



УДК 528.854 (630.4)

DOI: 10.17238/issn0536-1036.2019.5.9

КЛАССИФИКАЦИЯ И ОЦЕНКА СОСТОЯНИЯ СМЕШАННЫХ ДРЕВОСТОЕВ ПО АЭРОИЗОБРАЖЕНИЯМ СВЕРХВЫСОКОГО ПРОСТРАНСТВЕННОГО РАЗРЕШЕНИЯ*

Е.В. Дмитриев^{1,2}, канд. физ.-мат. наук, ст. науч. сотр.; ResearcherID: [E-4794-2014](#), ORCID: [0000-0001-5363-3934](#)

В.А. Козуб², соискатель

П.Г. Мельник^{3,4}, канд. с.-х. наук, доц.; ResearcherID: [E-7644-2014](#)

А.А. Соколов⁵, канд. физ.-мат. наук; ORCID: [0000-0001-9236-5864](#)

А.Н. Сафонова^{6,7}, аспирант; ResearcherID: [F-6811-2019](#), ORCID: [0000-0002-3290-2717](#)

¹Институт вычислительной математики РАН, ул. Губкина, д. 8, Москва, Россия, 119333; e-mail: yegor@mail.ru

²Московский физико-технический институт (национальный исследовательский университет), Институтский пер., д. 9, г. Долгопрудный, Московская область, Россия, 141701; e-mail: postbox-kozub@yandex.ru

³Московский государственный технический университет им. Н.Э. Баумана (Мытищинский филиал), 1-я Институтская ул., д. 1, г. Мытищи, Московская область, Россия, 141005; e-mail: melnik_petr@bk.ru

⁴Институт лесоведения Российской академии наук, ул. Советская, д. 21, с. Успенское, Московская область, Россия, 143030; e-mail: melnik_petr@bk.ru

⁵Laboratoire de Physico-Chimie de l'Atmosphère Université du Littoral Côte d'Opale, Maison de la Recherche en Environnement Industriel 2, 189A, Avenue Maurice Schumann, 59140 Dunkerque, France; e-mail: anton.sokolov@univ-littoral.fr

⁶Сибирский федеральный университет, пр. Свободный, д. 79, г. Красноярск, Россия, 660041; e-mail: safonova.nastya1@gmail.com

⁷Soft Computing and Intelligent Information Systems research group, University of Granada, E-18071 Granada, Spain; e-mail: safonova.nastya1@gmail.com

На сегодняшний день инвазия уссурийским полиграфом признана одним из основных факторов широкомасштабного усыхания сибирских лесов. Появление этого нового организма в пихтарниках серьезно ухудшило их состояние и привело к разнообразным экологическим эффектам в таежных экосистемах. В очагах массового размножения происходит снижение естественного биологического разнообразия, продуктивности лесов, изменение состава и структуры древесного и подчиненных ярусов. В данной работе предлагается методика определения степени поражения древостоев по цветным авиационным изображениям сверхвысокого разрешения (5...10 см на пиксель) с использованием методов машинного обучения. Методика включает в себя этапы предва-

* Статья опубликована в рамках реализации программы развития научных журналов в 2019 г., по материалам XIX Международной конференции молодых ученых «Леса Евразии – Южный Урал» (25–30 августа 2019 г., г. Челябинск).

рительной обработки, сегментации крон отдельных деревьев, обучаемой классификации и оценки степени поражения в соответствии со стандартными категориями. Для отработки методики использовались изображения тестовых территорий заповедника «Столбы» (Красноярский край), полученные с помощью аппаратуры, установленной на беспилотные летательные аппараты DJI Phantom 3 Pro и Yuneec Typhoon H в мае 2016 г. Для этапа построения обучающего ансамбля предложена методика фильтрации обучающих данных, которая позволила повысить точность расчетов на этапе классификации. Приведено обоснование разделения трех основных классов объектов на подклассы с использованием кластерного анализа. Наличие подклассов обусловлено наличием различных пород деревьев на тестовом участке. Проведено сравнение эффективности различных методов обучаемой классификации, используемых для решения данной задачи. Показано, что все рассмотренные методы позволяют достичь предельно высокой точности – почти 95 %. Расчет значений параметра каппа показывает, что классификации, проведенные с помощью всех рассмотренных методов, имеют отличное соответствие экспертным данным. Проведен анализ устойчивости обучения. Оценки полной вероятности ошибки, полученные методами кросс-валидации и переклассификации, отличаются менее чем на 0,1 %, что свидетельствует об отсутствии проблемы переобучения. Анализ соотношения точности и скорости обработки показал, что наиболее целесообразно использовать нормальный байесовский классификатор. Высокая точность классификации позволяет получить оценки 6 степеней поражения древостоев на тестовом участке. Полученные результаты в дальнейшем могут быть использованы для работы региональных служб по управлению лесным хозяйством.

Для цитирования: Дмитриев Е.В., Козуб В.А., Мельник П.Г., Соколов А.А., Сафонова А.Н. Классификация и оценка состояния смешанных древостоев по аэроизображениям сверхвысокого пространственного разрешения // Лесн. журн. 2019. № 5. С. 9–24. (Изв. высш. учеб. заведений). DOI: 10.17238/issn0536-1036.2019.5.9

Финансирование: Работа выполнена при финансовой поддержке РФФИ, проект № 19-01-00215 «Исследование оперативных возможностей гиперспектральных технологий ДЗЗ для решения региональных задач с использованием действующих и перспективных ГСК космического базирования».

Ключевые слова: дистанционное зондирование, изображения сверхвысокого разрешения, лес, оценка состояния лесов, деградация лесов, стволовые вредители, распознавание образов, тематическая обработка изображений.

Введение

В настоящее время происходит широкомасштабное поражение лесов на территории Российской Федерации, и одним из основных факторов этого процесса является уссурийский полиграф (*Polygraphus proximus* Blandford) – инвазивный вредитель-дендрофаг пихты [1, 6]. Данный вид стволовых вредителей изначально был обнаружен на территории Дальнего Востока на Сахалине, однако в последние годы ареал его обитания расширился и охватывает территорию Южной Сибири и европейскую часть России [1, 2]. Так, например, на территории Кемеровской области в период с 2004 по 2010 г. площадь очагов массового размножения уссурийского полиграфа увеличилась в 400 раз [2]. На сегодняшний день уссурийский полиграф наблюдается уже в 47 административных районах, в том числе Ленинградской и Московской областях [7, 8].

В основном массовому поражению уссурийским полиграфом подвержены различные виды пихты: белокорая, сахалинская, цельнолистная и бальзами-

ческая [3, 8, 16]. К повреждаемым деревьям (хотя и в гораздо меньшей степени) также относятся и некоторые другие виды семейства сосновых: ель сибирская, ель европейская, сосна обыкновенная, сосна кедровая корейская, лиственница сибирская [3, 8, 16, 17]. В первую очередь уссурийский полиграф заселяет деревья, поврежденные вследствие естественных и антропогенных факторов. Поражение здоровых деревьев говорит о недостатке пищевой базы при резком повышении численности вредителя.

Пихтовые леса являются не только ценным источником древесины, но и важным сырьем для медицинской промышленности. Появление уссурийского полиграфа значительно ухудшает состояние древостоев, приводит к разнообразным негативным экологическим эффектам, среди которых: снижение биоразнообразия, изменение состава и структуры светового и древесного ярусов, ухудшение микроклимата. Полиграф также участвует в распространении связанных вредителей и переносе фитопатогенных грибов, влияющих на сукцессии растительности. Описанные эффекты создают неблагоприятные условия для естественного возобновления пихтовой части полога и угрозу выпадения из состава насаждений пихты сибирской. С экономической точки зрения ущерб от инвазии жука уссурийского полиграфа значителен [3, 5].

В большинстве случаев возможности своевременного получения информации о новых очагах массового размножения уссурийского полиграфа сильно ограничены. В связи с этим сейчас происходит большой рост расстроенных пихтовых насаждений, непригодных для лесозаготовок и представляющих собой пожарную опасность. Оценка поражения лесов на данный момент осуществляется путем непосредственного выезда соответствующих специалистов на выборочные участки древостоя, которые могут считаться репрезентативными. По таким репрезентативным участкам дается оценка всего древостоя. В связи с тем что количество ресурсов, привлекаемых для таких работ, немного, указанные оценки обладают очень малой точностью. Автоматизация этих работ может сократить затраты на обнаружение очагов вредителя и повысить качество оценок.

В данной работе исследуется возможность получения точных оценок поражения древостоев дистанционными методами с использованием высокодетальных авиационных изображений. Подобного рода задача рассматривалась в работе [18]. В статье описывается двухэтапный метод определения степени поражения древостоев с использованием перспективной методики на основе сверточных нейронных сетей. Основной спорный момент касается постановки задачи. Классы поврежденных деревьев распознаются по их образам при рассмотрении сверху как отдельные объекты. Этапу распознавания предшествует поиск центров крон отдельных деревьев с использованием фильтра Гаусса, который имеет низкую точность: многие кроны либо пропускаются, либо объединяются с соседними в одну крону. В этом случае использование сверточных нейронных сетей, вероятно, потребует неоправданных вычислительных затрат.

Объекты и методы исследования

Методика, предлагаемая в данной работе, основана на альтернативном подходе, который включает в себя сегментацию крон отдельных деревьев и бо-

лее гибкий расчет степени поражения отдельных деревьев на основе результатов автоматизированной попиксельной классификации компонент лесного полога внутри выделенных сегментов. Для проведения численных экспериментов были использованы те же данные, что и в [18]. Для классификации пикселей изображения использовались различные классические методы: нормальный байесовский классификатор, метод k ближайших соседей и многоклассовый метод опорных векторов с гауссовым ядром.

Предлагаемый подход тематической обработки авиационных изображений сверхвысокого разрешения можно разделить на несколько этапов.

Этап 1. Сегментация крон отдельных деревьев на изображении лесного полога. Для решения данной задачи используемые изображения должны иметь достаточно высокое пространственное разрешение – размер пикселя должен быть существенно меньше характерного размера кроны на интересующем нас участке. В ходе реализации данного этапа на изображении выделяются контуры, представляющие собой границы крон отдельных деревьев. Пиксели вне контуров соответствуют межкроновому пространству. Для решения этой задачи может быть использовано как визуальное дешифрирование, так и один из известных алгоритмов автоматизированной сегментации лесного полога, например алгоритм, предложенный в работе [14]. Следует отметить, что при высокой сомкнутости полога указанный алгоритм и другие подобные алгоритмы могут иметь существенные ошибки, и в этом случае для уточнения контуров необходимо использовать визуальный анализ. Усовершенствование алгоритмов сегментации, применяемых на данном этапе, по сути, является отдельной достаточно сложной задачей.

Этап 2. Проверка однородности пикселей внутри контура. Однородность означает, что пиксели одного контура при заданном наборе признаков относятся преимущественно к одному классу. Для проверки выдвинутого предложения используется кластерный анализ, задачей которого является разделение имеющихся неклассифицированных данных на некоторое число классов (оно может указываться исследователем или определяться автоматически), исходя из заданных критериев сходства.

В нашей работе использовалась эффективная модификация метода k -средних [9, 15], состоящая в совместном использовании пакетного и последовательного алгоритмов. Задачей данного метода является разделение набора измерений на заданное число кластеров таким образом, чтобы максимизировать функционал, представляющий собой сумму расстояний от элементов выборки до центров соответствующих им классов

$$\Phi = \sum_{i=1}^k \sum_{x_j \in P_i} (x_j - c_i)^2,$$

где k – количество кластеров, $x_j \in P_i$ – элемент кластера P_j , c_i – центроид кластера P_i .

Качество кластеризации данных зависит от того, насколько удачно было выбрано количество кластеров k . Для визуализации использовалась горизонтальная гистограмма «силуэт» [11]. Эта гистограмма строится для каждого кластера, длина столбиков определяется формулой:

$$s_n(i) = \frac{\min_m \{b_m(i)\} - a(i)}{\max\{a(i), \min(b_m(i))\}}, (m = 1, \dots, k, m \neq n), \quad (1)$$

где $a(i)$ – среднее расстояние от i -го элемента n -го кластера до каждого из остальных элементов этого кластера; $b_m(i)$ – среднее расстояние до элементов m -го кластера ($m: m \neq n$). Из формулы (1) следует, что значение каждого столбца может быть от -1 до 1 , и чем оно выше, тем лучше вписывается объект в свой кластер по сравнению с остальными кластерами.

Этап 3. Обучаемая классификация пикселей и объектов внутри контуров по спектральным признакам. При классификации пикселей количество рассматриваемых классов может быть различным. При этом обязательно должны присутствовать следующие 3 класса: зеленые фитоэлементы деревьев (листва/хвоя), ветви и стволы, прочие объекты. При наличии внутри контура здорового дерева основная часть пикселей будет классифицирована как листва или хвоя, для пораженного дерева будут доминировать пиксели, классифицированные как ветви и стволы. Количество рассматриваемых классов может быть расширено, например, деревья могут быть классифицированы по породному составу и возрастным классам. При построении обучающей базы данных наряду с наземными измерениями необходимо учитывать результаты анализа, осуществляемого на этапе 2.

Для решения этой задачи нами использовались три базовых классификатора: нормальный байесовский классификатор (НБК), метод k -ближайших соседей (КБС) и многоклассовый метод опорных векторов (МОВ) [11, 13], имеющих различную постановку задачи, точность и вычислительную сложность.

При реализации НБК предполагается, что признаки распознаваемых объектов имеют нормальное распределение. Обучение данного классификатора состоит в получении оценок $\hat{\mu}_y$, векторов математических ожиданий и $\hat{\Sigma}_y$, ковариационных матриц для каждого распознаваемого класса $y \in Y$. Пусть x – классифицируемый вектор признаков, $class(x)$ – алгоритм классификации (функция, отображающая пространство признаков на конечное множество наименований объектов). Тогда алгоритм НБК имеет вид:

$$class(x) = \arg \max_{y \in Y} \left(\ln \det(\hat{\Sigma}_y^{-1}) - (x_y - \hat{\mu}_y)^T \hat{\Sigma}_y^{-1} (x - \hat{\mu}_y) \right).$$

Метод КБС определяет класс распознаваемого объекта на основании информации о ближайших к нему объектах, которые находятся в обучающей выборке. Пусть $u_i(x)$, ($i = 1, \dots, k$) – наименования k объектов, наиболее близких к классифицируемому вектору признаков x . Тогда алгоритм КБС имеет вид:

$$class(x) = \arg \max_{y \in Y} \sum_{i=1}^k [u_i(x) = y],$$

где квадратные скобки означают индикаторную функцию. В нашей работе для определения близости используется евклидова метрика. Для увеличения скорости поиска используется модификация КБС на основе kd -деревьев [12].

В классическом варианте метода опорных векторов решается задача бинарной классификации. При обучении данного метода решается задача максимизации расстояния между дискриминантными гиперплоскостями, которые проходят через несколько граничных векторов признаков распознаваемых классов из обучающего множества. Эти точки называются опорными векторами. В случае неразделимых классов дискриминантные гиперплоскости проходят через опорные точки с заданной допустимой погрешностью.

Решающее правило для бинарного МОВ имеет вид:

$$\text{class}(x) = \text{sign} \left(\sum_{x_i \in S} \lambda_i y_i(x_i, x) - \omega_0 \right),$$

где S – множество опорных векторов, λ_i – отличные от нуля параметры (коэффициенты Лагранжа), значения которых определяются на этапе обучения.

Для решения задач нелинейной классификации вместо скалярных произведений (x_i, x) используется специальное преобразование $K(x_i, x)$, называемое ядром. В нашей работе используется гауссово ядро:

$$K(x_i, x) = \exp \left(-\|x_i - x\|^2 / (2\sigma^2) \right).$$

Для перехода к случаю многоклассовой классификации используется метод декодирования, известный в зарубежной литературе как ECOC (Error Correcting Output Codes) [10]. Метод основан на минимизации средневзвешенных потерь при решении серии из L задач бинарной классификации, определяемых некоторой заданной матрицей кодирования C . Алгоритм классификации имеет вид:

$$\text{class}(x) = \arg \min_{y_i \in Y} \frac{\sum_{j=1}^L |C_{ij}| g_{y_i}(C_{ij}, S_j(x))}{\sum_{j=1}^L |C_{ij}|},$$

где g_{y_i} – функция потерь для бинарного МОВ, $S_j(x)$ – направленное расстояние от классифицируемого вектора x до дискриминантной поверхности, C_{ij} – элемент кодовой матрицы [10]. В нашем случае $g_{y_i}(s) = \max(0, 1 - y_i s)$, а в качестве кодовой матрицы используется кодирование по принципу «каждый против каждого» [10].

Этап 4. Модель оценок степени поражения. На данном этапе для каждого контура из тестового участка производится оценка поражения на основе результатов классификации пикселей этого контура, полученных на этапе 3. Введем обозначения. Пусть:

- X_1 – количество пикселей, отнесенных к классу зеленых фитоэлементов;
- X_2 – количество пикселей, отнесенных к классу «ветви и стволы»;
- X_3 – количество пикселей, отнесенных к классу «прочие объекты».

Если $X_3 \geq X_1 + X_2$, то полагаем, что оконтуренный объект не является деревом, иначе полагаем, что оконтуренный участок является деревом со степенью поражения:

$$t = \frac{X_2}{X_2 + X_1} \cdot 100 \%$$

При решении различных практических задач традиционно оперируют дискретными характеристиками состояния древостоев, определяемых по заданному набору признаков. Так, например, для лесов Московской области в работе [4] были введены 4 категории жизненного состояния: живые (без признаков ослабления), ослабленные, усыхающие и сухостой. Позже (для сибирских лесов) были предложены 6 категорий [5] по степени повреждения уссурийским полиграфом: 1 – здоровые, 2 – ослабленные, 3 – сильно ослабленные, 4 – усыхающие, 5 – свежий сухостой, 6 – старый сухостой.

Для пересчета степени поражения в категорию поражения нами использовалась линейная шкала. Такой подход был выбран в силу ограниченности используемых данных. Для нахождения более точного вида функции шкалирования необходимо провести дополнительные наземные исследования.

Результаты исследования и их обсуждение

Для отработки методики использовались изображения территории заповедника «Столбы», который находится вблизи г. Красноярска. Для получения изображений использовалась бесшовная ортомозаика цветных изображений с пространственным разрешением 5...10 см на пиксель, полученных в результате съемки с управляемого беспилотного летательного аппарата (БПЛА) DJI Phantom 3 Pro со встроенной камерой в июле 2016 г. и с гексакоптера Yuneec Turphoon H (с камерой CGO3+) в мае 2016 г. Положение тестовых участков на территории заповедника представлено на рис. 1.

Большая часть территории заповедника представляет собой средний горный пояс (500...800 м над уровнем моря), покрытый смешанным лесом с преобладанием пихты. Первое появление уссурийского полиграфа в природном заповеднике «Столбы» было зарегистрировано в 2011 г., и, по полученным в настоящее время оценкам, ущерб, нанесенный жуком, составляет около 25...30 % от площади, занимаемой пихтой. Породный состав тестового участка представлен семью основными таксонами в разных пропорциях: сосна (40 %), пихта (25 %), лиственница, кедр, ель, береза и осина.

Для детальной демонстрации результатов из имеющихся данных был выбран относительно небольшой тестовый участок, на котором достаточно хорошо представлены деревья различных пород и степеней поражений. Для обследуемой территории мы имели только общие наземные данные о породном составе и степени поражения. При проведении визуального анализа снимков наблюдалось разделение здоровых и пораженных деревьев минимум на 2 класса. Таким образом, было выдвинуто предположение о возможности выделения следующих 5 классов:

- здоровые деревья типа 1 (class A);
- здоровые деревья типа 2 (class B);
- пораженные деревья типа 1 (class C);
- пораженные деревья типа 2 (class D);
- прочие объекты (class E).

Контуры, используемые для формирования обучающей выборки, представлены на рис. 2. Общие данные о количестве контуров и пикселей каждого класса, которые попали в выборку, представлены в табл. 1.

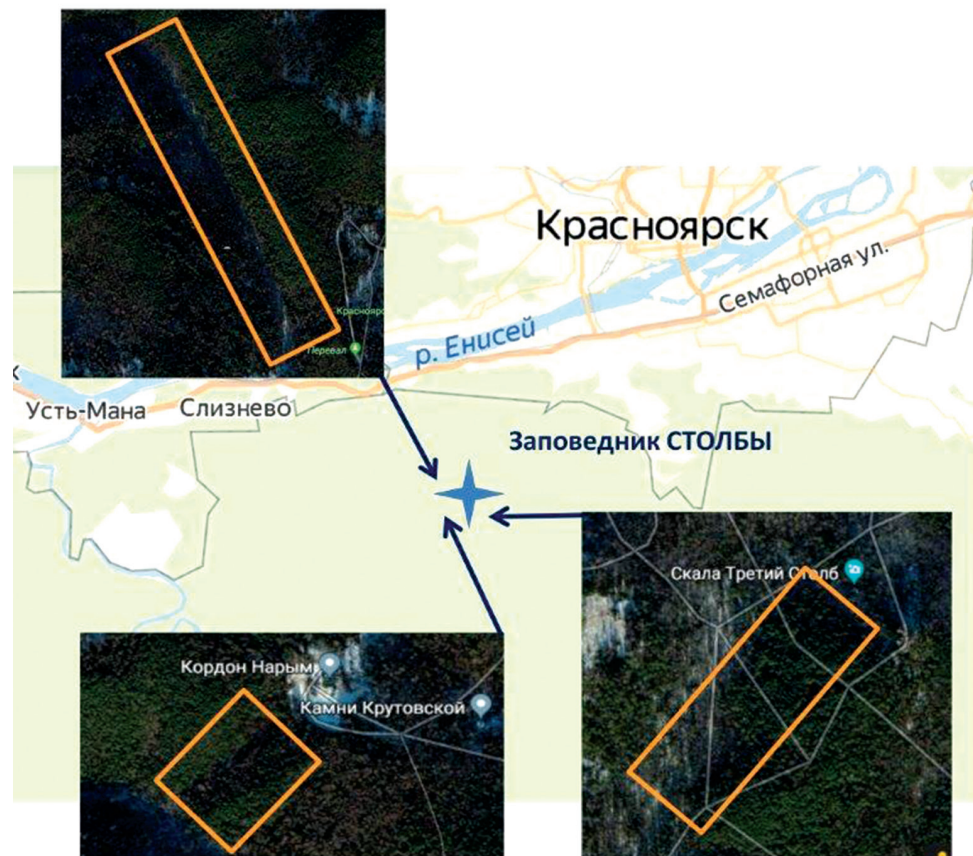


Рис. 1. Тестовые участки на территории заповедника «Столбы» (Красноярский край)
Fig. 1. Sample plots on the territory of the “Stolby” Nature Reserve (Krasnoyarskiy Region)

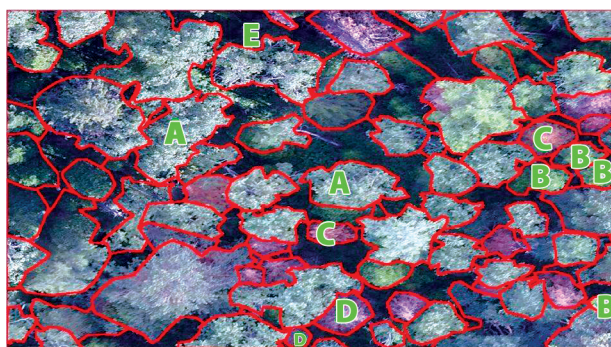


Рис. 2. Сегментация крон на тестовом изображении. Буквами отмечены контуры, пиксели которых использовались для построения обучающей выборки в соответствии с рассматриваемыми классами

Fig. 2. Segmentation of the tree crowns in the test image. The letters denote the contours which were used as sources of pixels for forming of the learning sample in accordance with the examined classes

Таблица 1

Параметры обучающей выборки

Класс	Кол-во контуров для обучающей выборки	Общее кол-во пикселей в контурах
A	2	70 259
B	4	26 347
C	2	15 244
D	2	11 520
E	1	43 596

При съемке разные компоненты полога располагались под разным углом к камере и имели разную освещенность. Из-за этого яркости разных частей одного объекта могут сильно различаться, однако отношение между яркостями разных каналов должны сохраниться. Поэтому целесообразно преобразовать яркости пикселей обучающей выборки таким образом, чтобы все они имели одинаковую интегральную яркость, а именно: для каждого пикселя поделить яркость каждого канала на интегральную яркость. Интегральная яркость рассчитывалась с помощью метода трапеций, центральные длины волн *RGB*-каналов были заданы следующим образом: красный – 660 нм, зеленый – 560 нм, синий – 485 нм.

Следует заметить, что интегральная яркость является нормой на данном подпространстве векторов яркостей пикселей пространства \mathbb{R}^3 . Поэтому, про-нормировав яркости всех пикселей, входящих в выборку, мы тем самым снизили размерность пространства, и, следовательно, всю информацию о яркости каждого пикселя можно получить, взяв две любые нормированные яркости из трех. В дальнейшем будут использоваться нормализованные яркости *R*- и *G*-каналов.

Для проверки предположения о разделимости классов «здоровые деревья» и «пораженные деревья» на подклассы – мы провели численные эксперименты с использованием метода *k*-средних. Схема расчетов состояла в следующем. Обучающие данные для классов A и B объединялись в единое множество, и далее к нему применялся метод *k*-средних при $k = 2$. В случае удачной кластеризации делается вывод, что по указанным признакам класс «здоровые деревья» разделяется на 2 подкласса. Аналогичным образом проводится тест для классов C и D.

Результаты этого исследования представлены на рис. 3 и 4. На рис. 3а представлены пиксели контуров, по которым формировались обучающие данные для классов A и B. В верхней части представлена часть *RGB*-изображения, соответствующая данным пикселям. В нижней части – результат кластерного анализа: красный – класс A, синий – класс B.

Как мы можем видеть, результаты кластерного анализа полностью согласуются с первоначальным предположением о разбиении на два подкласса – более 90 % пикселей было классифицировано верно. Кроме того, следует отметить, что основные артефакты расположены по краям контуров, и, скорее всего, это свидетельствует о наличии незначительных ошибок при сегментации крон деревьев. Аналогичные выводы мы можем сделать и для классов C и D (рис. 4а).

Качество кластеризации подтверждается также и силуэтами кластеров, представленными на рис. 3б и 4б. Как можно видеть, отрицательные значения отсутствуют, силуэты имеют форму, близкую к прямоугольной, и соотношение количества столбиков для каждой пары классов согласуется с данными табл. 1. Эксперименты с большим числом подклассов показали значительно худший результат, таким образом, рассмотрение указанных выше 5 классов мы сочли обоснованным.

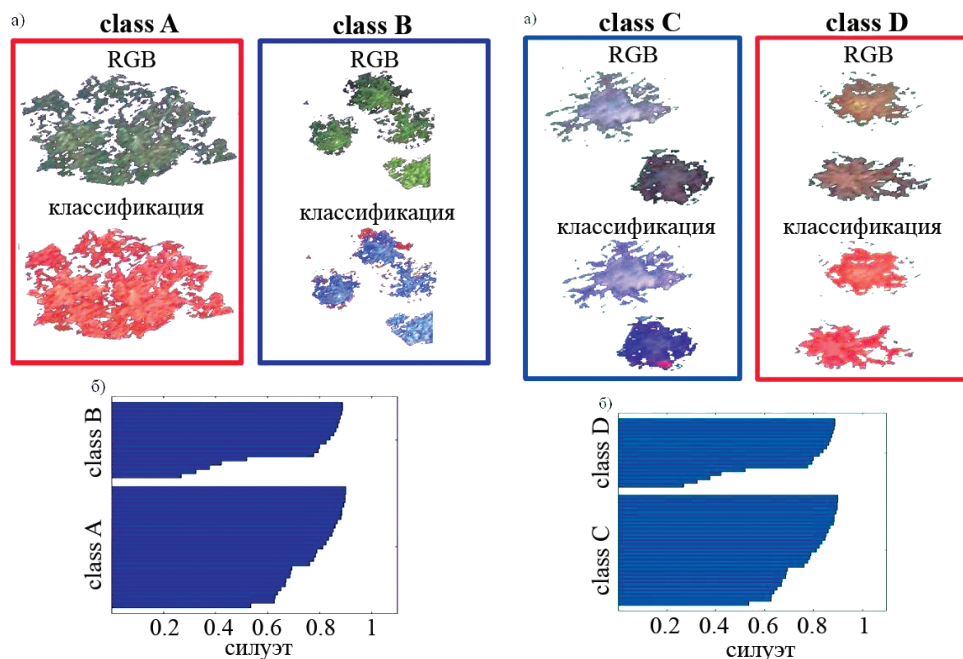


Рис. 3. Обоснование разделимости класса здоровых деревьев на подклассы – тип 1 и тип 2: а) результаты кластерного анализа методом k -средних при разделении на 2 класса; б) силуэты кластеров

Fig. 3. Substantiation of discriminability of the vigorous trees class to the subclasses type 1 and type 2: а) results of the cluster k -means analysis by fractionation to 2 classes; б) shadow figures of the clusters

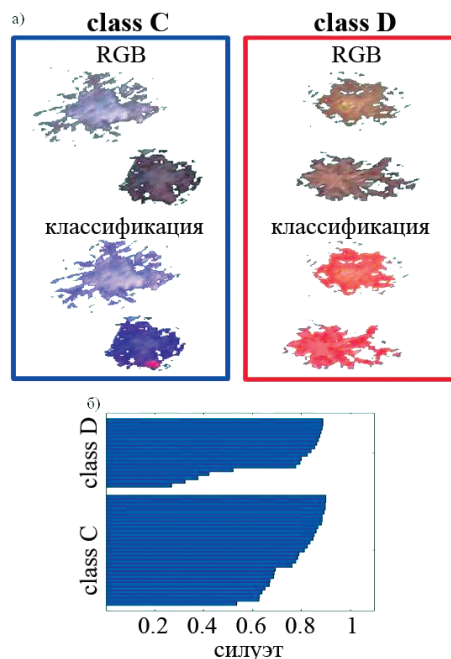


Рис. 4. Обоснование разделимости класса пораженных деревьев на подклассы – тип 1 и тип 2: а) результаты кластерного анализа методом k -средних при разделении на 2 класса; б) силуэты кластеров

Fig. 4. Substantiation of discriminability of the affected trees class to the subclasses type 1 and type 2: а) results of the cluster k -means analysis by fractionation to 2 classes; б) shadow figures of the clusters

Полученные таким образом обучающие данные были использованы для проведения обучаемой попиксельной классификации тестового изображения. Рассмотрено 3 различных типа классификаторов, описанных выше: НБК, КБС и МОВ, для которых проведено сравнение точности и скорости расчета. Результаты классификации контуров крон деревьев приведены на рис. 5. Полученные результаты находятся в хорошем соответствии с визуальным анализом RGB -изображения. Контуров пораженных деревьев классифицированы верно, исключение составляют 2 узких контура, экспертная классификация является спорной. Можно также заметить, что несколько контуров в левой

части изображений были отнесены к классу E, что является положительным результатом, поскольку они соответствуют породе, которая не была включена в обучающую базу данных.

Для оценки ошибок рассматриваемых алгоритмов классификации использовались методы кросс-валидации k -fold [13] и переклассификации (применение классификатора непосредственно к полной обучающей выборке). Близость ошибок переклассификации и кросс-валидации свидетельствует об устойчивости обучения и определяет способность к обобщению.

Общими характеристиками качества обучаемой классификации являются полная вероятность ошибки TE и каппа Коэна κ [13]. Значения этих параметров для 3 рассматриваемых методов классификации представлены в табл. 2. TE определяется как отношение количества ошибочно классифицированных объектов к общему количеству объектов, TE_{res} означает TE, полученную методом переклассификации, а TE_{cv} означает TE, полученную методом кросс-валидации. По данным табл. 2 мы можем заключить, что все рассмотренные методы имеют достаточно высокую точность – 95 %. Близость значений TE_{res} и TE_{cv} свидетельствует об отсутствии проблемы переобучения для всех классификаторов.

Параметр κ представляет собой меру фактического согласия между экспертными данными и результатами автоматической классификации. κ может принимать значения от -1 до 1 . Нулевое значение соответствует полностью случайной классификации. Принято, что значения κ можно шкалировать следующим образом: $[1; 0,8)$ – отлично; $[0,8; 0,6)$ – хорошо; $[0,6; 0,4)$ – приемлемо; $[0,4; 0,2)$ – посредственно; $[0,2; 0,0)$ – плохо; $[0,0; -1]$ – очень плохо. Таким образом, все рассмотренные методы показали отличное соответствие экспертным данным.

Таблица 2

Общие характеристики классификации

Общие характеристики качества	Методы		
	НБК	КБС	МОВ
TE_{res}	0,056	0,054	0,056
TE_{cv}	0,058	0,056	0,056
κ	0,909	0,925	0,922

Значения поклассовых ошибок представлены в табл. 3. Параметр OE означает ошибку пропуска

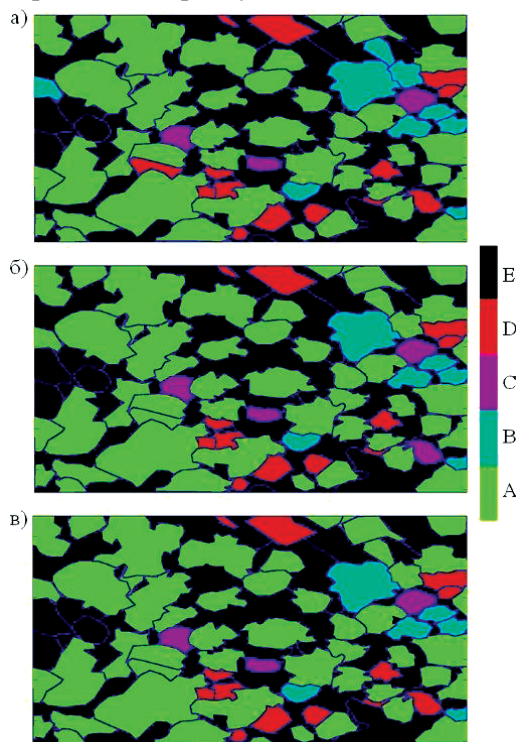


Рис. 5. Классификация контуров крон деревьев различными методами: а) нормальный байесовский классификатор (НБК); б) метод k -ближайших соседей (КБС); в) многоклассовый метод опорных векторов (МОВ)

Fig. 5. Classification of the tree crowns by the various methods: а) normal Bayes classifier, б) k -nearest neighbors algorithm, в) multi-class support vector machine

цели, которая показывает, какую долю объектов заданного экспертом класса классификатор отнес к другим классам. СЕ – ошибка ложной тревоги, показывающая долю объектов, которые классификатор отнес к некоторому заданному классу, но которые, по мнению эксперта, относятся к другим классам.

Таблица 3

Поклассовые ошибки

Метод	Ошибки	Класс пораженных деревьев				
		А	В	С	Д	Е
НБК	ОЕ	0,078	0,099	0,032	0,028	0,044
	СЕ	0,053	0,162	0,011	0,072	0,041
КБС	ОЕ	0,032	0,146	0,025	0,036	0,049
	СЕ	0,074	0,083	0,014	0,046	0,018
МОВ	ОЕ	0,029	0,156	0,027	0,033	0,054
	СЕ	0,080	0,082	0,013	0,048	0,014

Наилучшим образом классифицируются пораженные деревья обоих типов. Различие ошибок для рассматриваемых методов в основном обусловлено разным соотношением между ОЕ и СЕ. Основная часть ошибок возникает при разделении типов 1 и 2. Таким образом, задача оценки степени поражения безотносительно породного состава может быть решена с большей точностью.

Для сравнения вычислительной эффективности классификаторов мы приводим табл. 4, в которой указаны относительные скорости обучения и обработки. Несмотря на то что КБС и МОВ показали несколько меньшие ошибки, вносимые улучшения очевидно являются не столь существенными по сравнению с различиями в скорости обработки. Таким образом, для решения данной задачи достаточно использовать более эффективный НБК.

Таблица 4

Временные затраты на этапе классификации

Временные затраты	НБК	КБС	МОВ
Время обучения (усл. ед.)	1	2,5	900
Время обработки (усл. ед.)	1	10	400

Результаты оценки степени поражения для различных методов представлены в табл. 5 и на рис. 6.

Таблица 5

Распределение деревьев на тестовом участке по категориям поражения

Метод	Категории поражения						МП
	1	2	3	4	5	6	
НБК	53	2	0	1	3	13	33
КБС	54	2	0	0	3	13	33
МОВ	55	2	0	0	3	13	32

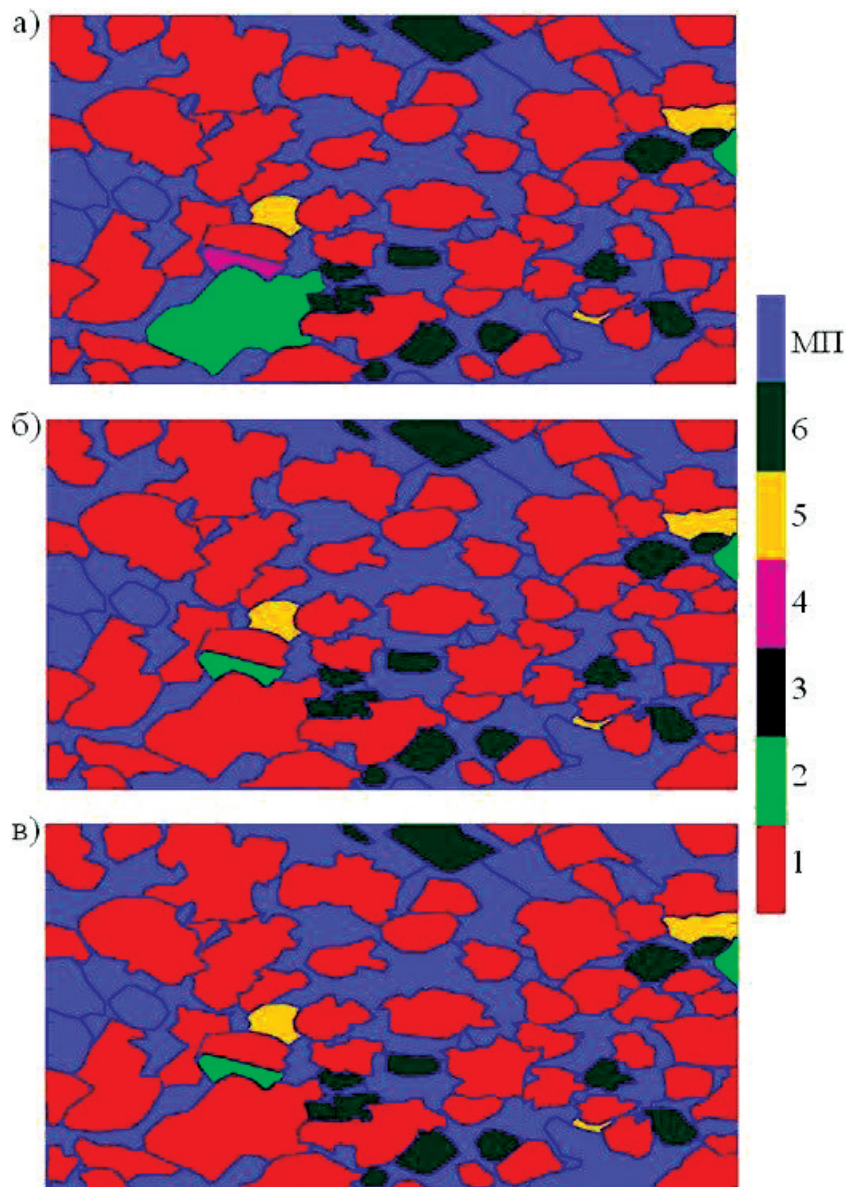


Рис. 6. Оценка степени поражения при использовании различных классификаторов: а) НБК; б) КБС; в) МОВ. Цифрами на цветовой шкале обозначены категории поражения: 1 – здоровые; 2 – ослабленные; 3 – сильно ослабленные; 4 – усыхающие; 5 – свежий сухостой; 6 – старый сухостой. МП означает межкروновое пространство

Fig. 6. Assessment of extent of injury by the use of various classifiers: а) normal Bayes classifier, б) k -nearest neighbors algorithm, в) multi-class support vector machine. The figures on the color scale indicate the categories of infestation: 1 – vigorous, 2 – weakened, 3 – severely weaken, 4 – drying out, 5 – immature snag, 6 – old snag. MP denotes the inter-crown space

Более 50 % тестового участка занято здоровыми деревьями. Основная часть пораженных деревьев отнесена к категории старого сухостоя. При срав-

нении данных на рис. 5 и 6 можно видеть, что категории 6 соответствует класс пораженных деревьев типа 2, при этом категории 5 соответствуют и тип 1, и тип 2 пораженных деревьев. Это свидетельствует о том, что при наличии подробных наземных данных методика распределения степени поражения по категориям 5 и 6, скорее всего, будет скорректирована введением характерных спектральных признаков старого и свежего сухостоя. Однако при имеющихся данных валидировать такую методику не представляется возможным.

СПИСОК ЛИТЕРАТУРЫ / REFERENCES

1. Акулов Е.Н., Кулинич О.А., Пономарев В.Л. Полиграф уссурийский – новый инвазийный вредитель хвойных лесов России // Защита и карантин растений. 2011. № 7. С. 34–36. [Akulov E.N., Kulinich O.A., Ponomarev V.L. *Poly-graphus proximus* is a New Invasive Pest of Softwood Forests in Russia. *Zashchita i karantin rasteniy*, 2011, no. 7, pp. 34–36].
2. Гниненко Ю.И., Клюкин М.С. Уссурийский короед на территории России // Защита и карантин растений. 2011. № 11. С. 32–34. [Gninenko Yu.I., Klyukin M.S. *Polygraphus proximus* in the Territory of Russia. *Zashchita i karantin rasteniy*, 2011, no. 11, pp. 32–34].
3. Керчев И.А. Экология полиграфа уссурийского *Polygraphus proximus* Blandford (Coleoptera; Curculionidae, Scolytinae) в Западно-Сибирском регионе инвазии // Российский журнал биологических инвазий. 2014. № 2. С. 80–94. [Kerchey I.A. Ecology of Four Eyed Fir Bark Beetle *Polygraphus proximus* Blandford (Coleoptera; Curculionidae, Scolytinae) in the West-Siberian Region of Invasion. *Rossiyskiy zhurnal biologicheskikh invaziy* [Russian Journal of Biological Invasions], 2014, no. 2, pp. 80–94].
4. Кобельков М.Е., Чуканов М.А., Хотин Д.В. Категории состояния основных лесобразующих пород Московской области. М., 2000. 40 с. [Kobel'kov M.E., Chukanov M.A., Khotin D.V. Status Categories of the Main Forest Forming Species of the Moscow Region. Moscow, 2000. 40 p.].
5. Кривец С.А., Керчев И.А., Бисирова Э.М., Пашенова Н.В., Демидко Д.А., Пет'ко В.М., Баранчиков Ю.Н. Уссурийский полиграф в лесах Сибири (распространение, биология, экология, выявление и обследование поврежденных насаждений): метод. пособие. Томск; Красноярск, 2015. 48 с. [Krivets S.A., Kerchey I.A., Bisirova E.M., Pashenova N.V., Demidko D.A., Pet'ko V.M., Baranchikov Yu.N. *Four-Eyed Fir Bark Beetle in Siberian Forests (Distribution, Biology, Ecology, Detection and Survey of Damaged Stands): Study Guide*. Tomsk, 2015. 48 p.].
6. Криволицкая Г.О. Семейство Scolytidae (Ipidae) – Короеды / Определитель насекомых Дальнего Востока России. Т. 3, ч. 3. Владивосток: Дальнаука, 1996. С. 312–317. [Krivolitskaya G.O. Family Scolytidae, Bark Beetles. *Key to the Insects of the Russian Far East*. Vol. 3, part 3. Vladivostok, Dal'nauka Publ., 1996, pp. 312–317].
7. Мандельштам М.Ю., Поповичев Б.Г. Аннотированный список видов короедов (Coleoptera, Scolytidae) Ленинградской области // Энтомологическое обозрение, 2000. Т. 79, вып. 3. С. 599–618. [Mandel'shtam M.Yu., Popovichev B.G. Annotated List of Bark Beetles (Coleoptera, Scolytidae) of Leningrad Province. *Entomologicheskoye obozreniye* [Entomological Review], 2000, vol. 79, iss. 3, pp. 599–618].
8. Чилаксаева Е.А. Первая находка *Polygraphus proximus* Blandford, 1894 (Coleoptera, Scolytidae) в Московской области // Бюлл. Московского общества испытателей природы. Отдел биологический. 2008. Т. 113, вып. 6. С. 39–42. [Chilakhsaeva E.A. First Record of *Polygraphus proximus* (Coleoptera: Scolytidae) in Moscow Province. *Byulleten' Moskovskogo obshchestva ispytateley prirody. Otdel biologicheskoy* [Bulletin of Moscow Society of Naturalists. Biological Series], 2008, vol. 113, iss. 6, pp. 39–42].

9. Barbakh W., Fyfe C. Online clustering algorithms. *International Journal of Neural Systems*, 2008, vol. 18, no. 3, pp. 1–10.
10. Dietterich T.G., Bakiri G. Solving Multiclass Learning Problems via Error-Correcting Output Codes. *Journal of Artificial Intelligence Research*, 1995, vol. 2, pp. 263–286.
11. Duda R.O., Hart P.E., Stork D.G. *Pattern Classification*. 2nd Edition. New York, NY, Wiley-Interscience, 2000. 688 p.
12. Freidman J.H., Bentley J.L., Finkel R.A. An Algorithm for Finding Best Matches in Logarithmic Expected Time. *ACM Transactions on Mathematical Software*, 1977, vol. 3, iss. 3, pp. 209–226. DOI: [10.1145/355744.355745](https://doi.org/10.1145/355744.355745)
13. Hastie T., Tibshirani R., Friedman J. *The Elements of Statistical Learning*. New York, NY, Springer, 2001. 536 p. DOI: [10.1007/978-0-387-21606-5](https://doi.org/10.1007/978-0-387-21606-5)
14. Jing L., Hua B., Noland T., Li J. An Individual Tree Crown Delineation Method Based on Multi-Scale Segmentation of Imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2018, vol. 70, pp. 88–98.
15. MacQueen J.B. Some Methods for Classification and Analysis of Multivariate Observations. *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*. Vol. 1: Statistics. Berkeley, CA, University of California Press, 1967, pp. 281–297.
16. Nijijima Y. Revision und Neubeschreibung der Polygraphus-Arten (Coleoptera, Ipidae) in Japan. *Insecta Matsumurana*, 1941, vol. 15(4), pp. 123–135.
17. Nobuchi A. Bark-Beetles Injurious to Pine in Japan. *Bulletin of the Government Forest Experiment Station*, 1966, vol. 185, pp. 1–49.
18. Safonova A., Tabik S., Alcaraz-Segura D., Rubtsov A., Maglinets Y., Herrera F. Detection of Fir Trees (*Abies sibirica*) Damaged by the Bark Beetle in Unmanned Aerial Vehicle Images with Deep Learning. *Remote Sensing*, 2019, vol. 11(6), art. 643, pp. 1–19. DOI: [10.3390/rs11060643](https://doi.org/10.3390/rs11060643)

CLASSIFICATION AND ASSESSMENT OF THE STATE OF MIXED FORESTS FROM VERY HIGH SPATIAL RESOLUTION AIRBORNE IMAGES*

E.V. Dmitriev^{1,2}, Candidate of Physics and Mathematics, Senior Research Scientist; ResearcherID: [E-4794-2014](https://orcid.org/0000-0001-5363-3934), ORCID: [0000-0001-5363-3934](https://orcid.org/0000-0001-5363-3934)

V.A. Kozub², External Student

P.G. Melnik^{3,4}, Candidate of Agriculture, Assoc. Prof.; ResearcherID: [E-7644-2014](https://orcid.org/0000-0001-9236-5864)

A.A. Sokolov⁵, Candidate of Physics and Mathematics; ORCID: [0000-0001-9236-5864](https://orcid.org/0000-0001-9236-5864)

A.N. Safonova^{6,7}, PhD Student; ResearcherID: [F-6811-2019](https://orcid.org/0000-0002-3290-2717), ORCID: [0000-0002-3290-2717](https://orcid.org/0000-0002-3290-2717)

¹Marchuk Institute of Numerical Mathematics of the Russian Academy of Sciences, ul. Gubkina, 8, Moscow, 119333, Russian Federation; e-mail: yegor@mail.ru

²Moscow Institute of Physics and Technology (National Research University), per. Institutskiy, 9, Dolgoprudny, Moscow Region, 141701, Russian Federation; e-mail: postbox-kozub@yandex.ru

³Mytishchi Branch of Bauman Moscow State Technical University, ul. 1-ya Institutskaya, 1, Mytishchi-5, Moscow Region, 141005, Russian Federation; e-mail: melnik_petr@bk.ru

⁴Institute of Forest Science, Russian Academy of Sciences, ul. Sovetskaya, 21, Uspenskoe, Moscow Region, 143030, Russian Federation; e-mail: melnik_petr@bk.ru

*The article is published within the framework of implementation of the scientific journals development program in 2019, based on proceedings of XIX International conference of the young scholars “Forests of Euro-Asia – The Southern Ural” (August 25–30, 2019, Chelyabinsk).

⁵Laboratoire de Physico-Chimie de l'Atmosphère Université du Littoral Côte d'Opale, Maison de la Recherche en Environnement Industriel 2, 189A, Avenue Maurice Schumann, 59140 Dunkerque, France; e-mail: anton.sokolov@univ-littoral.fr

⁶Siberian Federal University, prosp. Svobodny, 79, Krasnoyarsk, 660041, Russian Federation; e-mail: safonova.nastya1@gmail.com

⁷Soft Computing and Intelligent Information Systems research group, University of Granada, E-18071 Granada, Spain; e-mail: safonova.nastya1@gmail.com

At present, the invasion by Ussuri polygraphus (*Polygraphus proximus* Blandf) is considered as one of the main factors of large-scale drying of Siberian forests. The appearance of this new organism in fir trees has led to seriously worsening their condition and a variety of ecological effects in taiga ecosystems. The strong decrease of natural biological diversity, forest productivity, changes in the composition and structure of tree and subordinate layers may occur in the centers of mass reproduction. In this paper, we propose a method for determination of category of forest damage from very high spatial resolution color airborne images (5–10 cm per pixel) using machine learning methods. The method includes the stages of preprocessing, segmentation of crowns of individual trees, the classification and assessment of the forest damage in accordance with conventional standards. The images of several test plots of Stolby Nature Reserve (Krasnoyarsk Territory), obtained with the help of equipment installed on unmanned aerial vehicles DJI Phantom 3 Pro and Yuneec Typhoon H in May 2016, were used for testing the method proposed. The filtering method proposed for the stage of constructing a training set made it possible to increase the accuracy at the classification stage. The substantiation of division of the three main classes of objects into subclasses using cluster analysis is given. The presence of subclasses is caused by presence of various tree species in the test plot. A comparison of the efficiency of various supervised classification methods used for solving this problem is performed. It is shown that all the considered methods allow us to achieve a sufficiently high accuracy, about 95%. The calculation of the Cohen's kappa coefficient shows that the classifications carried out with the help of all the considered methods have excellent agreement with the expert data. The analysis of the stability of training is carried out. Estimates of the total probability of error obtained by methods of cross-validation and resubstitution differ by less than 0.1%, which indicates the absence of the problem of overtraining. The joint analysis of accuracy and processing speed has shown that it is most appropriate to use the normal Bayesian classifier. High classification accuracy allows us to obtain estimates of 6 categories of forest damage in the test plot. The results obtained can be potentially used by regional forest management services.

For citation: Dmitriev E.V., Kozub V.A., Melnik P.G., Sokolov A.A., Safonova A.N. Classification and Assessment of the State of Mixed Forests from Very High Spatial Resolution Airborne Images. *Lesnoy Zhurnal* [Forestry Journal], 2019, no. 5, pp. 9–24. DOI: 10.17238/issn0536-1036.2019.5.9

Funding: the work is supported by RFBR, project № 19-01-00215 «Investigation of operative opportunities of hyper-spectral technologies of remote sensing of the Earth to solve regional problems using updated hyper-spectral cameras from space».

Keywords: remote sensing, very high resolution images, forests, forest state assessment, forest degradation, stem pests, pattern recognition, thematic image processing.

Поступила 30.07.19 / Received on July 30, 2019
