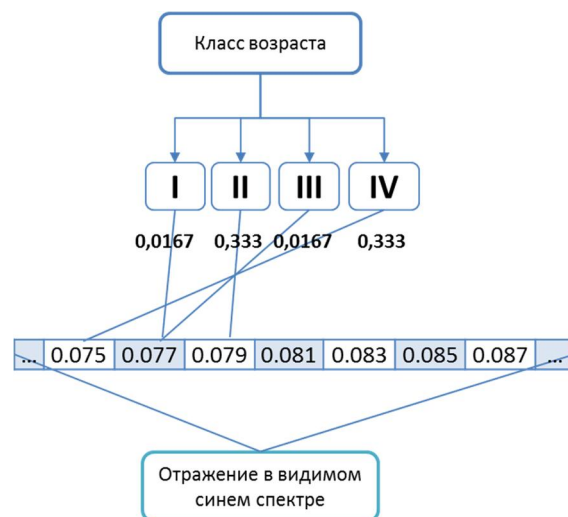


Рис. 1 – Структурные связи свойств «класс возраста» и «коэффициент отражения объекта в видимом синем спектре»

Представим полученные данные в виде структурных схем, демонстрирующих связь основных таксационных показателей и дешифровочных признаков лесных насаждений на примере класса возраста.

Структурные связи свойств «класс возраста» и «коэффициент отражения в видимом синем спектре» представлены на рис. 1.

Аналогичным образом построены структурные схемы для выделенных ранее основных таксационных показателей и всех дешифровочных признаков. Разработанные схемы можно

представить в виде продукционных правил. Введем следующие обозначения используемых спектральных каналов: видимый синий – VB (visible blue), видимый зеленый - VG (visible green), видимый красный – VR (visible red), ближний инфракрасный – NIR (near infra-red). Таксационный показатель «класс возраста» обозначим AC (age class). Критерий определения таксационного параметра укажем в скобках. Для таксационного параметра «Класс возраста» = I правило на основе продукционной модели будут выглядеть следующим образом:

IF VB = 41(0,0167) & VG = 23(0,25) & VR = 14(0,333) & NIR = 47(0,25) THEN AC := I

Для проверки работы алгоритма было выбрано опытное Емцовское участковое лесничество Северного (Арктического) федерального университета. В результате применения разработанной методики были получены карты лесных насаждений обширной территории, сформированной автоматически с указанием пород деревьев и классов возраста (рис. 2).

Проводилась экспериментальная оценка достоверности данных, получаемых в результате применения разработанной методики. Для выполнения классификации может быть использован различный математический аппарат: метод ближайшего соседа, методы кластерного анализа, интеллектуальные и нейросетевые методы. В работе достоверность оценивалась по результатам работы нейросетевого алгоритма классификации. Обучение производилось два раза. Сначала экспертно выделялись эталонные области с указанием породы и возраста насаждений и производилась классификация. Затем в качестве дополнительной информации для алгоритма нейрокомпьютерной сети использовались разработанные правила определения основных таксационных показателей и повторно производилось обучение сети. Далее рассчитывались показатели, характеризующие достоверность классификации.

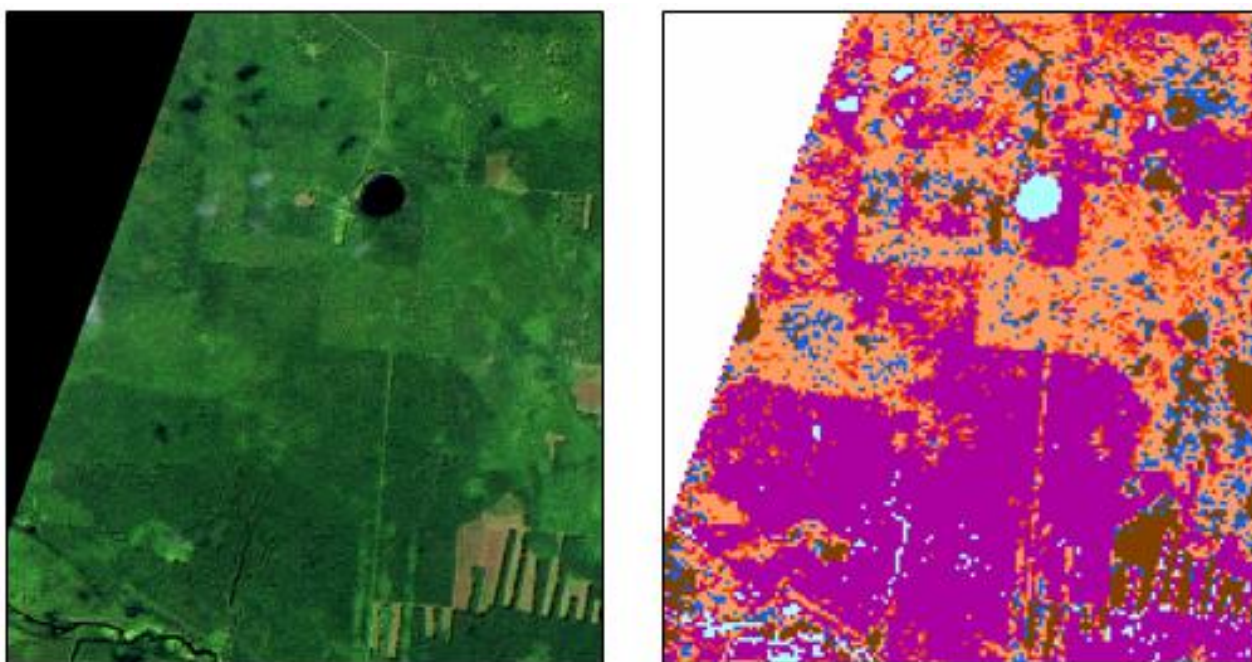


Рис.2– Снимок и карта, сформированная в результате классификации

Показатели наглядно представляются в виде матрицы ошибок, которую рассчитывают по контрольным точкам (не совпадающими с эталонными участками, использовавшимися для создания обучающих выборок). В матрице указываются классы, установленные по данными наземного исследования, а также данные, полученные в результате классификации по снимку. В матрице возможно при этом указать процент достоверности по числу правильно классифицированных пикселей.

Для оценки общей достоверности классификации используют коэффициент «каппа», рассчитываемый по формуле:

$$k = \frac{N \sum_{i=1}^n x_{ii} - \sum_{i=1}^r x_{i+} x_{+i}}{N^2 - \sum_{i=1}^r x_{i+} x_{+i}},$$

где x_{ii} – диагональные элементы матрицы ошибок; x_{i+} - суммарное количество пикселей по строке i ; x_{+i} – суммарное количество пикселей по столбцу i ; N – общее количество пикселей в матрице; n – количество классов.

Качество классификации считается неудовлетворительным при $k < 0,4$, удовлетворительным при $0,4 \leq k < 0,75$, хорошим – при $k \geq 0,75$. По значению k можно напрямую сравнивать результаты работы алгоритмов для одного и того же набора контрольных участков на снимке.

При обучении использовался следующий набор классов: 1 - Ель, средневозрастные; 2 - Ель, спелые и перестойные; 3 - Сосна, средневозрастные; 4 - Сосна, спелые и перестойные; 5 - Береза, молодняк; 6 - Береза, спелые и перестойные; 7 - Водоемы; 8 – Вырубки.

При классификации на основе выделенных эталонных областей получен коэффициент $k=0,6657$, следовательно, результаты можно считать удовлетворительным, но по ряду классификационных групп достоверность едва превышает 50%, что является достаточно низким показателем. Необходимо совершенствование предложенной системы классификации.

Далее проводилось классификация породно-возрастного состава насаждений по тому же спутниковому изображению, но с применением зависимостей, продемонстрированных в ходе реализации методики. Так же, как и в предыдущем случае использовался алгоритм нейрокompьютерной сети, но, помимо данных собранных из эталонных участков, на входы и выходы сети подавались коэффициенты отражения пикселей и варианты их тематической интерпретации. Для данного метода коэффициент $k=0,7977$.

Таким образом, в результате применения оценок, полученных в ходе реализации методики, удалось повысить достоверность данных, получаемых в результате выполнения алгоритма классификации, на 19,83 %.

Таким образом, в исследовании получены следующие научные результаты:

- На основе анализа структурных характеристик спутниковых снимков определен состав дешифровочных признаков отдельных таксационных параметров лесных насаждений обзорных и детальных спутниковых снимков.

- Установлены взаимосвязи основных дешифровочных признаков и таксационных показателей лесных территорий. Разработаны продукционные правила и структурные модели процесса тематической обработки спутниковых снимков лесных территорий, позволяющие автоматизировать процесс дешифрирования таежных лесов.

- На базе проведенных исследований были разработаны методика и алгоритм автоматизированного дешифрирования спутниковых снимков лесных территорий для решения практических задач лесного хозяйства.

Исследование проводится при поддержке грантов Российского фонда фундаментальных исследований, проекты № 14-07-31076, 14-07-98801.

Список литературы

1. Алешко Р.А., Гурьев А.Т. Структурное моделирование взаимосвязей дешифровочных признаков спутниковых снимков и таксационных параметров лесных насаждений — Труды СПИИРАН. Вып. 29 (2013). С. 180–189.

2. Алешко Р.А., Гурьев А.Т. Методика тематического дешифрирования спутниковых снимков лесных территорий на основе структурных моделей //Известия Вузов. Приборостроение. 2013. Т.56. №7. С. 76–77.

3. Алешко Р.А., Бекмешев А.Ю., Васендина И.С., Гурьев А.Т., Карлова Т.В., Шошина К.В., Щеников В.С. Система мониторинга и управления пространственными гетерогенными объектами (на примере Соловецкого архипелага) // Вестник БГТУ. - 2014. - №3(43). - С.104-108.

4. Варфоломеев Ю.А., Гурьев А.Т., Алешко Р.А. Методические и технические аспекты космического мониторинга биоповреждения и усыхания еловых лесов // Лесн. журн.-2010. -№5. -С. 149-156. - (Изв. высш. учеб.заведений).

5. Варфоломеев Ю.А., Гурьев А.Т., Плехов О.Г., Алешко Р.А. Высокотехнологичное проектирование строительства и реконструкции дорог с непрерывным жизненным циклом в лесах с биоповреждениями // Лесн. журн. - 2011. -№2. - С. 145-152. - (Изв. высш. учеб.заведений).

6. Гурьев А.Т., Алешко Р.А., Васендина И.С., Шошина К.В., Щеников В.С. Разработка геоинформационной системы на базе программного обеспечения с открытым исходным кодом // Вестник БГТУ. - 2014. - №3(43). - С.114-118.