

УДК 004.9

## КЛАСТЕРИЗАЦИЯ ГИПЕРСПЕКТРАЛЬНЫХ ДАННЫХ МОНИТОРИНГА ОБЪЕКТОВ ТРАНСПОРТНОЙ ИНФРАСТРУКТУРЫ

© 2014 Т.И. Михеева, А.А. Федосеев

Самарский государственный аэрокосмический университет им. академика С.П. Королева  
(национальный исследовательский университет)

Поступила в редакцию 17.12.2013

Рассматриваются вопросы обработки данных дистанционного зондирования Земли с целью обнаружения и идентификации объектов природного и антропогенного происхождения. Приводятся результаты исследований, проведённых по гиперспектральному снимку территории Самарской области.

*Ключевые слова:* дистанционное зондирование, спектральная сигнатура, смешанный пиксел, атмосферная коррекция, метод независимых компонент, согласованная фильтрация

Объемы массивов данных, накопленных за время использования информационных систем, настолько внушительны, что человеку часто не по силам проанализировать их. Необходимость проведения такого анализа вполне очевидна, т.к. в «сырых данных» заключены знания, которые являются уникальным фундаментом для проведения исследований и могут быть использованы в дальнейшем для принятия решений. Глубокий анализ имеющихся данных позволяет понять природу происходящих процессов, найти и изучить неочевидные зависимости между явлениями. Для проведения автоматического анализа данных мониторинга объектов используется технология Data Mining, позволяющая в «сырых данных» обнаружить ранее неизвестные, нетривиальные, практически полезные и доступные для интерпретации знания. Data Mining использует сложный статистический анализ и моделирование для нахождения моделей и отношений, скрытых в базе данных.

Data Mining является мультидисциплинарной областью, возникшей и развивающейся на базе достижений прикладной статистики, распознавания образов, методов искусственного интеллекта, теории баз данных и др. Отсюда обилие методов и алгоритмов, реализованных в различных действующих системах Data Mining. Многие из таких систем интегрируют в себе сразу несколько подходов. Алгоритмы, используемые в Data Mining, требуют большого количества вычислений. Раньше это являлось сдерживающим фактором широкого практического применения Data Mining, однако сегодняшний рост

производительности современных процессоров снял остроту этой проблемы. В последние годы наблюдается тенденция применения средств искусственного интеллекта в методах Data Mining. Методы искусственного интеллекта, основанные на нейронных сетях, являются менее требовательными к входному набору данных. Интеграция систем поддержки принятия решений, базирующихся на геоинформационных технологиях, с методами исследования данных повышает ценность географической информации. Геоинформационные системы (ГИС) являются хорошей средой для внедрения методов искусственного интеллекта.

Основная задача, к которой, так или иначе, относятся остальные, описанные ниже, это. Одна из основных задач, решаемых в ГИС, – построение слоя – означает заполнение его недостающих частей или построение слоя полностью по информации, имеющейся в других слоях, на основе нахождения некоторой функциональной зависимости между параметрами, полученными эмпирическим путем, и скрытыми теоретическими параметрами, определяющими существенные характеристики каждого конкретного объекта.

Эффективное управление объектами и мониторинг состояния современной дорожно-транспортной инфраструктуры связано с возможностью получения достоверной и точной информации о состоянии её статических объектов, типичным примером которых являются дороги. Развитие систем дистанционного зондирования позволило использовать материалы гиперспектральной съёмки («сырые данные») для решения задач мониторинга объектов транспортной инфраструктуры.

Средствами получения гиперспектральной информации являются датчики, размещённые как на космических аппаратах (КА), так и на воз-

*Михеева Татьяна Ивановна, доктор технических наук, профессор кафедры информационных систем и технологий.  
E-mail: mikheevati@its-spc.ru  
Федосеев Александр Андреевич, аспирант.  
E-mail: mikheevati@its-spc.ru*

душных носителях. В результате съёмки имеется набор данных видимого, ближнего инфракрасного и коротковолнового диапазона спектра, из которых необходимо выделить специальную информацию об объектах транспортной инфраструктуры.

При обработке гиперспектральных данных возникает ряд проблем, связанных с влиянием атмосферы, а также с наличием в пикселе изображения сочетания нескольких видов материалов (проблема «смешанного пиксела»).

Гиперспектральные данные содержат информацию не только о поверхности, но и о состоянии атмосферы. При прохождении солнечного излучения через атмосферу происходит изменение его интенсивности, вызванное влиянием присутствующих в атмосфере газов и аэрозолей. Изменение интенсивности излучения происходит главным образом за счёт процессов поглощения и рассеяния. Так, для устранения этого искажающего фактора и перевода изображения из единиц спектральной яркости в единицы спектрального коэффициента отражения производится специальная процедура атмосферной коррекции. Существует два класса алгоритмов атмосферной коррекции: эмпирические и физические. Первый тип алгоритмов основан на использовании статистических характеристик изображения и осуществляет скорее калибровку, а не компенсацию. Восстановить истинные параметры атмосферы в момент съёмки можно только на основе физических моделей. Классическая атмосферная коррекция осуществляется в три этапа. На первом восстанавливаются параметры атмосферы, главным образом, характеристики аэрозоля (дальность видимости и, по возможности, «тип» аэрозоля) и величина столба водяных паров. Так как существующие методы позволяют восстановить аэрозоли лишь по весьма ограниченному набору типов поверхностей (вода и очень тёмная почва), обычно рассчитывается только средняя дальность видимости по сцене. Спектральная сигнатура водяных паров довольно специфична, что позволяет оценить их количество для каждого пиксела сцены. Второй шаг - решение уравнения радиационного переноса для заданного типа аэрозоля и столба водяных паров с целью пересчёта зафиксированного датчиком излучения в коэффициенты отражения поверхности. Заключительный этап - дополнительная обработка, так называемая «спектральная доводка», устраняющая оставшиеся артефакты.

В основу современной технологии Data Mining положена концепция паттернов, отражающих фрагменты многоаспектных взаимоотношений в данных. Паттерны представляют собой закономерности, свойственные подвыборкам

данных, которые могут быть компактно выражены в понятной человеку форме. Построение паттернов производится методами, не ограниченными рамками априорных предположений о структуре выборки и виде распределений значений анализируемых показателей.

Основными методами автоматизированного спектрального анализа являются классификация и обнаружение целей – отнесение объектов к одному из заранее известных классов. Классификация является одной из проблем, решаемых в рамках анализа географических данных. Часто она является конечной целью исследования, но для детальной интерпретации необходимо найти связь между группами, а также степень зависимости от других данных, присутствующих на карте. Разновидностью классификационных задач является кластеризация – один из видов, построенный на группировке объектов на основе их свойств или данных, описывающих сущность объектов. Объекты внутри кластера должны быть «похожими» друг на друга и отличаться от объектов, вошедших в другие кластеры. Чем больше похожи объекты внутри кластера и чем больше отличий между кластерами, тем точнее выполнена кластеризация. Это задачи классификации без учителя, таксономии, типизации. Этот тип классификации основан преимущественно на опытных данных.

Основной проблемой спектральной селекции является неоднородность содержащегося в пикселе материала. Этот факт объясняется, во-первых, тем, что естественные поверхности редко состоят из однородного материала. Во-вторых, суммарное излучение от всех объектов в пределах элемента пространственного разрешения фиксируется аппаратурой как одиночный пиксел изображения. Смешанные пикселы оказывают значительное влияние на дешифрирование изображений. Использование целочисленных методов классификации смешанных пикселов приводит к ошибкам в определении площадей, т.к. каждый пиксел может быть отнесён только к одному конкретному классу, хотя при низком пространственном разрешении фактически он содержит два и более класса. Кроме того, смешанный пиксел может оказаться не согласованным ни с одним из эталонов или, напротив, иметь сходство со спектральной сигнатурой другого эталона, который в действительности отсутствует в этом пикселе. Проблема смешанного пиксела не решается простым улучшением пространственного разрешения, т.к. при этом появляются детали изображения, не регистрируемые ранее, т.е. вводятся новые спектральные классы. Более точные результаты дают методы субпиксельной обработки. Алгоритмы спектрального разделения ис-

пользуют широкий набор математических технологий для определения исходных компонентов смеси и оценки их процентного содержания.

Известным методом спектрального разделения является согласованная фильтрация. Согласованные фильтры повышают контраст цели относительно фона, который может быть представлен в виде структурированной или стохастической модели [1]. На выходе формируется изображение, в котором значения пикселей соответствуют процентному содержанию искомого эталона в элементах сцены (в интервале от 0,0 до 1,0). Другой путь преобразования данных с целью увеличения различий между фоном и объектами интереса - декорреляция изображений, полученных в различных спектральных каналах, с использованием метода главных компонент (РСА). Однако малые объекты интереса не вносят существенного вклада в статистику и могут не проявиться в новых компонентах. Для решения этой проблемы используется метод независимых компонент (ICA), позволяющий с большей вероятностью обнаружить редкие объекты. Он удобен тем, что не требует знания исходных источников и основан на предположении, что среди компонентов смеси не более чем один имеет гауссовское распределение, а также что они статистически независимы, т.е. ни один из них не передаёт никакой информации другим компонентам. Помимо независимости, для того чтобы гарантировать единственное решение, необходимо чтобы количество наблюдаемых компонентов было не меньше, чем число исходных сигналов [2].

В качестве исходных данных выбран снимок с КА EO-1 (датчик NuRegion) территории Самарской области. Аппаратура NuRegion производит съёмку в 242 каналах в диапазоне от 420 до 2400 нм с пространственным разрешением 30 м.

На рис. 1 показан пример обнаружения до-

рожного покрытия по эталону, взятому из спектральной библиотеки Геологической службы США (USGS). Поскольку проекция пиксела для датчика составляет 30 м, заведомо понятно, что все дороги представляют собой субпиксельные объекты, чаще всего смешанные с растительностью или грунтом. Согласованная фильтрация позволяет найти их на изображении.

Метод независимых компонент с помощью двумерных диаграмм рассеяния позволил выделить на гиперспектральном снимке эталоны четырёх объектов интереса (рис. 2).

По этим спектральным эталонам с помощью метода согласованной фильтрации были выявлены сами объекты (рис. 3). Полученные результаты показывают возможности обнаружения, классификации и мониторинга объектов транспортной инфраструктуры по космическим снимкам среднего разрешения. Для тщательного и подробного анализа рекомендуется использовать данные высокого пространственного разрешения.

В рамках проекта по созданию интеллектуальной транспортной системы (ИТС) города разработана автоматизированная система идентификации зависимостей в географических данных. Система использует геоинформационные технологии, методы искусственного интеллекта и имеет средства интеграции с ИТС: общие структуры данных, интерфейсы, модули и т.д. Программная часть ИТС – это средства хранения и обработки информации. Средства хранения информации включают в себя ГИС «ITSGIS», содержащую электронную карту, и базу данных с атрибутивной информацией. Средства обработки информации – подсистемы «ITSGIS» обеспечивают различные виды анализа «сырых данных», среди которых анализ отклонений, анализ значимости, т.е. выявление наиболее нехарактерных паттернов (шаблонов).

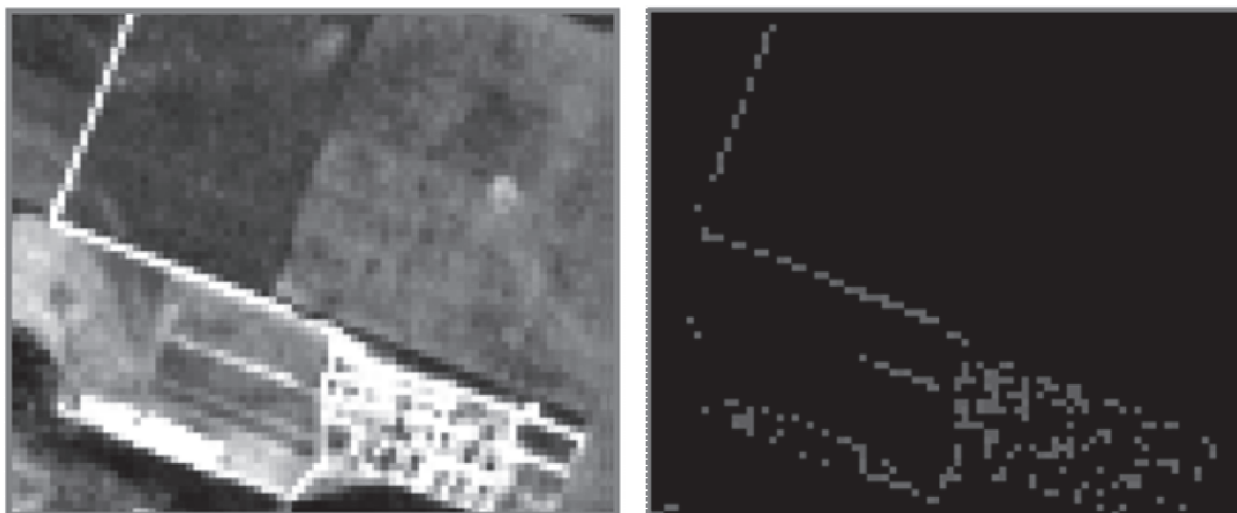


Рис. 1. Обнаружение дорожного покрытия методом согласованной фильтрации

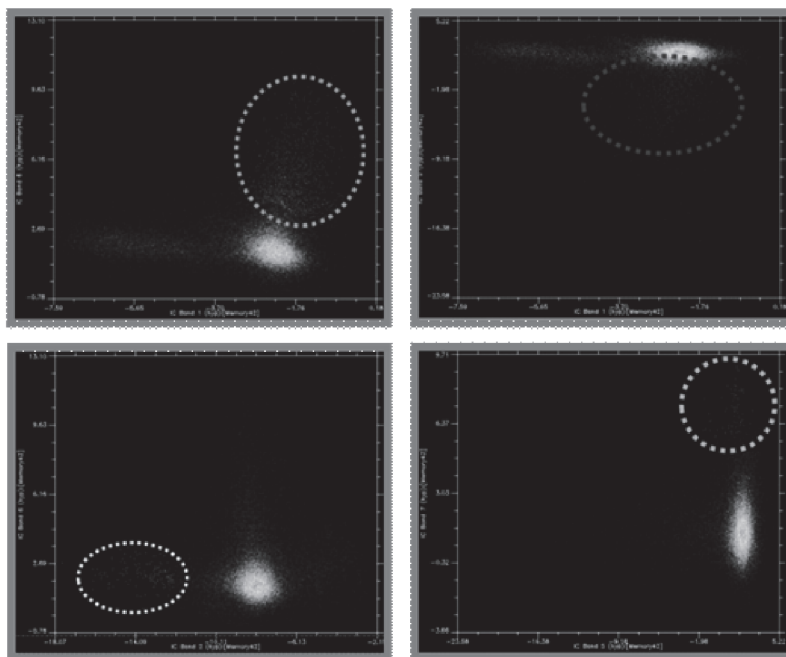


Рис. 2. Выделение эталонов объектов интереса

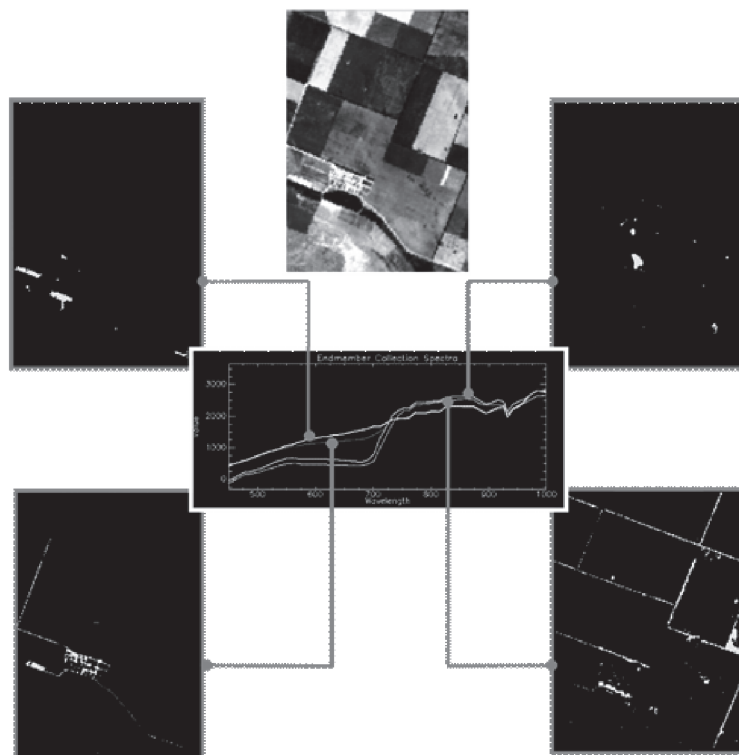


Рис. 3. Обнаружение объектов интереса по спектральным характеристикам

Процесс восстановления данных, восполнение пробелов в них осуществляется путем построения нового слоя по существующим слоям. При этом исследуется вопрос, какие данные (входные сигналы) являются доминирующими в процессе принятия нейросетью решения, а какие нет. В дальнейшем, последовательно убираются из рассмотрения те слои, которые в наименьшей степени влияют на восстановление отсутствующей информации. Значимость слоя формируется из

значимости его точек. Анализ значимости областей из слоев, участвующих в качестве входов, дает представление о территориальном распределении значимости.

Еще одна подсистема выполняет построение модели изолинейных изображений по регулярным и нерегулярным точкам, модели трехмерной визуализации. Расчет производится по содержащимся в базах данных числовым характеристикам. Моделироваться могут, как изображения

действительного рельефа или непрерывного поля, современного или с учетом динамических изменений, так и воображаемые поверхности, построенные по одному или нескольким показателям, например, поверхность интенсивности дорожного движения, распределение дорожно-транспортных происшествий, плотность транспортных потоков на улично-дорожной сети, плотность населения и т.п.

ГИС «ITSGIS» при решении задач кластеризации гиперспектральных данных мониторинга объектов транспортной инфраструктуры обеспечивает следующий набор функций:

- обучение нейросети на имеющихся в ГИС данных и построение аналитического выражения зависимости данных, находящихся на разных слоях электронной карты, в виде полинома (описательная модель);

- определение изменений в слоях карты при изменении данных в каком-либо одном из взаимосвязанных слоев (предсказательная модель);

- поиск объектов с похожей топологией в пределах одного слоя карты.

Система «ITSGIS» расширяется в следующих направлениях: реализации различных методов обучения нейросетей, использование различных архитектур нейросетей и расширение функциональности. Это позволит более детально проводить анализ огромных массивов данных.

## СПИСОК ЛИТЕРАТУРЫ

1. Михеева Т.И. Data Mining в геоинформационных технологиях // Вестник Самарского гос. техн. ун-та. Серия «Технические науки» №41. Самара: СамГТУ, 2006. С.96-99.
2. West J.E. Matched Filter Stochastic Background Characterization for Hyperspectral Target Detection / College of Science. Rochester Institute of Technology, 2005.
3. Robila S.A. Investigation of Spectral Screening Techniques for Independent Component Analysis Based Hyperspectral Image Processing // URL: <http://www.cs.uno.edu/~stefan> (дата обращения 15.11.2013).
4. Кухаренко Б.Г. Алгоритмы анализа компонент гиперспектральных изображений // Информационные технологии, Приложение. 2013. № 6. 32 с.
5. Барсемян А.А., Куприянов М.С., Степаненко В.В. и др. Методы и модели анализа данных: OLAP и Data Mining. СПб.: БХВ-Петербург, 2004. 336 с.
6. Осовский С. Нейронные сети для обработки информации [пер. с польского И.Д. Рудинского]. М.: Финансы и статистика, 2002. 344 с.
7. Михеева Т.И. Применение инструментальных средств проектирования интеллектуальной транспортной системы // Труды 6 международной научно-практической конференции «Организация и безопасность дорожного движения в крупных городах» - Санкт-Петербург: С-ПБАДИ, 2004. С. 85-89.
8. Михеева Т.И. Построение математических моделей объектов улично-дорожной сети города с использованием геоинформационных технологий // Информационные технологии. 2006. №1. С.69-75.

## HYPERSPECTRAL DATA CLUSTERING OF MONITORING OF TRANSPORT INFRASTRUCTURE OBJECTS

© 2014 T.I. Mikheeva, A.A. Fedoseev

Samara State Aerospace University named after Academician S.P. Korolyov  
(National Research University)

Questions of data handling of remote sensing of Earth for the purpose of detection and identification of objects of a natural and anthropogenous origin are considered. Results of the researches conducted on a hyper spectral picture of the territory of the Samara region are given.

*Keywords:* remote sensing, the spectral signature, the mixed pixel, atmospheric correction, method of the independent components, the coordinated filtering

---

*Tatyana Mikheeva, Doctor of Technical Science, Professor at the Information Systems and Technology Department.*

*E-mail: mikheevati@its-spc.ru*

*Aleksandr Fedoseev, Postgraduate Student.*

*E-mail: mikheevati@its-spc.ru*