

АНАЛИЗ И СИНТЕЗ СИГНАЛОВ И ИЗОБРАЖЕНИЙ

УДК 528.854

КОМПЛЕКСИРОВАНИЕ КЛАССИФИКАТОРОВ
В ЗАДАЧЕ ТЕМАТИЧЕСКОЙ ОБРАБОТКИ
ГИПЕРСПЕКТРАЛЬНЫХ АЭРОКОСМИЧЕСКИХ
ИЗОБРАЖЕНИЙ*Е. В. Дмитриев¹, В. В. Козодеров², А. О. Дементьев¹, А. Н. Сафонова³

¹*Институт вычислительной математики РАН,
119333, Москва, ул. Губкина, 8*

²*Московский государственный университет им. М. В. Ломоносова,
119991, Москва, Ленинские горы, 1*

³*Институт космических и информационных технологий
Сибирского федерального университета,
660074, г. Красноярск, ул. Академика Киренского, 26, корп. 1
E-mail: yegor@mail.ru*

Рассмотрена задача тематической обработки гиперспектральных аэрокосмических изображений. Предлагаемая методика, основанная на алгоритме декодирования, позволяет построить эффективный алгоритм многоклассовой классификации на основе комбинации стандартных алгоритмов бинарной классификации различной сложности. Приведено обоснование выбора конфигурации алгоритма классификации. С помощью тестовых расчётов на основе модельных и реальных данных показана эффективность использования предлагаемого подхода для распознавания объектов по гиперспектральным изображениям.

Ключевые слова: дистанционное зондирование, гиперспектральные изображения, распознавание образов, тематическая обработка, классификация.

DOI: 10.15372/AUT20180301

Введение. В последние годы широко развивается новая технология дистанционного зондирования, сочетающая в себе возможности получения информации об исследуемых объектах на основе пространственного анализа цифровых изображений и методов спектроскопии. Соответствующие подходы были названы гиперспектральными технологиями (ГСТ), включающими в себя гиперспектральные авиационные и космические измерения в различных спектральных диапазонах, а также математические методы обработки и анализа получаемых таким образом данных (гиперспектральных изображений) [1].

Основное преимущество гиперспектральных технологий состоит в том, что используется наиболее подробная информация о спектральных характеристиках исследуемого объекта в пределах заданной точности. При этом отпадает необходимость в дополнительных предположениях относительно свойств объекта, введение которых зачастую вносит значительные неопределённости в решение существующих задач. Высокое спектральное разрешение гиперспектральных приборов позволяет решать задачи классификации и определения характеристик исследуемых объектов на новом качественном уровне. Вместо традиционных вегетационных индексов всё чаще используются узкополосные гиперспектральные индексы, более высокая эффективность которых отмечается различными авторами [2].

*Работа выполнена при финансовой поддержке Российского научного фонда (проект № 16-11-00007) и Российского фонда фундаментальных исследований (проект № 16-01-00107).

Так, в частности, применение гиперспектральных измерений даёт возможность повысить качество оценки состояния растительного покрова за счёт более точного определения положения границ поглощения пигментов [3]. В работе [4] показаны преимущества гиперспектральных вегетационных индексов спутника EO-1 Hyperion для оценки биомассы посевов перед аналогичными многоспектральными индексами, полученными со спутников IKONOS, GeoEye-1, WorldView-2, Landsat ETM+ и Terra MODIS.

При обработке гиперспектральных измерений возникает ряд известных технических сложностей, связанных с накоплением больших объёмов информации и коррелированностью данных. Качество тематической обработки гиперспектральных изображений зависит от многих факторов: прозрачности атмосферы и наличия облачности, характеристик измерительной аппаратуры, качества предварительной обработки, точности геопривязки дистанционных и наземных данных. При достаточно высоком качестве измерений и коррекции появляется необходимость построения эффективного алгоритма классификации, который обеспечивал бы высокую точность и скорость вычислений [5–7].

Существуют различные методы, которые могут быть использованы для распознавания наземных объектов по гиперспектральным аэрокосмическим изображениям. В последние годы опубликован ряд работ, посвящённых сравнительному анализу различных методов классификации. Так, например, в [5] представлены результаты распознавания видового состава тропических лесов на основе гиперспектральных изображений со спутника EO-1 Hyperion. Для обработки применялись классификаторы на базе метода спектрального угла, искусственных нейронных сетей и метода опорных векторов (SVM). Обнаружено, что последние два метода дали наилучшие результаты для восьми рассматриваемых классов растительности. Также было показано, что метод опорных векторов оказался существенно менее подвержен негативному влиянию избыточности спектральной информации.

Оценка зависимости качества распознавания лесной растительности по гиперспектральным данным от спектрального разрешения и сложности классификатора получена в работе [6]. В качестве тестовой информации использовались данные 126-канального гиперспектрометра AISA Eagle. Спектральное разрешение менялось с 9,2 до 36,8 нм. Сравнивались линейный и квадратичный дискриминантные анализы, а также метод SVM. В [8] метод опорных векторов сопоставлялся с методом случайных лесов. В качестве тестовой территории был выбран участок, расположенный на севере г. Карлсруэ в федеральном штате Баден-Вюртемберг (Германия) и содержащий древостой, характерные для Центральной Европы. В [8] было отмечено, что оба указанных классификатора можно использовать для распознавания породного состава, однако метод случайных лесов имеет некоторые преимущества.

В данной работе рассматриваются возможности построения комплексных многоклассовых классификаторов, созданных на основе комбинирования базовых классификаторов различной сложности: метрические и байесовские классификаторы, метод SVM. Для этого используются методы взвешенного голосования и некоторые результаты теории кодирования. Для проведения численных экспериментов применяются модельные данные и гиперспектральные изображения тестовых участков с наличием наземной информации.

Методика классификации. В качестве основного метода распознавания используется метод декодирования [9]. Метод позволяет решать задачу многоклассовой классификации на базе проведения серии бинарных классификаций и последующего голосования и может быть сформулирован следующим образом.

Пусть $x \in R^M$ — признаки классов, а $y \in R^1$ — наименования классов. Введём матрицу $C = (c_{ij})$ размера $K \times L$ (далее кодовая матрица), столбцы которой соответствуют различным способам обучения некоторого бинарного классификатора (или, лучше сказать, некоторым задачам бинарной классификации, построенным на основе имеющихся классов), а строки — распознаваемым классам. Элементы матрицы принимают следую-

	l_1	l_2	l_3	l_4	l_5	l_6	l_7	HD	ED
y_1	-1	-1	1	1	-1	1	1	1	2
y_2	-1	1	0	0	0	0	0	5/2	$\sqrt{5}$
y_3	1	-1	-1	-1	1	1	-1	6	$\sqrt{24}$
y_4	0	-1	1	0	1	-1	1	4	$\sqrt{14}$
z	-1	1	1	1	-1	1	1		

Рис. 1. Пример голосования в алгоритме декодирования. Два столбца справа обозначают расстояния между кодом z классифицируемого вектора признаков и кодами классов. Жирным шрифтом (1, 2) выделены минимальные расстояния в соответствии с метриками HD и ED

щие значения: 1 — объект, -1 — альтернатива, 0 — прочие объекты, не участвующие в классификации. Таким образом, каждому классу ставится в соответствие некоторый код (строка матрицы C). В связи с этим при построении матрицы C имеют место ограничения:

- допустимые значения элементов — только числа -1, 0 и 1;
- каждый столбец содержит, как минимум, один элемент 1 и один элемент -1;
- все различные векторы-столбцы не лежат на одной прямой;
- все различные векторы-строки не совпадают;
- для любых двух классов должен существовать столбец, в котором эти классы разделяются.

Пусть имеется некоторое обучающее множество из N пар элементов x_i и y_i ($i = 1, \dots, N$) и ставится задача классифицировать новое измерение x . Для наглядности рассмотрим пример, представленный на рис. 1. В данном примере имеется четыре класса и задана кодовая матрица, определяющая $L = 7$ задач бинарной классификации l_1, \dots, l_7 . Пусть классификаторы l_1, \dots, l_7 обучены на некотором заданном множестве. Последовательно применяя их к измерению x , получим кодовый вектор z . Сравнивая z с кодами классов y_1, \dots, y_4 (соответствующие строки кодовой матрицы) с использованием выбранной метрики, например расстояния Хэмминга $HD(z, y_j) = \sum_{i=1}^L \frac{1}{2}[1 - \text{sign}(z^i y_j^i)]$ или рассто-

яния Евклида $ED(z, y_j) = \sqrt{\sum_{i=1}^L (z^i - y_j^i)^2}$, выбираем класс y_i с кодом, наиболее близким к коду z , в соответствии с выбранной метрикой. Данный класс и является результатом классификации.

Результат алгоритма декодирования будет зависеть от выбора кодовой матрицы. Стандартные кодовые матрицы представлены на рис. 2. Матрицы подразделяются по типу кодирования (бинарное и тернарное) и обеспечивают различную точность решения задач классификации при разных объёмах вычислений. Необходимо подчеркнуть, что в случае тернарного кодирования используемые базовые классификаторы должны обеспечить возможность отказа от классификации (значение 0 на выходе). Точность классификации напрямую связана с различием кодов классов. Матрица, имеющая наибольшее различие кодов, гарантирует лучшую точность.

Одним из перспективных подходов к построению оптимальных кодовых матриц является метод случайных матриц, который состоит в следующем. Задаётся число L (обычно пропорционально логарифму числа классов) и случайным образом генерируется большое число кодовых матриц в соответствии с указанными выше правилами. Из полученных матриц выбирается такая, которая имеет максимально большое минимальное попарное

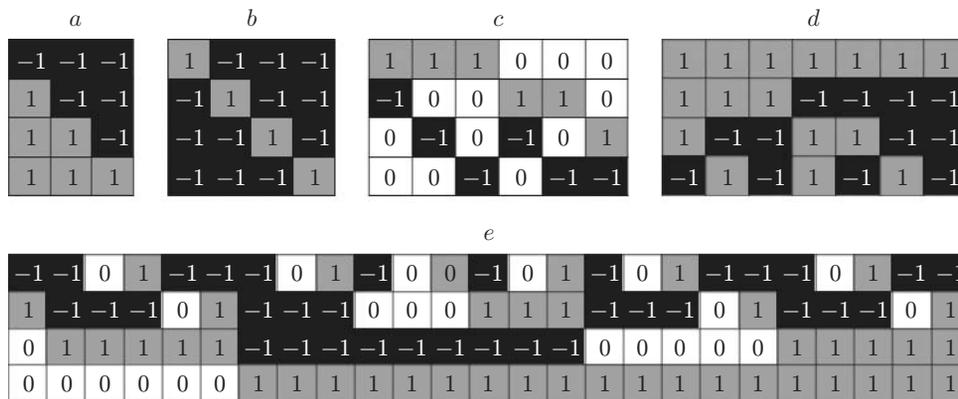


Рис. 2. Стандартные кодовые матрицы: *a* — порядковое кодирование (ПК), *b* — один против всех (ОПВ), *c* — каждый против каждого (КПК), *d* — полная бинарная (ПБ), *e* — полная тернарная (ПТ)

расстояние между строками в смысле метрики Хэмминга, т. е.

$$D(k_1, k_2) = \frac{1}{2} \sum_{j=1}^L |c_{k_1j} - c_{k_2j}|.$$

Приведём некоторые рекомендации по выбору кодовых матриц для решения различных задач. Минимальные попарные расстояния между кодами стандартных кодовых матриц представлены в табл. 1. Порядковое кодирование состоит из верхнетреугольной и нижнетреугольной матриц, соответствующих объектам 1 и -1 . Длина кодов определяется как $L = K - 1$. Данная матрица обеспечивает самую высокую скорость обучения и классификации, но она имеет минимальное расстояние между кодами соседних классов. Матрица ОПВ имеет почти одинаковую с ПК длину кодов $L = K$, однако разница между кодами в 2 раза больше. Эти две матрицы предпочтительны при большом числе классов ($K > 100$).

Минимальное попарное расстояние между кодами для матрицы КПК увеличивается с ростом K и, начиная с $K = 4$, становится значительно больше, чем в случае матрицы ОПВ. Однако увеличение длины кода происходит по квадратичному закону, что делает затруднительным решение задач классификации с большим количеством классов. Например, уже при $K = 100$ мы должны решить 4950 бинарных задач классификации всего лишь для одной реализации вектора признаков. Таким образом, эта матрица предпочтительнее в случае среднего числа классов $10 < K < 100$.

Матрицы ПБ и ПТ обеспечивают хорошо различимые, но при этом достаточно длинные коды распознаваемых классов. Минимальное попарное расстояние между кодами

Таблица 1

Минимальные попарные расстояния между кодами для различных кодовых матриц

Кодовые матрицы	Число классов						
	2	3	4	5	6	7	8
ПК	1,00	1,00	1,00	1,00	1,00	1,00	1,00
ОПВ	1,00	2,00	2,00	2,00	2,00	2,00	2,00
КПК	1,00	2,00	3,50	5,50	8,00	11,00	14,50
ПБ	1,00	2,00	4,00	8,00	16,00	32,00	64,00
ПТ	1,00	4,00	14,50	49,00	158,50	499,00	1544,50

ми быстро растёт по сравнению с другими схемами кодирования. Однако длины кодов указанных матриц, которые могут быть выражены формулами $L = 2^{K-1} - 1$ и $L = (3^K - 2^{K+1} + 1)/2$ соответственно, также увеличиваются и быстро становятся настолько большими, что делает классификацию невозможной в пределах разумного времени. В частности, для $K = 15$ длины кодов для матриц ПБ и ПТ равны 32767 и 7141686 соответственно. Таким образом, при решении практических задач нам нужно найти компромисс между временем вычислений и точностью используемого классификатора.

Точность и скорость работы метода декодирования также во многом определяется эффективностью базовых алгоритмов бинарной классификации. Задача сравнения методов классификации почвенно-растительного покрова по данным гиперспектрального аэрокосмического зондирования рассматривалась, например, в [2]. Эффективность методов распознавания трудноразличимых типов растительности с учётом предварительной пространственной обработки и постобработки картосхем попиксельной спектральной классификации исследовалась в [10]. Эффективность совместного использования спектральных и текстурных признаков в задаче определения породного состава смешанных древостоев по авиационным гиперспектральным изображениям высокого пространственного разрешения продемонстрирована в [11].

В рамках данного исследования больший интерес представляет анализ точности и скорости работы базовых классификаторов и их модификаций. Нелинейные классификаторы, как правило, имеют преимущество по точности за счёт использования более сложных дискриминантных поверхностей. При этом они проигрывают в скорости и устойчивости результатов обучения. Квадратичный байесовский классификатор (БК) имеет более низкую скорость расчётов по сравнению с методом ближайших эталонов, которая падает с увеличением размерности признакового пространства. Различие в скорости классификации может также зависеть от числа реализаций в обучающем ансамбле. В частности, это имеет место в методе SVM с мягким зазором из-за увеличения числа опорных векторов.

В работе [12] было проведено сравнение эффективности различных модификаций метрических и байесовских классификаторов и метода SVM в задаче классификации смешанных древостоев. Наиболее быстрым оказался метод ближайших спектральных эталонов. Нормальный БК показал лучшую точность при достаточно высокой скорости расчётов. Байесовский классификатор на основе гауссовских смесей (ГС) и метод опорных векторов с гауссовским ядром позволили получить решение наиболее сложных задач классификации видового и возрастного составов смешанных лиственных древостоев с наилучшей точностью в отличие от других базовых классификаторов. Кроме этого, следует отметить, что при добавлении новых реализаций в обучающую выборку соотношение между оценками точности базовых алгоритмов может меняться.

В исходной обучающей базе данных, как правило, присутствуют объекты, имеющие разную степень различимости. Таким образом, при решении задач бинарной классификации, определяемых кодовой матрицей, целесообразно использовать классификаторы различной сложности. Пример комбинирования базовых классификаторов представлен на рис. 3.

Составные классы — водные объекты, почва и растительность — имеют хорошую делимость. Для решения большей части соответствующих задач классификации достаточно использовать слабые, но более быстрые классификаторы, например метод ближайшего эталона. Если точность таких классификаторов немного ниже требуемой, их можно улучшить за счёт применения бустинга [13]. Для разделения луговой растительности, хвойных и лиственных древостоев, как правило, достаточно точности нормального БК или SVM с квадратичным ядром. Для разделения лиственных таксонов, например берёзы и ольхи, необходимо использовать более сложные нелинейные классификаторы, такие как SVM с гауссовским ядром, метод парзенковского окна или ГС. При этом необходимо учитывать,

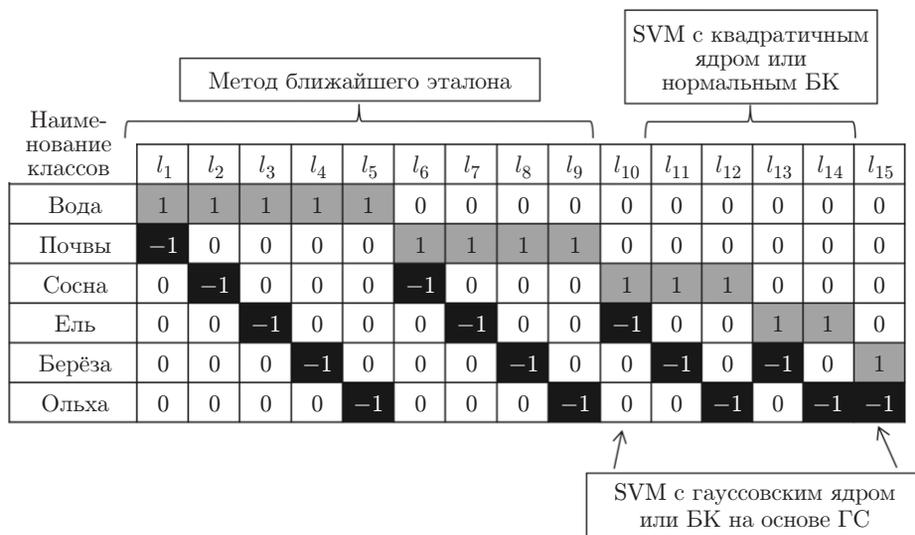


Рис. 3. Комбинирование базовых классификаторов на основе метода декодирования

что для обучения такого рода классификаторов в основном требуется достаточно большой объём обучающих данных для каждого класса.

Численные эксперименты. Для отработки предлагаемой методики использовались авиационные гиперспектральные изображения территории Савватьевского лесничества (Тверская область). Гиперспектрометр АВ-ВД разработки НПО «Лептон» (Москва, г. Зеленоград) позволяет получать изображения в 287 спектральных каналах видимого и ближнего инфракрасного диапазонов. Для выравнивания соотношения сигнал/шум было проведено сведение спектральных каналов к единому разрешению 5 нм. Таким образом, для обработки использовалось 87 объединённых каналов в указанном диапазоне. Пространственное разрешение составило 1,1 м при высоте полёта 2 км. Подробная информация об измерительной кампании и характеристики гиперспектрометра АВ-ВД представлены в работе [14].

Наземная информация, используемая для обучения алгоритма и валидации результатов классификации, представлена табличными и картографическими данными. Лесные насаждения разбиты на крупные участки — кварталы, которые подразделяются на более мелкие — выделы. Для каждого выдела имеются табличные лесотаксационные описания, включающие в себя породный состав и возраст древостоя. Для построения обучающего ансамбля в первую очередь брались однородные древостои с сомкнутым пологом. Однако такие участки можно найти не для всех имеющихся пород и возрастов. Для расширения обучающего набора данных мы использовали также выделы со смешанным породным и возрастным составом. При этом доминантная порода должна занимать не менее 80 % площади выдела и существенно отличаться от субдоминантной породы по своим спектральным характеристикам. Для обеспечения возможности применения нелинейных классификаторов полученная обучающая база содержит не менее 500 реализаций признаков для каждого класса. Различные породы деревьев представлены выборками, содержащими более 1000 реализаций.

Для обоснования выбора кодовой матрицы и ядра для метода SVM мы провели несколько тестовых экспериментов по классификации модельных данных вышеописанным методом. Рассматривались три класса с признаками, полученными на основе модели ГС с использованием реальных спектров растительности, и фоновый класс на основе равномерного распределения. Преимущество применения модельных данных для тестовых экспериментов заключается в том, что в этом случае мы знаем точные функции распределения

Таблица 2

Полная вероятность ошибки классификации модельных классов для различных ядер и кодовых матриц					
Ядро	Кодовые матрицы				
	ПК	ОПВ	КПК	ПБ	ПТ
Линейное	0,75	0,57	0,53	0,54	0,56
Квадратичное	0,48	0,46	0,38	0,41	0,38
Кубическое	0,45	0,43	0,32	0,38	0,36
Гауссовское	0,34	0,3	0,3	0,3	0,3

вероятности признаков рассматриваемых классов. Эти функции могут быть использованы для построения оптимального байесовского классификатора, который можно рассматривать в качестве эталона при тестировании других классификаторов, поскольку он обеспечивает минимальное значение полной вероятности ошибки. В нашем случае ошибка БК составила 0,3.

Результаты оценки полной вероятности ошибки классификации на основе метода кроссвалидации (скользящего контроля) представлены в табл. 2. Как видно, применение линейного ядра приводит к наибольшим ошибкам классификации, которые не превышают ошибки случайного выбора для всех рассмотренных схем кодирования. Ситуация улучшается при использовании квадратичного и кубического ядер. Можно отметить, что кодовые матрицы КПК, ПБ и ПТ обеспечивают значительно более точные результаты. Выбор гауссовского ядра приводит к минимальным ошибкам для всех рассмотренных схем кодирования. В этом случае применение любой матрицы кодирования, представленной в табл. 2, за исключением ПК, позволяет достичь минимально возможного значения ошибки классификации.

Решение рассматриваемой задачи предусматривает наличие порядка 60 классов. Таким образом, нами была выбрана кодовая матрица КПК (требует порядка 1770 задач бинарной классификации). В качестве наиболее точного бинарного классификатора, используемого для разделения схожих классов, выбран метод SVM с гауссовским ядром.

Характерной проблемой, возникающей при тематической обработке гиперспектральных изображений вследствие избыточности спектральной информации, является «проклятие размерности». Решение данной проблемы состоит в выделении наиболее информативных спектральных диапазонов. В [15] предложена методика отбора наиболее информативных признаков, устойчивая к малым изменениям обучающего множества. В результате применения этого метода для данных гиперспектрометра АВ-ВД и метода классификации были выбраны каналы с центральными длинами волн 523,6, 642,8, 709,7, 719,1, 790,4 и 834,9 нм.

В рамках предлагаемой работы мы ограничимся анализом ошибок распознавания основных типов объектов на территории тестового участка. Для этого после определения всего исходного множества классов была проведена их перегруппировка в составные классы: водная поверхность, открытые почвы, искусственные объекты, луговая растительность, берёзовые, осиновые и сосновые древостой. Характеристики точности тематической обработки данных прибором АВ-ВД с помощью предлагаемой методики сведены в табл. 3 и 4.

Матрица ошибок классификации, представленная в табл. 3, получена с использованием метода кроссвалидации по обучающим данным. Табл. 4 содержит априорные и независимые оценки вероятностей ошибочной классификации для указанных объектов и полную вероятность ошибки. При получении априорной оценки (метод переклассификации — МП) обучающие и тестовые ансамбли совпадают. Формирование независимых оценок основано

Таблица 3

**Матрица ошибок классификации гиперспектральных изображений АВ-ВД
для основных типов объектов тестового участка**

Предсказанный классификатор	Истинные объекты						
	Вода	Искусственные объекты	Почвы	Сосна	Берёза	Осина	Травы
Вода	4117	0	0	0	0	0	0
Искусственные объекты	0	833	4	0	0	0	0
Почвы	0	4	3335	0	0	0	0
Сосна	0	0	0	17675	568	4	267
Берёза	0	0	0	479	3583	182	39
Осина	0	0	0	1	224	985	0
Травы	0	0	0	246	36	0	2942

на методе случайных контрольных выборок (МСКВ), т. е. тестовые данные не участвуют в обучении. Хорошее соответствие априорной и независимой оценок свидетельствует об отсутствии проблем переобучения и «проклятия размерности».

Наилучшим образом удаётся классифицировать водные объекты (точная классификация). Вероятность ошибки классификации открытых почв и искусственных объектов составляет менее 0,5 %. Отдельные пиксели, соответствующие почвам, могут быть ошибочно классифицированы как искусственные объекты и наоборот. По данным табл. 3 можно видеть хорошую различимость хвойных древостоев, мелколиственных древостоев и луговой растительности. Вероятность ошибки распознавания сосновых древостоев и луговой растительности не превышает 10 %. Луговая растительность может быть неверно классифицирована как сосновый или берёзовый древостой и наоборот. Это связано с тем, что обучающая база данных охватывает весь диапазон возрастных классов берёзы и сосны от молодняка до перестойных древостоев. Наибольшие ошибки возникают при распозна-

Таблица 4

**Априорная и независимая оценки ошибки распознавания
основных типов объектов тестового участка**

Наименование объектов	Ошибки	
	Априорная оценка (МП)	Независимая оценка (МСКВ)
Вода	0	0
Искусственные объекты	0	0,0048
Почвы	0	0,0012
Сосна	0,0330	0,0453
Берёза	0,0908	0,1634
Осина	0,0927	0,1860
Травы	0,0267	0,0875
Общая ошибка	0,0347	0,0697

вании осинового древостоя. Полная вероятность ошибочной классификации составляет порядка 7 %.

Дополнительно к оценкам точности предлагаемого комплексного метода классификации были получены характеристики точности и скорости классификации с помощью базовых алгоритмов. Приведём характерные значения для тестовых экспериментов с использованием метода ближайшего эталона, нормального байесовского классификатора и метода SVM с гауссовским ядром для классификации берёзовых и сосновых древостоев одного возрастного класса. Скорость классификации для метода ближайшего эталона в данном случае оказывается в 4–6 раз выше, чем для нормального БК, и в 10–20 раз выше, чем для метода SVM с гауссовским ядром. При этом вероятность ошибочной классификации зачастую оказывается выше более чем на порядок. При классификации хорошо различимых объектов точности указанных методов совпадают, однако скорость работы метода ближайшего эталона оказывается в 3–5 раз выше.

Результаты сопоставления данных тематической обработки и наземных лесотаксационных данных показаны на рис. 4. На рис. 4, *a* приведены границы 13 выделов и соответствующие данные по породному составу. Значения среднеквадратичной ошибки восстановления породного состава для указанных выделов с учётом градаций освещённости

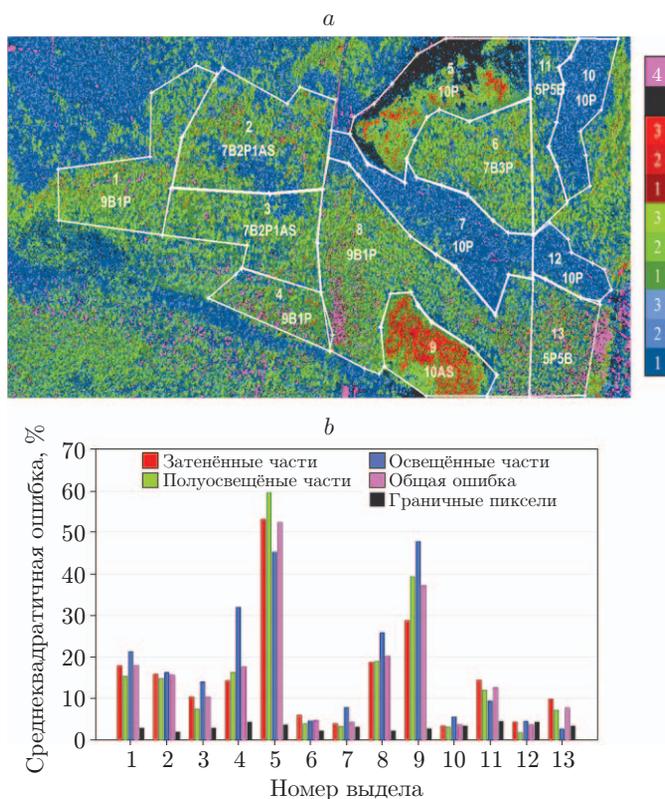


Рис. 4. Валидация результатов классификации породного состава древостоев по данным гиперспектрометра АВ-ВД: *a* — результаты классификации с наложением границ лесных выделов (верхняя цифра — номер выдела) и указанием данных по породному составу (B — берёза, P — сосна, AS — осина); цифрами на цветовой шкале обозначены градации освещённости лесного полога (1, 2, 3 — полностью затенённые, полуосвещённые и полностью освещённые части, 4 — нераспознанные пиксели (отказ от классификации)); *b* — ошибки распознавания породного состава лесных выделов для различных градаций освещённости лесного полога, общая ошибка классификации и естественный уровень ошибки за счёт граничных пикселей

лесного полога даны на рис. 4, *b*. Для большинства выделов ошибка не превышает 20 %. При этом уровень ошибок, возникающих за счёт граничных пикселей, составляет 3–5 %. Средняя ошибка, взвешенная на площадь выделов для всех градаций освещённости, равна 15 %, ошибка наземных данных — 10 %.

Можно видеть, что состав чистых древостоев (содержащих только одну породу) воспроизводится более точно, чем состав смешанных древостоев. Исключением являются выделы 5 и 9. Выдел 5 согласно данным лесотаксации содержит чистый молодой сосновый древостой. Значительное несоответствие объясняется тем, что на территории данного выдела производилась вырубка, однако план по посадкам сосен не был реализован. Поэтому территория выдела начала зарастать естественным образом в соответствии с породами прилегающих выделов: южная часть — берёзой, северная — сосной. Если исключить из рассмотрения этот выдел, то ошибка дистанционной инвентаризации породного состава тестовой территории будет сопоставима с ошибкой наземных данных.

Заключение. Предложена методика тематической обработки гиперспектральных изображений. Основу методики составляет алгоритм декодирования, который решает задачу многоклассовой классификации на базе серии решения задач бинарной классификации. Модификация алгоритма декодирования позволяет эффективно использовать преимущества стандартных алгоритмов классификации. Приведены рекомендации по выбору кодовой матрицы в зависимости от числа классов. Тестовые эксперименты на модельных данных показали, что точность предлагаемой методики классификации достигает точности оптимального классификатора. Проведены численные эксперименты по тематической обработке авиационных гиперспектральных изображений. Показано соответствие априорной и независимой оценок ошибки классификации. Оценка полной вероятности ошибочной классификации различных природных и искусственных объектов составляет менее 7 %. Предлагаемая методика может быть эффективно использована при создании и обновлении лесотаксационных карт для определения породного состава смешанных древостоев.

СПИСОК ЛИТЕРАТУРЫ

1. **Shukla A., Kot R.** An overview of hyperspectral remote sensing and its applications in various disciplines // *IRA-Intern. Journ. Appl. Sci.* 2016. **5**, Is. 2. P. 85–90.
2. **Козодеров В. В., Дмитриев Е. В., Каменцев В. П.** Когнитивные технологии дистанционного зондирования в природопользовании. Тверь: Изд-во Тверского гос. ун-та, 2016. 280 с.
3. **Борзов С. М., Потатуркин О. И.** Классификация типов растительного покрова по гиперспектральным данным дистанционного зондирования Земли // *Вестн. НГУ. Сер. Информационные технологии.* 2014. **12**, вып. 4. С. 13–22.
4. **Marshall M., Thenkabail P.** Advantage of hyperspectral EO-1 Hyperion over multispectral IKONOS, GeoEye-1, WorldView-2, Landsat ETM+, and MODIS vegetation indices in crop biomass estimation // *ISPRS Journ. Photogrammetry and Remote Sensing.* 2015. **108**. P. 205–218.
5. **Vyas D., Krishnayya N. S. R., Manjunath K. R. et al.** Evaluation of classifiers for processing Hyperion (EO-1) data of tropical vegetation // *Intern. Journ. Appl. Earth Observation and Geoinform.* 2011. **13**, N 2. P. 228–235.
6. **Dalponte M., Bruzzone L., Vescovo L., Gianelle D.** The role of spectral resolution and classifier complexity in the analysis of hyperspectral images of forest areas? // *Remote Sensing of Environment.* 2009. **113**, N 11. P. 2345–2355.
7. **Остриков В. Н., Плахотников О. В., Кириенко А. В.** Обработка гиперспектральных данных, получаемых с авиационных и космических носителей // *Современные проблемы дистанционного зондирования Земли из космоса.* 2013. **10**, № 2. С. 243–251.

8. **Ghosh A., Fassnacht F. E., Joshi P. K., Koch B.** A framework for mapping tree species combining Hyperspectral and LiDAR data: Role of selected classifiers and sensor across three spatial scales // Intern. Journ. Appl. Earth Observation and Geoinform. 2014. **26**, N 1. P. 49–63.
9. **Dietterich T. G., Bakiri G.** Solving multiclass learning problems via error-correcting output codes // Journ. Artificial Intelligence Research. 1995. **2**. P. 263–286.
10. **Борзов С. М., Потатуркин О. И.** Исследование эффективности спектрально-пространственной классификации данных гиперспектральных наблюдений // Автометрия. 2017. **53**, № 1. С. 32–42.
11. **Melnikov P. V., Pestunov I. A., Rylov S. A.** Comparison of spectral-spatial classification methods for hyperspectral images of high spatial resolution // Journ. Siberian Federal University. Eng. & Technol. 2017. **10**, Is. 6. P. 805–811.
12. **Kozoderov V. V., Dmitriev E. V.** Testing different classification methods in airborne hyperspectral imagery processing // Opt. Express. 2016. **24**, N 10. P. A956–A965.
13. **Freund Y., Schapire R. E.** A decision-theoretic generalization of on-line learning and an application to boosting // Journ. Computer and System Sci. 1997. **55**, N SS971504. P. 119–139.
14. **Дмитриев Е. В., Козодеров В. В.** Проблемы гиперспектрального авиационного мониторинга почвенно-растительного покрова // Журнал СФУ. Сер. Техника и технологии. 2015. **8**, № 8. С. 1083–1095.
15. **Kozoderov V. V., Kondranin T. V., Dmitriev E. V., Sokolov A. A.** Retrieval of forest stand attributes using optical airborne remote sensing data // Opt. Express. 2014. **22**, N 13. P. 15410–15423.

Поступила в редакцию 6 февраля 2018 г.
