

Автоматическое детектирование просек под линиями электропередач на снимках Sentinel-2 с помощью методов машинного обучения и компьютерного зрения

Я. О. Бахрамхан¹, Д. М. Ермаков^{2,3}, Е. С. Подольская^{1,4}

¹ *Национальный исследовательский университет «Высшая школа экономики»
Москва, 109028, Россия*

E-mails: ybahramhan@gmail.com, podols_kate@mail.ru

² *Институт космических исследований РАН, Москва, 117997, Россия
E-mail: d.m.ermakov@cosmos.ru*

³ *Фрязинский филиал Института радиотехники и электроники
им. В. А. Котельникова РАН, Фрязино, Московская обл., 141190, Россия*

⁴ *Центр по проблемам экологии и продуктивности лесов им. А. С. Исаева РАН
Москва, 117997, Россия*

Регулярный мониторинг состояния линий электропередач (ЛЭП) необходим для бесперебойного электроснабжения населённых пунктов и объектов инфраструктуры. Однако на лесных территориях с низкой плотностью населения и дорог задача постоянного мониторинга усложняется. Для дистанционного мониторинга просек под ЛЭП в труднодоступных лесных территориях настоящим исследованием предлагается алгоритм, выделяющий лесные просеки на космических снимках Sentinel-2. Такой алгоритм может рассматриваться в качестве первого шага автоматизации дистанционного мониторинга состояния просек под ЛЭП и, в частности, детектирования изменений (прокладки, зарастания, расчистки просек под ЛЭП и т.д.). Логика работы алгоритма имеет наглядную интерпретацию и основывается на поиске объектов интереса в пространстве дешифровочных признаков. На основе спектральных характеристик модель логистической регрессии выделяет участки, потенциально относящиеся к просекам, затем вероятностное преобразование Хафа находит линейные структуры на бинарной маске этих участков. Разработанный алгоритм является полностью автоматическим — он не требует индивидуальной настройки параметров для каждого снимка. Для оценки результатов созданы метрики качества выделения просек. Работа алгоритма локально устойчива — алгоритм корректно определяет области расположения просек под ЛЭП на всех использованных спутниковых изображениях в пределах региона исследования.

Ключевые слова: просеки, ЛЭП, дистанционный мониторинг, признаки дешифрирования, интерпретируемость алгоритма, преобразование Хафа, логистическая регрессия

Одобрена к печати: 13.05.2025

DOI: 10.21046/2070-7401-2025-22-4-11-26

Введение

Регулярный мониторинг объектов энергетической инфраструктуры, включая линии электропередач (ЛЭП) и лесные просеки под ними, необходим для предотвращения аварий на высоковольтных линиях. В частности, зарастание лесных просек высокорослой растительностью вызывает серьёзные проблемы и перебои в электроснабжении. Для предотвращения подобных случаев проводят наземный осмотр лесных просек под ЛЭП, однако периодический мониторинг лесных просек на месте затруднён в регионах России с высокой лесистостью и низкой плотностью населения и дорожной сети.

Для мониторинга состояния ЛЭП и просек под ними применяются разнообразные данные дистанционного зондирования Земли. В задачах исследования лесных просек используют снимки радио- и оптического диапазонов, данные лазерной съёмки и съёмки в тепловом диапазоне. Стоит отметить, что выбор пространственного разрешения данных играет ключевую роль в выделении как ЛЭП, так и лесных просек под ними. На снимках высокого разрешения очертания ЛЭП и просек под ними более точные, а на данных среднего и низкого разрешения границы лесных просек довольно сглажены.

Первые работы по исследованию отражения волн радиодиапазона от ЛЭП в 1990-х гг. носили преимущественно экспериментально-поисковый характер. В одной из таких публикаций (Sarabandi et al., 1994) исследователи изучили обратное рассеяние радиоволн от кабелей ЛЭП в диапазонах электромагнитных волн С, К и Х на радиолокационных изображениях с размером пикселя около 12 м и зависящим от синуса угла падения радара. Это исследование получило продолжение в работе (Sarabandi, Park, 1999) о рассеянии радиоволн в W-диапазоне кабелем ЛЭП, поверхность которого покрыта слоем воды и льда. Информация о точном разрешении прибора в статье отсутствует.

Более современные работы по использованию данных радиолокационной съёмки для выделения ЛЭП направлены на практическое применение данных определённого спутника или на съёмку с использованием наземных воздушных средств для выделения ЛЭП на конкретной территории. Так, в работе (Schwarz et al., 2009) с помощью частотно-временного разложения выделены пилоны ЛЭП на снимках радиодиапазона спутника TerraSAR-X с разрешением от 1 до 3 м в зависимости от режима съёмки. В исследовании (Lin et al., 2012) пилоны высоковольтных ЛЭП распознаны при съёмке высокого разрешения в Р-диапазоне.

Оптические снимки начали применять несколько позднее по сравнению с данными радиосъёмки. В качестве одной из первых можно назвать публикацию (Bernstein, Di Gesù, 1999) о распознавании высоковольтных линий на снимках с размером пикселя 1 м. Работы 2000–2010 гг. по выделению ЛЭП на снимках видимого диапазона больше сосредоточены на создании стереопар изображений — это позволяет определить не только пилоны, но и высоту растительности на просеке под ЛЭП. Среди таких исследований — работа по использованию снимков спутника Quickbird для расчёта высоты растительности на просеке (Moeller, 2006); статья (Kobayashi et al., 2009), посвящённая созданию стереопары для поиска высоко-рослой растительности под ЛЭП; исследование по применению анализа стереовидения для идентификации древесной растительности (Ahmad et al., 2011). Во всех приведённых работах использовались данные высокого разрешения в пределах первых единиц метра.

В исследованиях по детектированию как высоковольтных линий, так и просек под ними по спутниковым данным применяются алгоритмы автоматической и полуавтоматической обработки.

В одной из ранних работ по использованию алгоритмов полуавтоматического выделения ЛЭП на оптических снимках (Bernstein, Di Gesù, 1999) исследователи предложили методику поиска линий электропередач с помощью эквализации изображения для уменьшения влияния неравномерного освещения и морфологического анализа. Как установлено в указанной работе, алгоритм отлично показал себя на тестовых участках. Тем не менее авторы отметили зависимость результатов от типа наземного покрова.

В исследованиях 2000-х гг. стали активно применять методы компьютерного зрения, хотя большинство алгоритмов продолжали оставаться полуавтоматическими. В работе (Tong et al., 2009) для выделения ЛЭП на снимках высокого разрешения, полученных наземными воздушными средствами, применён алгоритм Марра–Хилдрета в сочетании с преобразованием Хафа. В статье (Mu et al., 2009) по аналогии с предыдущей использовано преобразование Хафа, но в качестве детектора границ выбран фильтр Габора. Алгоритм успешно распознавал ЛЭП на снимках с размером пикселя 4 см, однако вместе с ними он часто выделял другие линейные объекты.

В современных работах исследователи чаще используют автоматические алгоритмы машинного обучения. Так, в публикации (Mahdi Elsiddig Haroun et al., 2021a) использована кластеризация К-средних на снимках высокого разрешения с беспилотного летательного аппарата. Предварительный перевод цветового пространства изображений из RGB в HSV (*англ.* Hue, Saturation, Value) позволил увеличить контрастность на границах объектов, что повысило точность выделения участков зарастания просек под ЛЭП. В другом исследовании (Mahdi Elsiddig Haroun et al., 2021b) был применён метод опорных векторов для классификации изображений высокого разрешения. Авторы использовали широкий набор признаков, среди которых есть признаки матриц совместной встречаемости на уровне серого GLCM (*англ.* Gray-level Co-occurrence Matrices), локальные бинарные шаблоны LBP (*англ.*

Local Binary Patterns) и статистические моменты трёх цветовых пространств — RGB, HSV и CIELAB (точнее, CIE 1976 L*a*b*). Самый лучший по метрикам качества классификатор на основе метода опорных векторов получен при использовании всех признаков за исключением признаков LBP (*англ.* Local Binary Patterns).

Настоящим исследованием предлагается автоматический алгоритм выделения лесных просек под ЛЭП на космических снимках Sentinel-2. Комплексный алгоритм состоит из интерпретируемых методов машинного обучения и компьютерного зрения, которые позволяют заменить ручное дешифрирование лесных просек и ускорить процесс их распознавания. Особенно это актуально для больших территорий, где затраты времени на детекцию лесных просек значительно сокращаются, а качество выделения остаётся высоким.

В начале статьи изложен опыт применения данных дистанционного зондирования и, в частности, спутниковых снимков Sentinel-2 для мониторинга состояния лесных просек под ЛЭП, а также методы машинного обучения и компьютерного зрения, которые используются для автоматического выделения объектов энергетической инфраструктуры в лесных ландшафтах. Далее описывается спецификация нашего алгоритма, призванного учесть недостатки предыдущих методик выделения линейных объектов. В последней части исследования рассказывается про разработку алгоритма и полученные результаты, а также отмечаются его преимущества и недостатки.

Использованные данные и их предварительная обработка

С точки зрения изменения структуры лесных просек интересен район расположения Богучанской гидроэлектростанции (ГЭС) на востоке Красноярского края, где произошло значительное увеличение числа лесных просек из-за строительства электростанции (*рис. 1*).

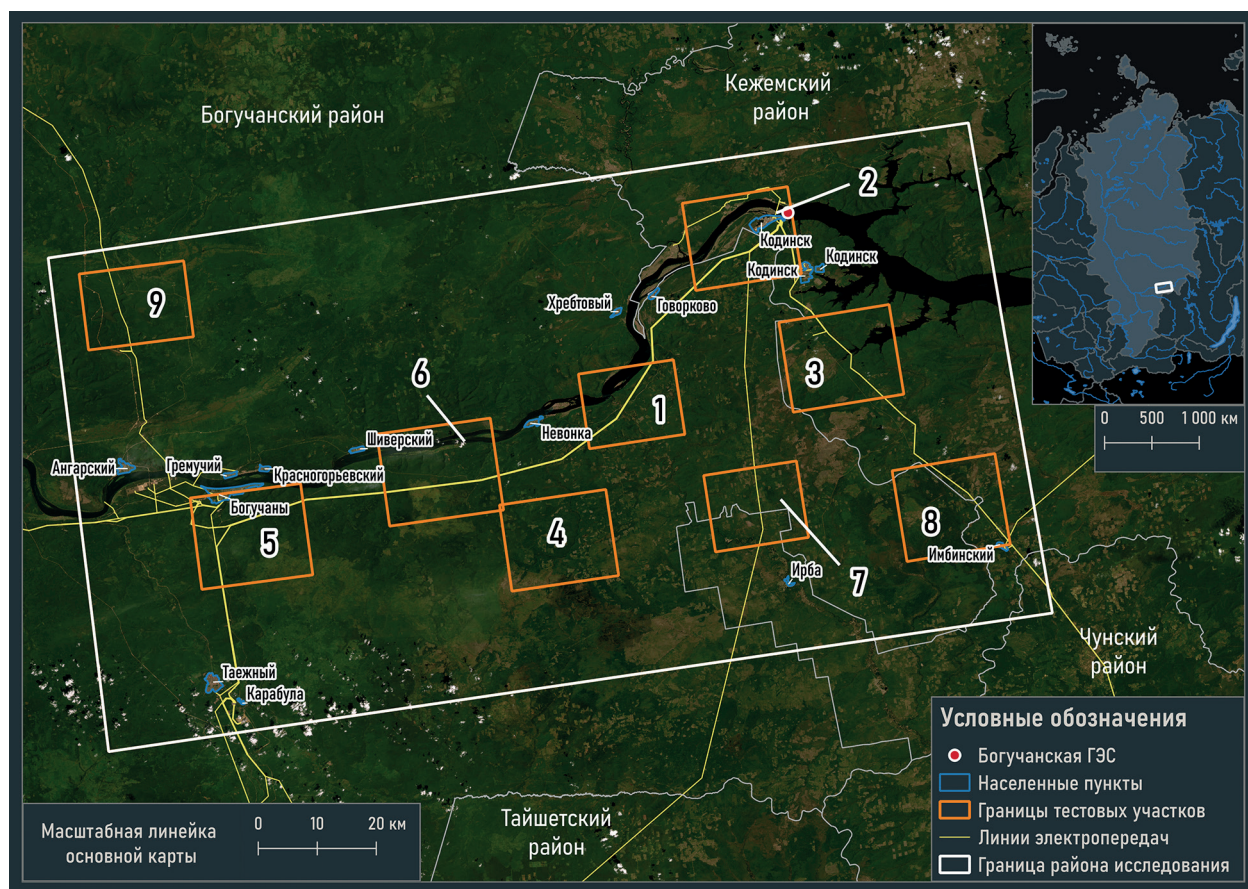


Рис. 1. Карта района исследования и расположения тестовых участков

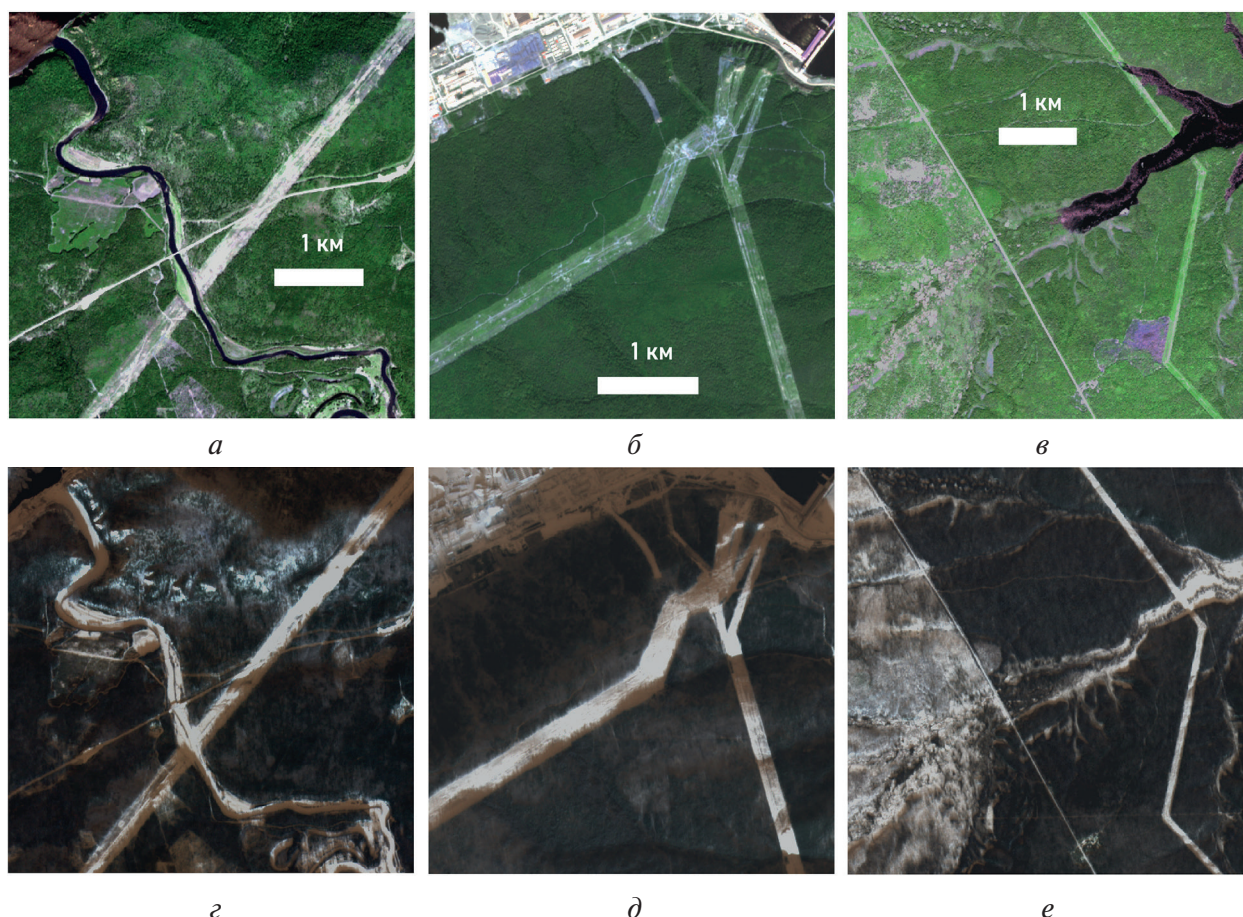


Рис. 2. Примеры летних (а–в) и зимних (г–е) космических снимков на экспериментальные участки

Выбор данной территории также обусловлен высокими показателями лесистости и низкой плотностью населения, что соответствует целям исследования. Всего на территорию интереса подготовлены 18 снимков Sentinel-2, девять из которых получены в тёплое время года (с мая по сентябрь), а остальные девять — в холодное (с января по февраль). Таким образом, на каждый из девяти участков собрано по два снимка. Снимки получены с помощью системы Google Earth Engine (Gorelick et al., 2017).

Для эксперимента по выбору модели машинного обучения отобрано девять участков, три из них были выбраны для разметки. Длина стороны границы этих трёх участков составляет в среднем 5 км. Отобранные тестовые участки демонстрируют высокую степень репрезентативности. В частности, на большинстве из них присутствуют автомобильные дороги. Участки № 1 (рис. 2а, г) и 3 (см. рис. 2в, е) характеризуются наличием лесных опушек, а также участков с открытой почвой и разреженной растительностью. Вторым участком (см. рис. 2б, д) выделяется расположенной на нём городской застройкой.

Лесная просека под ЛЭП обладает такими дешифровочными признаками, как высокая спектральная яркость в коротковолновом инфракрасном диапазоне SWIR (*англ.* short wave infrared) и линейная форма. Эти особенности проявляются как на летних, так и на зимних изображениях. На летних снимках в видимом диапазоне просека имеет ярко-зелёный цвет, что контрастирует с тёмно-зелёным цветом леса. В зимний период просека, покрытая снегом, заметно выделяется на фоне тёмного леса.

На основе дешифровочных признаков отобраны признаки, обеспечивающие выделение лесных просек под ЛЭП от других объектов. Среди признаков дешифрирования для использования в разрабатываемом алгоритме выбраны следующие:

- каналы видимого спектра, а также ближнего NIR (*англ.* near-infrared) и дальнего инфракрасного диапазонов летних изображений, полученных спутником Sentinel-2;

- каналы видимого спектра зимних изображений Sentinel-2;
- спектральные индексы, позволяющие подчеркнуть спектральные характеристики объектов на снимках.

На основе набора данных каналов зимнего и летнего снимка созданы три спектральных индекса: индекс растительности с поправкой на почву SAVI (*англ.* Soil-Adjusted Vegetation Index) (Huete, 1988), нормализованный разностный водный индекс NDWI (*англ.* Normalized Difference Water Index) (McFeeters, 1996), нормализованный разностный индекс городской застройки NDBI (*англ.* Normalized Difference Built-up Index) (Xu, 2008). Стоит отметить, что инфракрасный и коротковолновый инфракрасный каналы зимнего снимка исключены из анализа, так как лесные просеки в этих каналах оказывались под тенью от высоких деревьев из-за низких углов падения солнечных лучей в зимнее время. В результате использовано 11 признаков, включая 8 каналов изображений и 3 спектральных индекса.

Распределение спектральной яркости в ряде каналов не отражает действительное распределение яркости из-за наличия бликов от городской застройки. Для решения проблемы использовано контрастирование с насыщением. После этого проведена предварительная нормализация всех входных признаков.

Методические основы алгоритма выделения лесных просек под линиями электропередач

Алгоритмы, описанные выше, опробованы на небольших тестовых участках, и некоторые из них обладают низкой степенью интерпретируемости. К примеру, алгоритм из статьи (Bernstein, Di Gesù, 1999) необходимо адаптировать к другим изображениям, так как он применялся исключительно на тестовых данных, и поэтому у него наблюдается высокая чувствительность к наземному покрову и погодным условиям. Алгоритм из публикации (Mahdi Elsiddig Haroun et al., 2021b) использует избыточное количество признаков, причём многие из них не поддаются интуитивной трактовке, поэтому алгоритм не поддаётся интерпретации. Преобразование цветового пространства из RGB в HSV в статье (Mahdi Elsiddig Haroun et al., 2021a) также усложняет интерпретацию алгоритма.

Идея настоящего алгоритма заключается в том, чтобы модель классификации выделяла просеку на основе её спектральных характеристик в класс низкорослой растительности или лугов, а вероятностное преобразование Хафа выделило просеку как кусочно-линейный объект на маске лугов. Модель логистической регрессии оказалась самой точной в составлении карты классификации и в выделении класса лугов в частности, а вероятностное преобразование Хафа широко применяется в задаче выделения линейных объектов на изображении (Бахрамхан и др., 2024).

В качестве меток классов для карты классификации выбраны следующие:

- класс № 1 — водные объекты;
- класс № 2 — леса;
- класс № 3 — редколесья;
- класс № 4 — луга (или свободные от растительности земли);
- класс № 5 — антропогенные объекты.

Отбор пикселей изображений Sentinel-2, относящихся к тому или иному классу, проведён в программе QGIS. С помощью кривой обучения выбрано оптимальное соотношение обучающей и тестовой выборки пикселей, равное 80:20.

Для получения карты классификации использована логистическая регрессия, обученная на отобранной выборке пяти классов. Выборка состоит из пикселей, поэтому прогноз метки даётся для отдельного пикселя. Из карты классификации извлекается класс лугов (№ 4). Маска лугов может быть частично фрагментирована вследствие шума в исходных данных и ошибок классификации. Поэтому для бинарной маски лугов применяется двоичное

открытие (Chen, Haralick, 1995) — серия операций математической морфологии для связывания близких фрагментов маски на бинарном изображении без увеличения её внешних размеров. После этого очищенная от шума бинарная маска подвергалась гауссову фильтру размытия с параметром сигма, равным пяти. Это позволило исказить формы объектов на маске, однако лесные просеки под ЛЭП не потеряли линейность формы. Для размытого изображения применялась бинаризация на основе порогового значения, которая привела изображение к бинарному виду. Последний шаг перед использованием вероятностного преобразования Хафа — поиск границ с помощью расчёта разности между полученной маской и результатом применения двоичной эрозии на этом результате. Двоичная эрозия (Chen, Haralick, 1995) стирает граничные пиксели бинарной маски следующим образом: если в заданном по радиусу скользящем окне находится хотя бы один нулевой пиксель, то значение центрального пикселя равно нулю. Таким образом, в результирующее изображение попадают все пиксели объектов, которые не относятся к границе объектов. На основе расчёта разности между исходным изображением и результатом двоичной эрозии можно получить маску границ объектов. Наконец, на маске границ применяется вероятностное преобразование Хафа (Duda, Hart, 1972), которое выделяет линейные структуры на изображении. Выделенные линии относятся к просекам под ЛЭП.

Для количественной оценки результатов выделения просек под ЛЭП необходимо применить метрику качества выделения. Однако широко используемые в машинном обучении метрики совпадения областей (такие как «пересечение к объединению», IoU (*англ.* Intersection over Union)) не применимы для нашей задачи, поскольку отсутствуют исходные данные о границах лесных просек, а есть только данные о положениях ЛЭП. Поэтому для оценки качества выделения лесных просек под ЛЭП применены сразу две собственные метрики: метрика близости и метрика непрерывности выделения.

Метрика близости показывает среднюю удалённость выделенных границ просек от ЛЭП. Для расчёта этой метрики совершаются следующие шаги:

- 1) растеризация векторного слоя линий электропередач от OpenStreetMap (<https://planet.osm.org>) (*рис. 3а и 4а*, см. с. 17);
- 2) расчёт матрицы евклидова расстояния от пикселей, принадлежащих к ЛЭП, до остальных пикселей участка (см. *рис. 3б и 4б*);
- 3) маскирование матрицы расстояний выделенными границами лесных просек (см. *рис. 3в и 4в*);
- 4) расчёт медианы значения расстояния в пикселях и перевод в метры (1 пиксель изображения Sentinel имеет пространственное разрешение 10 м).

Медиана выбрана в качестве меры центральности ввиду того, что распределение значений расстояний от ЛЭП до пикселя матрицы значительно отклоняется от нормального. Для проверки на нормальность распределения значений расстояний после маскирования выбраны тесты Андерсона—Дарлинга и Жарка—Беры (*табл. 1*). По результатам статистических тестов для всех участков распределение расстояний отклоняется от нормального на уровне значимости p -value ниже 0,01, поэтому гипотеза о нормальности распределения для всех участков отклоняется. Выбор медианы в качестве меры центральности вполне обоснован.

В лучшем случае расстояние между ЛЭП и границами просек под ними достигает небольших значений, ведь просека практически точно выделена (см. *рис. 3*). В худшем случае заметно, что некоторые пиксели приобретают достаточно большое значение расстояния из-за того, что просеки выделены ошибочно вдалеке от действительных границ (см. *рис. 4*). Получившаяся метрика близости ненормированная — значение расстояния может увеличиваться до бесконечности, а наилучшие значения метрики варьируются от 2 до 4 пикселей (от 20 до 40 м), ведь границы лесных просек под ЛЭП расположены на определённом расстоянии от линий электропередач и не могут пересекаться ими. Для нормализации метрики близости из медианного расстояния в пикселях мы рассчитали обратное ему значение посредством деления единицы на исходное значение.

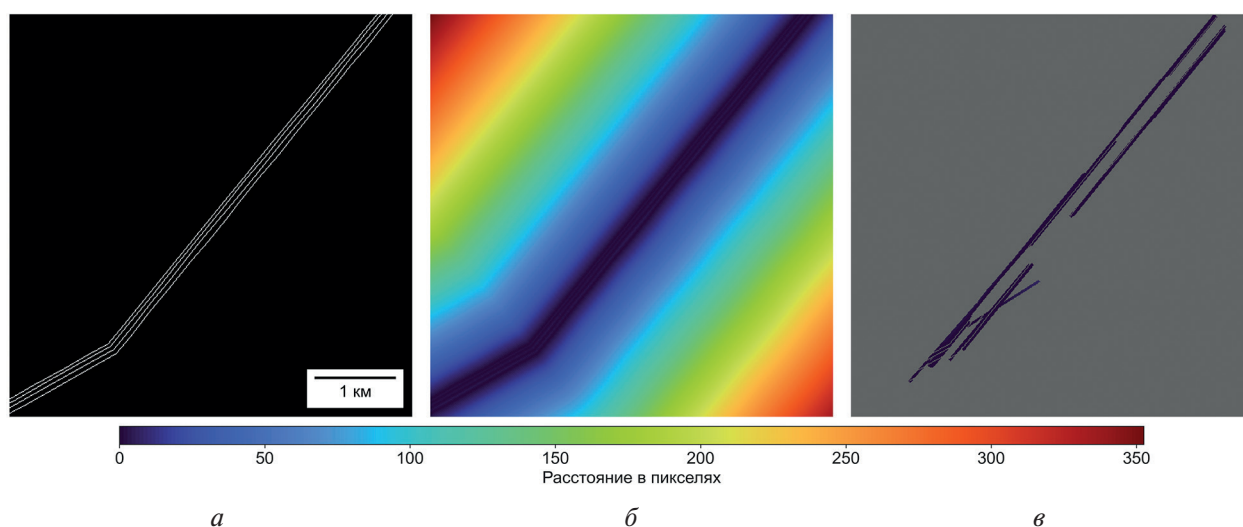


Рис. 3. Визуализация шагов 1–3 по расчёту метрики близости в лучшем случае: *a* — растрезированный слой линий электропередач; *б* — матрица расстояний; *в* — маскированная матрица расстояний

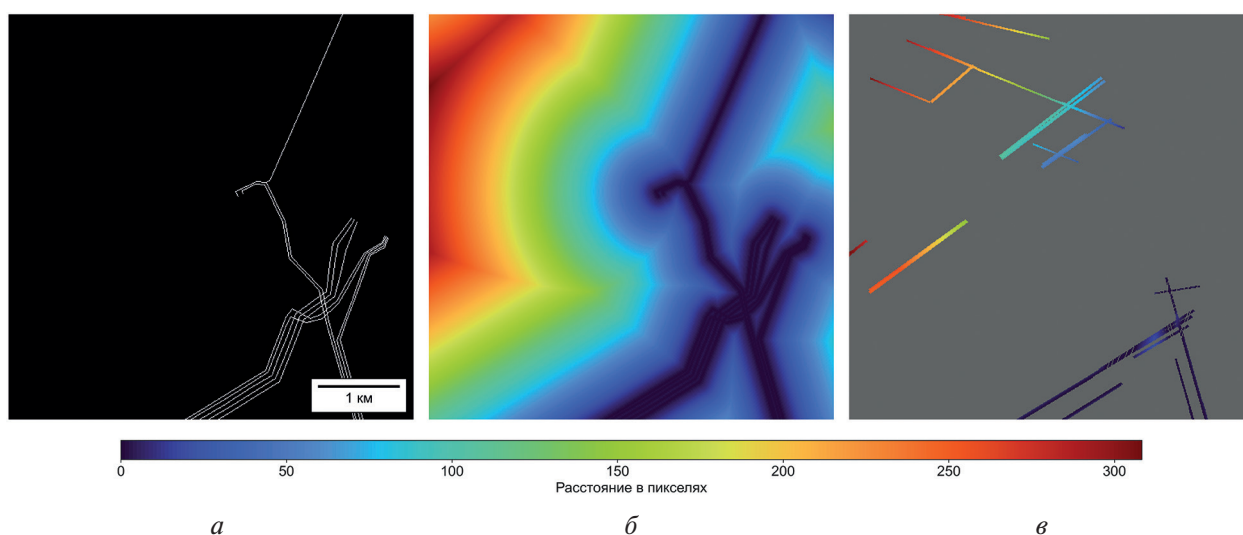


Рис. 4. Визуализация шагов 1–3 по расчёту метрики близости в худшем случае: *a* — растрезированный слой линий электропередач; *б* — матрица расстояний; *в* — маскированная матрица расстояний

Таблица 1. Результаты тестов на нормальность распределения маскированных значений расстояний для участков

Номер участка	Адренсона – Дарлинг (статистика; критическое значение для p -value = 0,01)	Жарка – Беры (статистика; p -value)
1	3806,7; 1,1	786657,5; 0,0
2	3656,8; 1,1	5655,9; 0,0
3	23,0; 1,1	82,4; 0,0
4	—	—
5	6941,8; 1,1	21241,2; 0,0
6	2881,8; 1,1	3638,8; 0,0
7	1519,2; 1,1	7193,0; 0,0
8	2162,6; 1,1	2898,1; 0,0
9	6702,3; 1,1	27111,0; 0,0

Метрика непрерывности выделения границ лесных просек показывает меру того, насколько непрерывно выделены лесные просеки на всей их протяжённости. Как и в случае с метрикой близости проводится растеризация векторного слоя, но не ЛЭП, а их пилонов в виде точек из OpenStreetMap. Перед растеризацией точечный слой пилонов подвергался операции исключения дублей и неинформативных вертексов векторного слоя пилонов ЛЭП, в результате чего сокращённый слой не потерял пространственную репрезентативность. С растеризованным слоем производится итерация по пилонам ЛЭП — для каждого пилона вычисляется матрица расстояний (рис. 5а–в), затем матрица маскируется и находится минимальное значение расстояния среди немаскированных значений (см. рис. 5г–е). Рассчитывается среднее из минимальных расстояний посредством суммирования минимальных расстояний и деления суммы на число пилонов.

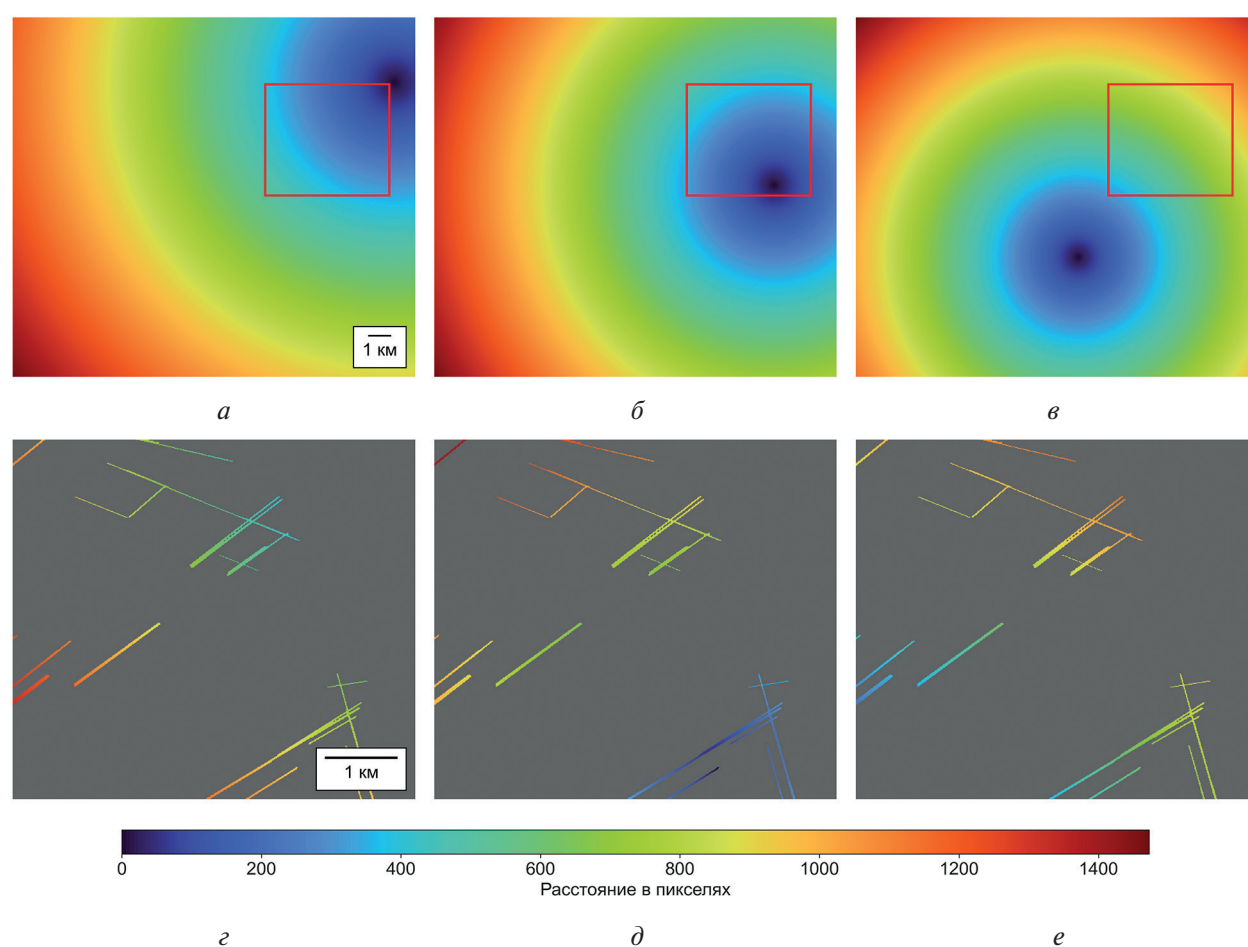


Рис. 5. Визуализация шагов по расчёту метрики непрерывности: а–в — матрицы расстояний для некоторых пилонов ЛЭП; г–е — маскированные матрицы. Прямоугольником обозначены границы врезок маскированных матриц

Как и в предыдущем случае метрика непрерывности ненормированная — в самом худшем исходе, когда просека выделяется на достаточно большом расстоянии от пилонов ЛЭП, она может возрастать до бесконечности. Для этого снова вычислено обратное значение расстояния.

Две созданные метрики обладают такими недостатками, как отсутствие учёта кривизны земной поверхности при расчёте расстояния, однако в нашем случае при небольших размерах тестовых участков кривизной поверхности можно пренебречь. Кроме того, метрики невозможно вычислить при отсутствии векторных данных о расположении ЛЭП и их пилонов, поэтому мы не смогли определить обе метрики на 4-й участок и метрику непрерывно-

сти на 7-й и 9-й участки. Точность метрик и погрешность расчётов зависит от пространственного разрешения изображения — при меньшем пространственном разрешении погрешность метрики существенно сокращается.

Результаты и обсуждение

Модель логистической регрессии позволяет дать понятную интерпретацию результатов. На *рис. 6* приведена матрица важности коэффициентов пяти уравнений регрессии, прогнозирующих вероятность принадлежности пикселя к классу, для которого строится уравнение. Объект интереса — класс лугов — характеризуется не только высокой спектральной яркостью в коротковолновом инфракрасном канале, но и высокой яркостью в синем канале зимнего снимка. Модель логистической регрессии успешно выявила реальную зависимость, продемонстрировав положительные коэффициенты для указанных признаков. Это свидетельствует о том, что с увеличением яркости в этих каналах возрастает вероятность того, что пиксель принадлежит к классу лугов. Таким образом, модель пригодна для дальнейшего извлечения линейных объектов.

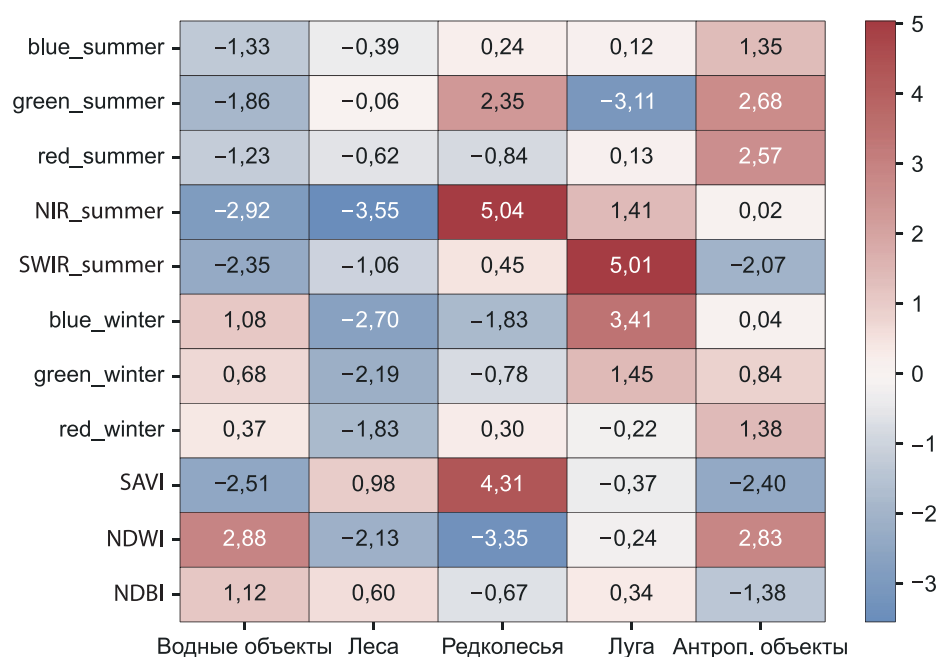


Рис. 6. Матрица важности коэффициентов модели логистической регрессии. По рядам расположены коэффициенты уравнений регрессии для конкретного признака, по столбцам — коэффициенты уравнения регрессии для конкретного класса. Условные обозначения каналов: summer — канал летнего снимка, winter — канал зимнего снимка, blue — синий, green — зелёный, red — красный

Наборы карт на каждый из трёх экспериментальных участков (*рис. 7–9*, см. с. 20–21) показывают последовательный процесс извлечения лесных просек из бинарной маски класса лугов. На участках алгоритм успешно распознал просеки практически на всей их протяжённости, поэтому качество выделения лесных просек можно назвать удовлетворительным. Стоит отметить, что алгоритм наиболее эффективно выявляет более прямолинейные участки лесных просек под ЛЭП, как, например, на 1-м участке (см. *рис. 7е*), а в местах изгиба алгоритм ошибается. Форма участка просеки на 1-м участке не обладает разветвлениями, как на 2-м (см. *рис. 8*), или значительными изгибами части лесной просеки, как на 3-м (см. *рис. 9*). На 2-м участке из-за высокого значения минимальной длины линии вероятностного преобразования Хафа алгоритм не распознаёт небольшой участок изгиба просеки под ЛЭП (см. *рис. 8г*).

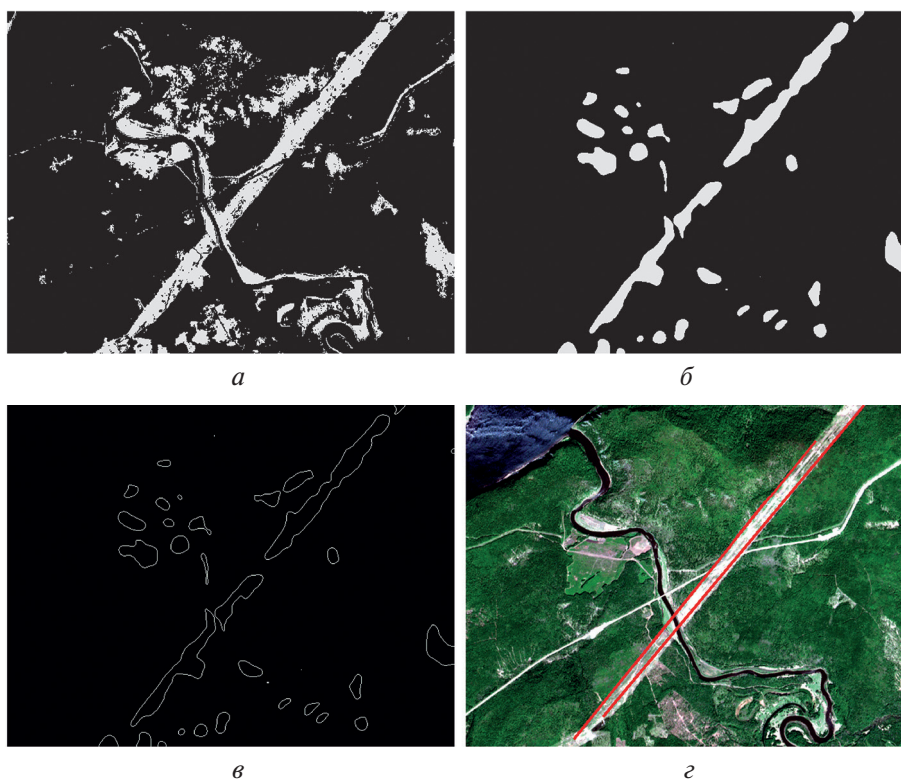


Рис. 7. Результат выделения лесных просек на участке № 1: *а* — маска лугов из карты классификации; *б* — маска лугов после гауссова размытия и бинаризации; *в* — результат разности между маской *б* и её двоичной эрозией; *г* — результат выделения лесных просек

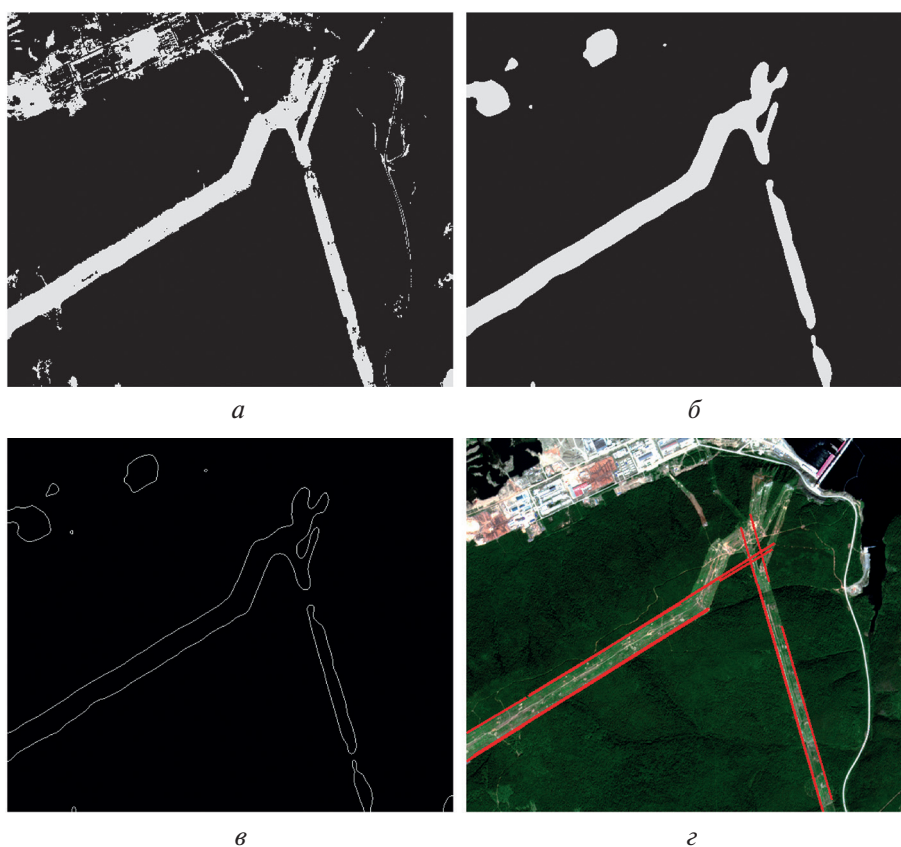


Рис. 8. Результат выделения лесных просек на участке № 2: *а* — маска лугов из карты классификации; *б* — маска лугов после гауссова размытия и бинаризации; *в* — результат разности между маской *б* и её двоичной эрозией; *г* — результат выделения лесных просек

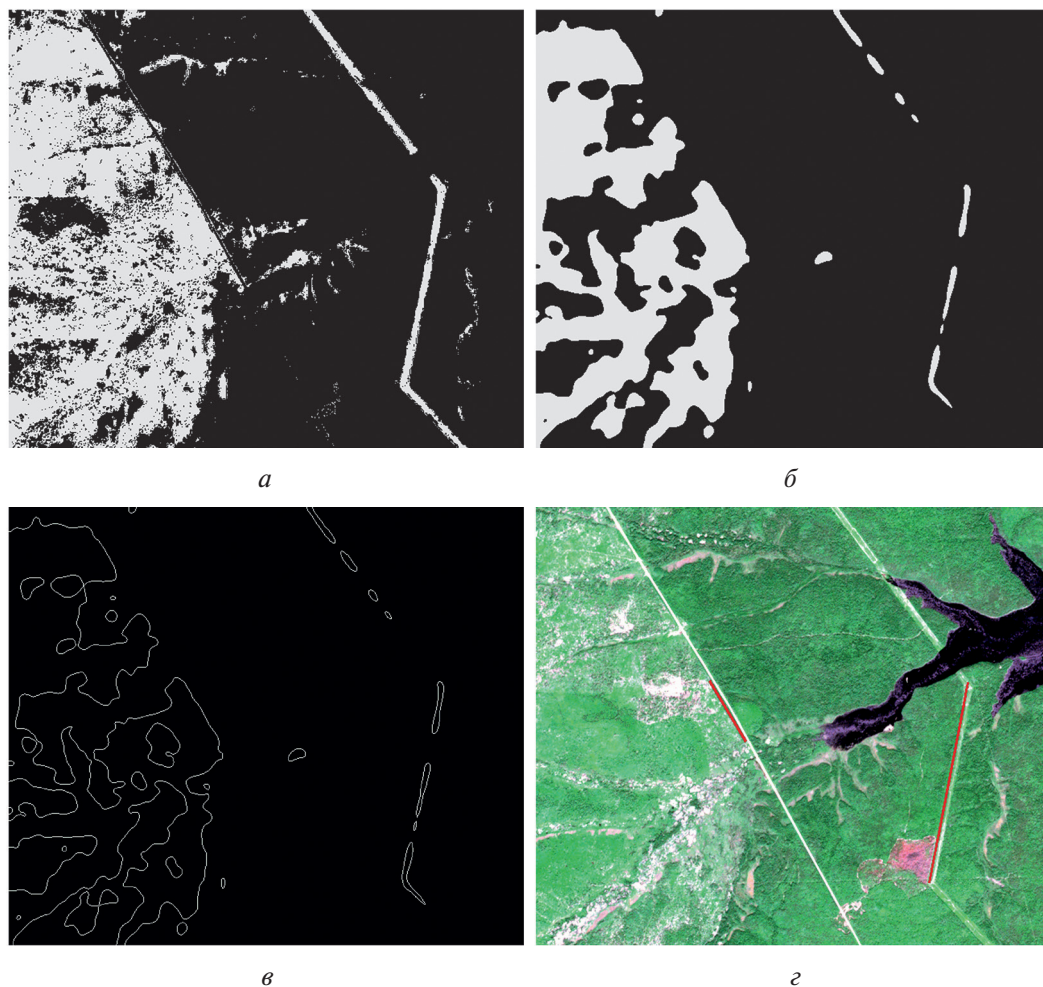


Рис. 9. Результат выделения лесных просек на участке № 3: *а* — маска лугов из карты классификации; *б* — маска лугов после гауссова размытия и бинаризации; *в* — результат разности между маской *б* и её двоичной эрозией; *г* — результат выделения лесных просек

На шести остальных участках проводилось апробирование алгоритма с целью тестирования разработанной методики на более разнообразных участках (рис. 10, см. с. 22). Так, на 4-м участке (см. рис. 10, 4) лесные просеки под ЛЭП отсутствуют, на 7-м участке (см. рис. 10, 7) лесная просека проходит через участки вырубок, границы которых аналогично лесной просеке линейны, а на 9-м участке (см. рис. 10, 9) просека пересекается автомобильными дорогами. Результаты выделения сопоставлены с данными о расположении линий электропередач от OpenStreetMap, ведь ЛЭП косвенно свидетельствуют о расположении лесных просек под ними.

В большинстве случаев алгоритм продемонстрировал относительно высокую эффективность в определении лесных просек под ЛЭП. На 4-м участке, где таковых объектов нет, алгоритм распознал «шумовые» линии на самом юге участка (см. рис. 10, 4), но их количество не критично. На 5, 7 и 8-м участках (см. рис. 10, 5, 7 и 8) алгоритмом успешно выделены лесные просеки практически на всей их протяжённости. Однако на 9-м участке алгоритм допустил серьёзные ошибки в идентификации лесных просек (см. рис. 10, 9). На изображении возникло значительное количество случайных линий, которые не соответствуют лесным просекам.

Результаты расчётов метрик приведены в табл. 2. Как говорилось в разделе про методологию алгоритма, метрика близости не была подсчитана для 4-го участка, так как на него отсутствовали данные о ЛЭП на момент подсчёта метрики. Аналогично метрика непрерывности не вычислена для 4, 7 и 9-го участков из-за отсутствия векторных данных о пилонах ЛЭП.

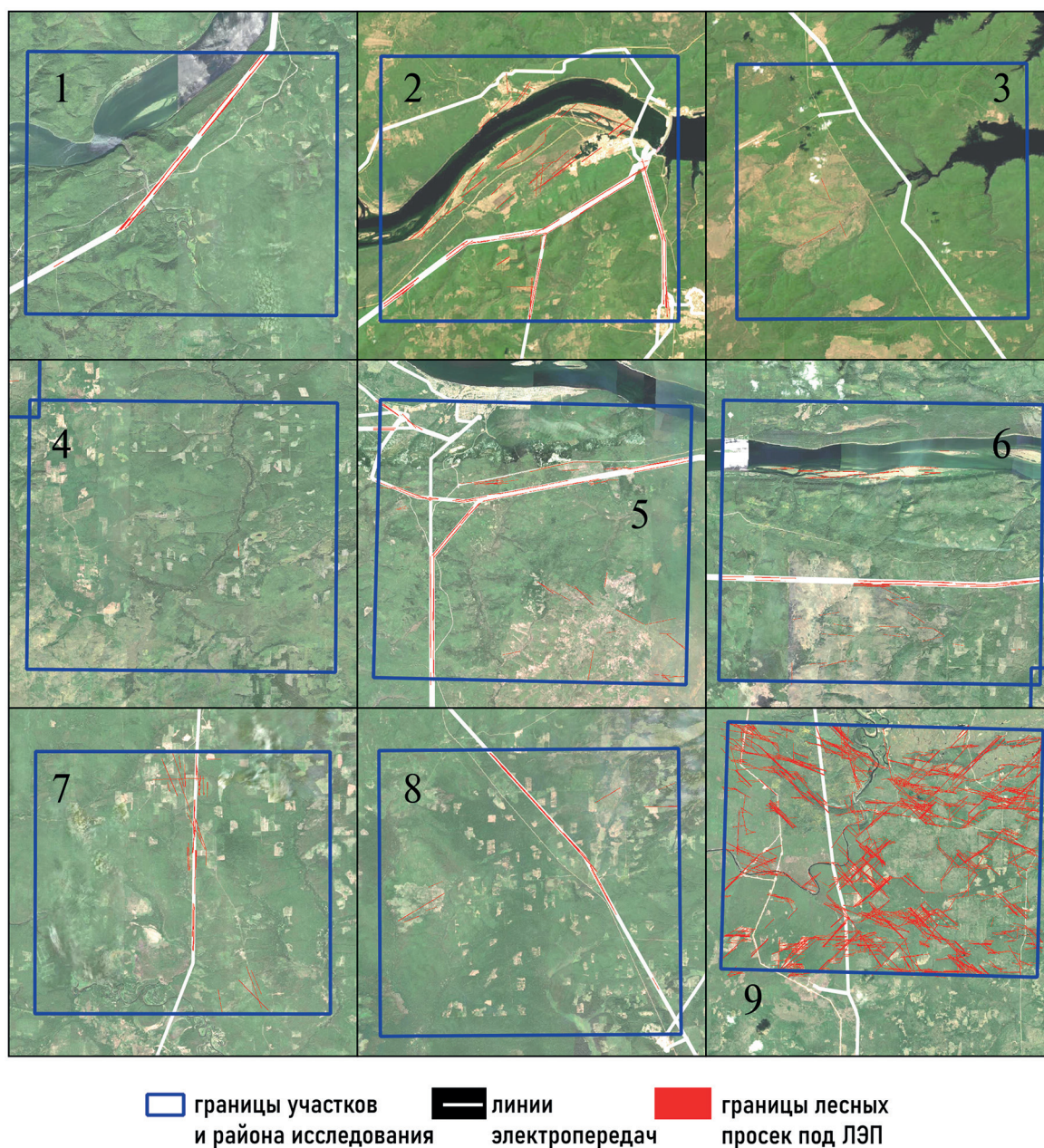


Рис. 10. Сравнение результатов выделенных лесных просек с данными о расположении ЛЭП от OpenStreetMap

Таблица 2. Результаты расчёта метрик близости и непрерывности

Номер тестового участка	Значение метрики близости	Качественная характеристика близости	Значение метрики непрерывности	Качественная характеристика непрерывности
1	0,354	Высокая	0,041	Высокая
2	0,033	Средняя	0,026	Средняя
3	0,002	Низкая	0,002	Низкая
5	0,167	Высокая	0,024	Средняя
6	0,015	Средняя	0,041	Высокая
7	0,031	Средняя	—	—
8	0,354	Высокая	0,021	Средняя
9	0,002	Низкая	—	—

Метрика близости достигла максимальных значений для 1, 5 и 8-го участков, где большинство выделенных границ лесных просек расположены практически на их действительных границах (табл. 2). В свою очередь, на 9-м участке значение метрики минимально, что вполне ожидаемо — довольно много линий проведено в отдалении от исходной лесной просеки. Аналогично низкое значение наблюдается на 2-м участке, где границы просек распознаны на территории города либо на свободных от растительности землях. Исходя из вышеперечисленного можно провести следующую качественную классификацию метрики близости:

- высокая степень близости (от 0,1);
- средняя степень близости (от 0,01 до 0,1);
- низкая степень близости (до 0,01).

Таким образом, на трёх участках из восьми метрика близости достигла высоких значений, на трёх из восьми — средних и двух из восьми — низких значений.

Метрика непрерывности показала максимальные значения на 1-м участке, где лесные просеки действительно проведены алгоритмом на всей её протяжённости. Довольно высокое значение метрики для 6-го участка, хотя на нём есть участки разрыва линий. Самое низкое значение метрики наблюдается у 3-го участка, возможно из-за длинного отрезка лесной просеки, на которой алгоритм не провёл линии. Для метрики непрерывности можно провести следующую классификацию:

- высокая степень непрерывности (от 0,04);
- средняя степень непрерывности (от 0,01 до 0,04);
- низкая степень непрерывности (до 0,01).

Таким образом, для 1-го участка алгоритм поиска лесных просек показал высокие значения двух метрик. Второй участок отличается средними, а 3-й — самыми низкими значениями метрик. При этом получились промежуточные варианты, в которых метрика близости показала более высокие результаты, чем метрика непрерывности (5-й, 8-й участки).

Созданный и реализованный алгоритм успешно выделяет лесные просеки практически на всех участках региона исследования. При этом вероятность переоценки числа лесных просек комплексным алгоритмом растёт при наличии больших по размерам участков без древесной растительности на изображении. Двоичное открытие и гауссовский фильтр не способны отфильтровать эти участки, из-за чего их границы подаются на вход вероятностному преобразованию Хафа. Тем не менее методика выделения лесных просек на основе логистической регрессии и преобразования Хафа оказалась работоспособной, и дальнейшее развитие алгоритма представляется перспективным.

Заключение

В рамках исследования выполнена задача разработки алгоритма распознавания лесных просек под ЛЭП в районе Богучанской ГЭС, основанного на логистической регрессии и вероятностном преобразовании Хафа, а также его апробации на различных участках. В отличие от алгоритмов, созданных в предыдущих исследованиях, предложенный метод обладает высокой интерпретируемостью и способен выделять просеки под ЛЭП на основе их дешифровочных признаков. Дополнительным достоинством данного алгоритма является его автоматизированный характер, что позволяет осуществлять выделение просек без необходимости настройки гиперпараметров. Однако текущая версия алгоритма протестирована на небольших по площади участках со сравнительно небольшим набором объектов, поэтому обеспечение устойчивости работы алгоритма на больших территориях остаётся задачей дальнейших исследований. Разрабатываемая автоматическая методика имеет перспективу использования в процедурах мониторинга состояния просек под ЛЭП и оценки рисков возникновения аварий, в том числе на основе выявления участков с интенсивным развитием древесной растительности, а также для формирования лётного задания беспилотным летательным аппаратам, производящим съёмку и контроль состояния ЛЭП.

В качестве продолжения исследования планируется апробация и внедрение нейросетевых алгоритмов для выделения просек и проведение сравнительного анализа и анализа устойчивости результатов, получаемых по двум различным технологиям. Помимо этого, планируется наполнение результатами детектирования просек под ЛЭП специализированной базы данных, позволяющей проводить их визуализацию и анализ интерфейсными средствами геоинформационных систем. На основе созданной методики выделения линейных объектов — просек под ЛЭП — возможно создание алгоритма для актуализации информации о сети лесных дорог. В целом разрабатываемая технология имеет перспективы применения для дистанционного мониторинга состояния объектов энергетической инфраструктуры (ЛЭП и просек под ними).

Репозиторий с кодом алгоритма и примером его применения доступен по ссылке https://github.com/yana-b27/clearings_extraction_algorithm.

Работа выполнена в рамках государственного задания ИКИ РАН (тема «Мониторинг», госрегистрация № 122042500031-8) и государственного задания ЦЭПЛ РАН им. А. С. Исаева по теме «Биоразнообразие и экосистемные функции лесов» (регистрационный номер НИОКТР 124013000750-1).

Литература

1. Бахрамхан Я. О., Ермаков Д. М., Подольская Е. С. Опыт разработки алгоритма выделения лесных просек под линиями электропередач в лесных ландшафтах на основе данных Sentinel-2 // Материалы 22-й Международ. конф. «Современные проблемы дистанционного зондирования Земли из космоса». М.: ИКИ РАН, 2024. С. 163. <https://doi.org/10.21046/22DZZconf-2024a>.
2. Ahmad J., Malik A. S., Xia L. Vegetation monitoring for high-voltage transmission line corridors using satellite stereo images // 2011 National Postgraduate Conf. 2011. 5 p. DOI: 10.1109/NatPC.2011.6136337.
3. Bernstein R., Di Gesù V. A combined analysis to extract objects in remote sensing images // Pattern Recognition Letters. 1999. V. 20. Iss. 11–13. P. 1407–1414. [https://doi.org/10.1016/S0167-8655\(99\)00126-9](https://doi.org/10.1016/S0167-8655(99)00126-9).
4. Chen S., Haralick R. M. Recursive erosion, dilation, opening, and closing transforms // IEEE Trans. Image Processing. 1995. V. 4. Iss. 3. P. 335–345. DOI: 10.1109/83.366481.
5. Duda R. O., Hart P. E. Use of the Hough transformation to detect lines and curves in pictures // Communications of the ACM. 1972. V. 15. Iss. 1. P. 11–15. <https://doi.org/10.1145/361237.361242>.
6. Gorelick N., Hancher M., Dixon M. et al. Google Earth Engine: Planetary-scale geospatial analysis for everyone // Remote Sensing of Environment. 2017. V. 202. P. 18–27. <https://doi.org/10.1016/j.rse.2017.06.031>.
7. Huete A. R. A soil-adjusted vegetation index (SAVI) // Remote Sensing of Environment. 1988. V. 25. Iss. 3. P. 295–309. [https://doi.org/10.1016/0034-4257\(88\)90106-X](https://doi.org/10.1016/0034-4257(88)90106-X).
8. Kobayashi Y., Karady G. G., Heydt G. T., Olsen R. G. The utilization of satellite images to identify trees endangering transmission lines // IEEE Trans. Power Delivery. 2009. V. 24. Iss. 3. P. 1703–1709. DOI: 10.1109/TPWRD.2009.2022664.
9. Lin Y., Hong W., Tan W. et al. Airborne circular SAR imaging: Results at P-band // 2012 IEEE Intern. Geoscience and Remote Sensing Symp. 2012. P. 5594–5597. DOI: 10.1109/IGARSS.2012.6352051.
10. Mahdi Elsiddig Haroun F., Mohamed Deros S. N., Din N. M. (2021a) Detection and monitoring of power line corridor from satellite imagery using RetinaNet and K-mean clustering // IEEE Access. 2021. V. 9. P. 116720–116730. DOI: 10.1109/ACCESS.2021.3106550.
11. Mahdi Elsiddig Haroun F., Mohamed Deros S. N., Bin Baharuddin M. Z., Din N. M. (2021b) Detection of vegetation encroachment in power transmission line corridor from satellite imagery using support vector machine: A features analysis approach // Energies. 2021. V. 14. Iss. 12. Article 3393. <https://doi.org/10.3390/en14123393>.
12. McFeeters S. K. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features // Intern. J. Remote Sensing. 1996. V. 17. Iss. 7. P. 1425–1432. <https://doi.org/10.1080/01431169608948714>.
13. Moeller M. S. Monitoring powerline corridors with stereo satellite imagery // Proc. MAPPS/ASPRS Fall Conf. 2006. 6 p.
14. Mu C., Yu J., Feng Y., Cai J. Power lines extraction from aerial images based on Gabor filter // Intern. Symp. Spatial Analysis, Spatial-Temporal Data Modeling, and Data Mining. 2009. V. 7492. P. 1081–1088.

15. Sarabandi K., Park M. Millimeter-wave radar phenomenology of power lines and a polarimetric detection algorithm // IEEE Trans. Antennas and Propagation. 1999. V. 47. Iss. 12. P. 1807–1813. <https://doi.org/10.1109/8.817656>.
16. Sarabandi K., Pierce L., Oh Y., Ulaby F. T. Power lines: Radar measurements and detection algorithm for polarimetric SAR images // IEEE Trans. Aerospace and Electronic Systems. 1994. V. 30. Iss. 2. P. 632–643. <https://doi.org/10.1109/7.272288>.
17. Schwarz G., Soccorsi M., Chaabouni-Chouayakh H. et al. Automated information extraction from high resolution SAR images: TerraSAR-X interpretation applications // 2009 IEEE Intern. Geoscience and Remote Sensing Symp. 2009. V. 4. P. 677–680. DOI: 10.1109/IGARSS.2009.5417467.
18. Tong W.-G., Li B.-S., Yuan J.-S., Zhao S.-T. Transmission line extraction and recognition from natural complex background // 2009 Intern. Conf. Machine Learning and Cybernetics. 2009. V. 4. P. 2473–2477. DOI: 10.1109/ICMLC.2009.5212202.
19. Xu H. A new index for delineating built-up land features in satellite imagery // Intern. J. Remote Sensing. 2008. V. 29. Iss. 14. P. 4269–4276. <https://doi.org/10.1080/01431160802039957>.

Automatic detection of power line clearings in Sentinel-2 images using machine learning and computer vision methods

Ya. O. Bakhramkhan¹, D. M. Ermakov^{2,3}, E. S. Podolskaia^{1,4}

¹ National Research University Higher School of Economics, Moscow 109028, Russia
E-mails: ybahramkhan@gmail.com, podols_kate@mail.ru

² Space Research Institute RAS, Moscow 117997, Russia
E-mail: d.m.ermakov@cosmos.ru

³ Kotelnikov Institute of Radioengineering and Electronics RAS, Fryazino Branch
Fryazino, Moscow Region 141190, Russia

⁴ Isaev Centre for Forest Ecology and Productivity RAS, Moscow 117997, Russia

Regular monitoring of the condition of power lines is imperative for ensuring uninterrupted power supply to settlements and infrastructure facilities. However, in forest areas with sparse population and limited road infrastructure, continuous monitoring is complicated. To address this challenge, the study proposes an algorithm that uses Sentinel-2 satellite images to identify forest clearings in remote, hard-to-reach forest areas. This algorithm marks a pioneering step in the automation of remote monitoring of power line clearings, particularly in the context of detecting changes (e.g., laying, overgrowing, clearing of power line clearings, etc.). The algorithm's logic is straightforward and based on the search for objects of interest in the space of decoding features. The logistic regression model applies spectral characteristics to identify areas potentially associated with clearings. Subsequently, the probabilistic Hough transform locates linear structures within the binary mask of these areas. The developed algorithm is fully automatic, eliminating the need for individual parameter tuning for each image. We developed a set of metrics to evaluate quality of clearings extraction as a measure of algorithm's efficiency. Performance of the algorithm is locally stable, it correctly identifies the areas of power line clearings in all satellite images of the study region that have been used.

Keywords: clearings, power lines, remote monitoring, interpretation features, interpretability of the algorithm, Hough transform, logistic regression

Accepted: 13.05.2025

DOI: 10.21046/2070-7401-2025-22-4-11-26

References

1. Bakhramkhan Ya. O., Ermakov D. M., Podolskaia E. S., Experience in developing an algorithm for identifying forest clearings under power lines in forest landscapes based on Sentinel-2 data, *Materialy 22-i Mezhdunarodnoi konferentsii "Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa"*

- (Proc. 22nd Intern. Conf. “Current Problems in Remote Sensing of the Earth from Space”), Moscow: IKI RAS, 2024, p. 163 (in Russian), <https://doi.org/10.21046/22DZZconf-2024a>.
2. Ahmad J., Malik A. S., Xia L., Vegetation monitoring for high-voltage transmission line corridors using satellite stereo images, *2011 National Postgraduate Conf.*, 2011, 5 p., DOI: 10.1109/NatPC.2011.6136337.
 3. Bernstein R., Di Gesù V., A combined analysis to extract objects in remote sensing images, *Pattern Recognition Letters*, 1999, V. 20, Iss. 11–13, pp. 1407–1414, [https://doi.org/10.1016/S0167-8655\(99\)00126-9](https://doi.org/10.1016/S0167-8655(99)00126-9).
 4. Chen S., Haralick R. M., Recursive erosion, dilation, opening, and closing transforms, *IEEE Trans. Image Processing*, 1995, V. 4, Iss. 3, pp. 335–345, DOI: 10.1109/83.366481.
 5. Duda R. O., Hart P. E., Use of the Hough transformation to detect lines and curves in pictures, *Communications of the ACM*, 1972, V. 15, Iss. 1, pp. 11–15, <https://doi.org/10.1145/361237.361242>.
 6. Gorelick N., Hancher M., Dixon M. et al., Google Earth Engine: Planetary-scale geospatial analysis for everyone, *Remote Sensing of Environment*, 2017, V. 202, pp. 18–27, <https://doi.org/10.1016/j.rse.2017.06.031>.
 7. Huete A. R., A soil-adjusted vegetation index (SAVI), *Remote Sensing of Environment*, 1988, V. 25, Iss. 3, pp. 295–309, [https://doi.org/10.1016/0034-4257\(88\)90106-X](https://doi.org/10.1016/0034-4257(88)90106-X).
 8. Kobayashi Y., Karady G. G., Heydt G. T., Olsen R. G., The utilization of satellite images to identify trees endangering transmission lines, *IEEE Trans. Power Delivery*, 2009, V. 24, Iss. 3, pp. 1703–1709, DOI: 10.1109/TPWRD.2009.2022664.
 9. Lin Y., Hong W., Tan W. et al., Airborne circular SAR imaging: Results at P-band, *2012 IEEE Intern. Geoscience and Remote Sensing Symp.*, 2012, pp. 5594–5597, DOI: 10.1109/IGARSS.2012.6352051.
 10. Mahdi Elsiddig Haroun F., Mohamed Deros S. N., Din N. M. (2021a), Detection and monitoring of power line corridor from satellite imagery using RetinaNet and K-mean clustering, *IEEE Access*, 2021, V. 9, pp. 116720–116730, DOI: 10.1109/ACCESS.2021.3106550.
 11. Mahdi Elsiddig Haroun F., Mohamed Deros S. N., Bin Baharuddin M. Z., Din N. M. (2021b), Detection of vegetation encroachment in power transmission line corridor from satellite imagery using support vector machine: A features analysis approach, *Energies*, 2021, V. 14, Iss. 12, Article 3393, <https://doi.org/10.3390/en14123393>.
 12. McFeeters S. K., The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features, *Intern. J. Remote Sensing*, 1996, V. 17, Iss. 7, pp. 1425–1432, <https://doi.org/10.1080/01431169608948714>.
 13. Moeller M. S., Monitoring powerline corridors with stereo satellite imagery, *Proc. MAPPS/ASPRS Fall Conf.*, 2006, 6 p.
 14. Mu C., Yu J., Feng Y., Cai J., Power lines extraction from aerial images based on Gabor filter, *Intern. Symp. Spatial Analysis, Spatial-Temporal Data Modeling, and Data Mining*, 2009, V. 7492, pp. 1081–1088.
 15. Sarabandi K., Park M., Millimeter-wave radar phenomenology of power lines and a polarimetric detection algorithm, *IEEE Trans. Antennas and Propagation*, 1999, V. 47, Iss. 12, pp. 1807–1813, <https://doi.org/10.1109/8.817656>.
 16. Sarabandi K., Pierce L., Oh Y., Ulaby F. T., Power lines: Radar measurements and detection algorithm for polarimetric SAR images, *IEEE Trans. Aerospace and Electronic Systems*, 1994, V. 30, Iss. 2, pp. 632–643, <https://doi.org/10.1109/7.272288>.
 17. Schwarz G., Soccorsi M., Chaabouni-Chouayakh H. et al., Automated information extraction from high resolution SAR images: TerraSAR-X interpretation applications, *2009 IEEE Intern. Geoscience and Remote Sensing Symp.*, 2009, V. 4, pp. 677–680, DOI: 10.1109/IGARSS.2009.5417467.
 18. Tong W.-G., Li B.-S., Yuan J.-S., Zhao S.-T., Transmission line extraction and recognition from natural complex background, *2009 Intern. Conf. Machine Learning and Cybernetics*, 2009, V. 4, pp. 2473–2477, DOI: 10.1109/ICMLC.2009.5212202.
 19. Xu H., A new index for delineating built-up land features in satellite imagery, *Intern. J. Remote Sensing*, 2008, V. 29, Iss. 14, P. 4269–4276, <https://doi.org/10.1080/01431160802039957>.