

УДК 528.852

Classification of the Hyperspectral Data of the Earth Remote Sensing

Aleksey A. Buchnev and Valeriy P. Pyatkin*

*Institute of Computational Mathematics
and Mathematical Geophysics SB RAS*

6 Akademician M.A. Lavrenteva, Novosibirsk, 630090, Russia

Received 22.01.2018, received in revised form 20.07.2018, accepted 07.11.2018

The system of classification of hyperspectral data of the Earth remote sensing is discussed. This system consists of automatic (cluster analysis) and supervised classifications. Applicability of Bayes classifier for the normally distributed feature vectors to processing of these data is estimated. Realization of the classifier of spectral-angular mapping is described.

Keywords: remote sensing, hyperspectral data, clustering, K-means algorithm, supervised classification, Bayesian classifier, spectral-angular mapping classifier.

Citation: Buchnev A.A., Pyatkin V.P. Classification of the hyperspectral data of the earth remote sensing, J. Sib. Fed. Univ. Eng. technol., 2019, 12(5), 536-541. DOI: 10.17516/1999-494X-0109.

Классификация гиперспектральных данных дистанционного зондирования Земли

А.А. Бучнев, В.П. Пяткин

*Институт вычислительной математики
и математической геофизики СО РАН*

Россия, 630090, Новосибирск, пр. Академика Лаврентьева, 6

Рассматривается система классификации гиперспектральных данных дистанционного зондирования Земли (ДЗЗ), включающая автоматическую (кластерный анализ) и контролируруемую классификации. Оценивается применимость классификатора Байеса для нормально распределенных векторов признаков к обработке этих данных. Описана реализация классификатора спектрально-углового картирования.

Ключевые слова: дистанционное зондирование, гиперспектральные данные, кластерный анализ, алгоритм K-средних, классификация с обучением, классификатор Байеса, классификатор спектрально-углового картирования.

© Siberian Federal University. All rights reserved

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

* Corresponding author E-mail address: aleksei.buchnev@mail.ru, pvp@ooi.sccc.ru

Эффективность дистанционных исследований Земли из космоса во многом определяется используемыми методами обработки данных дистанционного зондирования Земли (ДЗЗ). Функциональные возможности разработанного в ИВМиМГ СО РАН совместно с ФГБУ «НИЦ «Планета» Росгидромета РФ программного комплекса PlanetaMonitoring для обработки данных ДЗЗ дают возможность использовать при их обработке практически функционально полный набор операций [1]. Программный комплекс PlanetaMonitoring реализует технологии предварительной и тематической обработки многоспектральных (МС) данных оптического, инфракрасного и микроволнового диапазонов. В настоящее время на орбитах находится более ста космических аппаратов ДЗЗ, большая часть которых оснащена МС-сканерами, позволяющими получать МС-изображения с высоким (0,3-10 м) пространственным разрешением и с небольшим количеством спектральных каналов. В последние годы в практику ДЗЗ активно внедряется гиперспектральная (ГС) аппаратура, спутники оснащаются ГС-сканерами, обеспечивающими съемку в видимом и инфракрасном диапазонах спектра [2, 3]. Отличительной особенностью ГС-изображений является их низкое пространственное разрешение (несколько десятков метров) и высокое спектральное разрешение (несколько сотен, а в перспективе – несколько тысяч каналов). Определенный импульс исследованиям по созданию алгоритмов, методов и технологий для обработки ГС-спутниковых данных в нашей стране был получен в связи с успешным запуском в последние годы российских космических аппаратов, оснащенных ГС-аппаратурой. Это спутники серии «Ресурс-П» № 1, № 2 и № 3 (запущен 13.03.2016). Спутник оснащен МС-сканером (пространственное разрешение 3 м, 5 каналов в диапазоне 0,45-0,9 мкм и в панхроматическом режиме (0,58-0,8 мкм) – 0,9 м) и ГС-сканером (пространственное разрешение 30 м, до 255 каналов со спектральным разрешением 5-10 нм в диапазоне 0,4-1,1 мкм). Цель настоящего исследования – оценить возможность использования системы классификации (контролируемой и неконтролируемой), реализованной в программном комплексе PlanetaMonitoring, в классификации ГС-данных ДЗЗ.

Известно, что непосредственное использование традиционных алгоритмов для обработки данных ГС-сканеров приводит к определенным трудностям, связанным с большой размерностью векторов измерений [3]. В рамках создания программного комплекса PlanetaMonitoring по обработке данных ДЗЗ была разработана система классификации, включающая классификацию с обучением (контролируемую классификацию) [4] и автоматическую классификацию (кластерный анализ) [5]. Контролируемая классификация реализует стратегию максимального правдоподобия Байеса для нормально распределенных векторов признаков в классах. Оценки параметров классификатора (векторов средних и ковариационных матриц в классах) получаются на основе обучающих выборок. В основе кластерного анализа лежит использование алгоритма К-средних.

Эти классификаторы, разработанные для обработки данных ДЗЗ с относительно небольшим количеством спектральных диапазонов (МС-данные), были достаточно легко адаптированы для обработки ГС-образов, полученных сканерами с высоким спектральным разрешением. Однако такое прямое использование классификаторов сопровождается следующими негативными обстоятельствами:

1. Высокие вычислительные затраты, обусловленные большой размерностью векторов признаков N (например, сканер ГСА спутника Ресурс-П № 3 имеет $N=255$ спектральных

каналов). Расстояние Махалонобиса, используемое для оценки близости векторов признаков к центрам классов, является значением квадратичной формы с весовой матрицей размерности $N*N$. Расстояние Махалонобиса также один из типов метрик, используемых нами в реализации алгоритма К-средних.

2. Для классификации с обучением требуются большие объемы обучающих данных. Чтобы избежать сингулярностей при вычислении матриц, обратных к ковариационным, минимальный объем обучающих выборок в классах должен равняться $N+1$ вектору. Известно, что для получения хороших оценок ковариационных матриц в мультиспектральных данных ($N \sim 10$) объем обучающей выборки V в классе должен быть таким, чтобы выполнялось следующее условие для соотношения размерности: $V/N \geq 100$ [6]. Ясно, что для выполнения подобного условия в ГС-данных ($N \sim 100$) объем обучающей выборки должен быть увеличен не менее чем на порядок.

Этих затруднений можно избежать путем извлечения из ГС-данных признаков, позволяющих построить набор векторов, размерность которых соответствует размерности МС-образа. С этой целью полный набор спектральных диапазонов разбивается на блоки (соседние блоки могут пересекаться), каждый из которых представлен в результирующем векторе признаков одним значением. Это значение может быть: 1) средним по блоку; 2) максимальным значением в блоке; 3) главной компонентой в блоке; 4) центральным значением в блоке. Кроме того, предоставляется возможность выбора главных компонент исходного набора ГС-векторов данных в соответствии с упорядоченной по убыванию последовательностью собственных чисел ковариационной матрицы. Количество выбираемых главных компонент определяется суммарной дисперсией.

К полученному так МС-образу в полной мере применимы функции системы контролируемой классификации программного комплекса PlanetaMonitoring. Хотя подобное снижение размерности ГС-образов используется достаточно часто (см., например, [7]), применять его следует с осторожностью из-за возможных потерь полезной информации. В частности, метод главных компонент, являясь линейным преобразованием, может разрушать нелинейности, присутствующие в данных [8].

В последнее время в состав системы контролируемой классификации комплекса PlanetaMonitoring нами включен еще один классификатор: классификатор спектрально-углового картирования [9]. Первоначально разработанный специально для классификации ГС-данных этот классификатор не содержит какой-либо специфики этих данных и может быть использован для классификации и МС-данных. В качестве расстояния $\rho(x, y_i)$ вектора признаков x до прототипа y_i класса i берут угловое расстояние:

$$\rho(x, y_i) = \arccos\left(\frac{x^T y_i}{\|x\| \cdot \|y_i\|}\right).$$

Вектор x зачисляется в класс, для которого угловое расстояние $\rho(x, y_i)$ является минимальным. В качестве прототипов классов (ссылочных векторов) авторы алгоритма [8] использовали библиотеку лабораторных спектров. Мы в своей реализации алгоритма в качестве прототипов классов используем векторы средних, полученных на основе обучающих и контрольных выборок классов. Для работы с обучающими и контрольными полями применяется тот же инструментарий, что и для классификатора Байеса [4]. Классификатор ра-

ботает в двух режимах: тестовом и рабочем. По результатам тестового режима вычисляется вероятность правильной классификации и строится матрица ошибок. Анализируя соответствующие показатели и выполняя при необходимости корректировки обучающих выборок классов, можно добиться необходимого уровня правильной классификации. В рабочем режиме классификатор обрабатывает весь набор данных, используя построенные прототипы классов. При этом может быть взято заданное пороговое значение угла a : если для вектора признаков x минимальное значение угла $\min_{\rho}(x, y_i) > a$, то этот вектор заносится в класс отклоненных векторов.

Для вычислительных экспериментов применяли один из файлов известного набора данных спектрометра AVIRIS [10]. Размер файла 614*2678, размерность векторов измерений 210. Из этого файла были сформированы два набора векторов признаков размером 614*2678 векторов: один с размерностью векторов признаков 184 (на основе визуального анализа спектральных составляющих векторов измерений) и второй с размерностью векторов признаков 21 (взяты главные компоненты из каждых 10 последовательных спектральных составляющих). Вычисления производились в рамках модели параллельного программирования OpenMP под управлением ОС Windows-10 на компьютере с 4-ядерным процессором i3-2100. В таблице приведены затраты времени в 1 с на контролируруемую классификацию указанных наборов данных классификаторами Байеса и спектрально-углового картирования в зависимости от числа выделяемых классов.

На рис. 1-4 демонстрируются результаты работы классификаторов (показаны нижние фрагменты полных рисунков). На рис. 1 представлено изображение спектрального канала с номером 72 исходного набора данных. Рис. 2 – результат кластеризации набора векторов признаков размерности 184 (выделялось 10 кластеров). Рис. 3 – результат, полученный применением классификатора Байеса к набору векторов признаков размерности 184. Белым цветом выделен класс отклоненных векторов. Чтобы избежать затруднений, связанных с выбором обучающих и контрольных полей в пространстве многомерных векторов признаков, для ввода этих полей использовалась карта кластеризации рис. 2. Априорные вероятности классов задавались с использованием выходных данных операции кластеризации. Вероятность правильной классификации составила 0.86 (0.998 для обучающих и 0.722 для контрольных полей). Рис. 4 – результат, полученный применением спектрально-углового классификатора к набору векторов признаков размерности 184. Были использованы те же обучающие и контрольные поля, что и в построении рис. 3. Вероятность правильной классификации составила 0.737 (0.821 для обучающих и 0.653 для контрольных полей).

Таблица

Table

Размерность векторов признаков	Классификатор Байеса		Спектрально-угловой классификатор	
	15 классов	10 классов	15 классов	10 классов
184	246	152	3.60	2.40
21	3.65	2.50	0.50	0.34

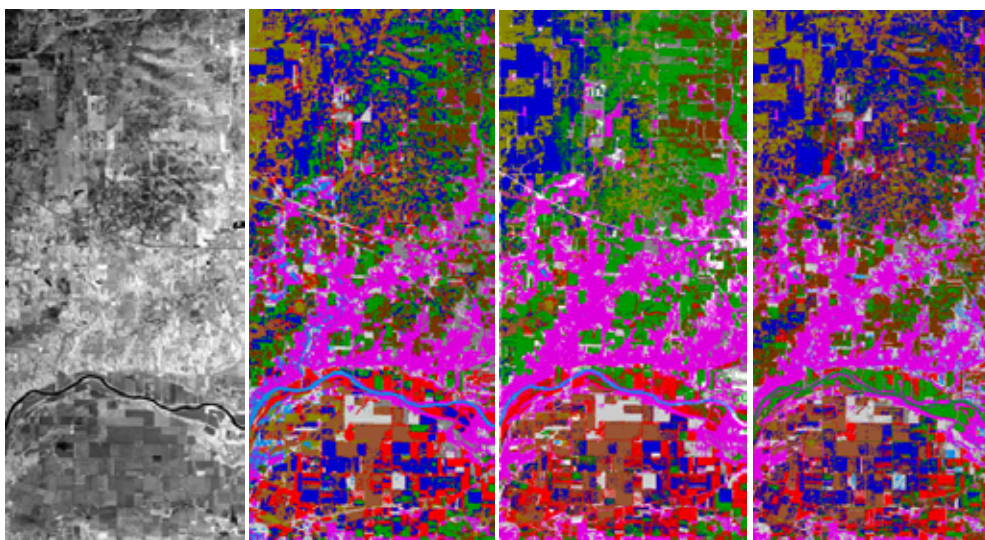


Рис. 1

Fig. 1

Рис. 2

Fig. 2

Рис. 3

Fig. 3

Рис. 4

Fig. 4

Включение технологии классификации ГС-данных ДЗЗ в состав программного комплекса PlanetaMonitoring позволяет расширить возможности системы классификации комплекса по построению карты тематических классов данных ДЗЗ различного спектрального разрешения.

Работа выполнена в рамках государственного задания ИВМиМГ СО РАН (проект 0315-2016-0003) при финансовой поддержке Российского фонда фундаментальных исследований (код проекта 16-07-00066).

Список литературы

[1] Асмус В.В., Бучнев А.А., Кровотынцев В.А., Пяткин В.П., Салов Г.И. PlanetaMonitoring: программный комплекс обработки спутниковых данных. *Проблемы информатики*, 2013, 3, 85-99. [Asmus V.V., Buchnev A.A., Krovotyntsev V.A., Pyatkin V.P., Salov G.I. Planetamonitoring: software complex processing of satellite data. *Problems of Informatics*, 2013, 7(3), 85-99 (in Russian)].

[2] Бондур В.Г. Современные подходы к обработке больших потоков гиперспектральной и многоспектральной аэрокосмической информации. *Исследование Земли из космоса*, 2014, 4-16. [Bondur V.G. Modern approaches to the processing of large flows of hyperspectral and multispectral aerospace information. *Issledovanie Zemli iz kosmosa*, 2014, 1, 4-16 (in Russian)].

[3] Шовенгердт Р.А. *Дистанционное зондирование. Модели и методы обработки изображений*. М.: Техносфера, 2010. 560 с. [Schowengerdt R.A. *Remote sensing. Models and methods for image processing*. Moskva, Technosfera, 2010, 560 p. (in Russian)].

[4] Асмус В.В., Бучнев А.А., Пяткин В.П. Контролируемая классификация данных дистанционного зондирования Земли. *Автометрия*, 2008, 4, 60–67. [Asmus V.V., Buchnev A.A.,

Pyatkin V.P. Supervised classification of the Earth remote sensing data. *Avtometriya*, 2008, 4, 60–67 (in Russian)].

[5] Асмус В.В., Бучнев А.А., Пяткин В.П. Кластерный анализ данных дистанционного зондирования Земли. *Автометрия*, 2010, 2, 58-66. [Asmus V.V., Buchnev A.A., Pyatkin V.P. Cluster analysis of the Earth remote sensing data. *Avtometriya*, 2010, 2, 58-66 (in Russian)].

[6] Асмус В.В. *Программно-аппаратный комплекс обработки спутниковых данных и его применение для задач гидрометеорологии и мониторинга природной среды. Дис...д-ра физ.-мат. наук.* Москва, 2002. 75 с. [Asmus V.V. *Software-hardware complex processing of satellite data and its application to problems of hydrometeorology and environmental monitoring.* Doctoral thesis, Moskva, 2002, 75 p. (in Russian)]

[7] Мельников П.В., Пестунов И.А., Рылов С.А. Экспериментальное сравнение методов классификации гиперспектральных изображений высокого пространственного разрешения по спектральным и пространственным признакам. *Региональные проблемы дистанционного зондирования Земли: сб. материалов III Междунар. научной конф.* Красноярск, СФУ, 2016, 28-33. [Melnikov P.V., Pestunov I.A., Rylov S.A. Experimental comparison of spectral-spatial classification methods of high spatial resolution hyperspectral images. *Regional problems of Earth remote sensing: proceedings of III International scientific conference*, Krasnoyarsk, SFU, 2016, 28-33 (in Russian)].

[8] J.P. Marques de Sa. *Pattern Recognition. Concepts, Methods and Applications.* Berlin, Springer, 2001, 318 p.

[9] Kruse, FA., Lefkoff A.B., Boardman J.W., Heidebrecht K.B., Shapiro A.T., Barloon P.J. and Goetz A.F.H. The Spectral Image Processing System (SIPS) – interactive vialization and analysis of imaging spectrometer data. *Remote Sensing of Environment*, 1993, 44, 145-163.

[10] AVIRIS: Airborne visible / Infrared Imaging Spectrometer [Electronic resource]. Access: https://aviris.jpl.nasa.gov/data/get_aviris_data.html