

Ананьев Н.С.
Ананьев С.Н.
Куренков Н.И.
НИИ Восход (Москва)

Об одном подходе к экспресс-анализу гиперспектральных изображений

Известно, что гиперспектральная съемка является одним из важнейших направлений развития средств космического, авиационного и наземного наблюдения. Так, результаты аэрокосмической гиперспектральной съемки используются для решения задач контроля загрязнений окружающей среды, в картографии, в исследованиях природных ресурсов, климатологии, океанографии, для мониторинга земной поверхности, для наблюдения за пожарами и др. [1,2,3,4]. Средства наземной гиперспектральной съемки используются для выявления малозаметных объектов на фоне объектов природного, включая растительность, происхождения. Существенное увеличение возможности получения полезной информации о целевых объектах при использовании гиперспектральных изображений (ГСИ) основывается на спектральных различиях целевых объектов и фонов. Однако реализация этого направления представляет собой серьезную проблему, обусловленную требованием обработки огромных объемов иконической (видовой) информации (десятки и сотни гигабайт) практически в масштабе реального времени.

Основными направлениями решения этой проблемы являются развитие и внедрение компьютерных технологий обработки изображений, способных извлекать полезную информацию из ГСИ, а также оптимизация числа спектральных каналов аппаратуры гиперспектральной съемки.

Анализ источников, в которых отражен опыт создания технологий обработки аэрокосмических изображений у нас в стране и за рубежом, показывает, что развитие по первому направлению осуществляется путем создания быстрых алгоритмов обработки ГСИ, одновременно обеспечивающих распараллеливание вычислений [5, 6, 7, 8]. По второму направлению - к

уменьшению числа спектральных каналов (признаков) для решения конкретных тематических задач без потери качества их решения [9]. Основанием для проведения исследований на последнем направлении являются результаты анализа информационного содержания различных снимков, которые показывают, что при гиперспектральной съемке наиболее ценная для интерпретации информация сосредоточена, как правило, в небольшом числе каналов. В [9] для формального решения этой задачи предложен специальный функционал, имеющий смысл информативности набора выбранных спектральных каналов. Функционал построен на основе дивергенции Кульбака-Лейбнера (взаимной информационной мере статистических распределений) и учитывающем разрешающую способность и отношение сигнал/шум гиперспектрального изображения. С его помощью были получены оценки информативности спектральных каналов ГСИ, на основе которых можно определить их состав [9]. В конечном итоге оценка качества получаемых таким образом изображений осуществляется визуально операторами, что позволяет уже окончательно принять решение о выборе конечного набора спектральных каналов.

Исходя из определяющей роли визуального выделения искомым объектов на изображениях выбранных спектральных каналов ГСИ предлагается подход к определению набора спектральных каналов на основе использования результатов автоматической кластеризации спектральных слоев ГСИ. Под спектральным слоем ГСИ будем понимать матрицу цифровых значений спектральной яркости в элементах разрешения на определенной длине волны излучения (строки $i = 1, 2, \dots, n$; столбцы $j = 1, 2, \dots, m$) для каждого из дискретных $k = 1, 2, \dots, r$ измерительных спектральных каналов гиперкуба. При этом оператору для визуального анализа предъявляются изображения, полученные после усреднения попавших в соответствующий кластер спектральных слоев ГСИ, по результатам анализа которых оператор делает заключение о возможности выделения на них целевых объектов. Отметим, что этот подход во многом аналогичен так называемому разведочному анализу,

который проводится с целью качественной оценки используемых классификационных процедур при решении задач анализа многомерных данных [11].

Основаниями для выбора такого подхода явились:

- предположение, что при высоком линейном разрешении на объекте наблюдения, когда в пределах размера элемента разрешения изображения (везде далее для его обозначения будем использовать термин «пиксел») имеется однородная по спектральному составу поверхность, пикселы, относящиеся к целевым объектам, наблюдаемым в некотором диапазоне спектра (кластере), могут быть объединены в связные фрагменты, форма контуров которых является признаком их распознавания;

- временные затраты на кластеризацию спектральных слоев (матриц спектральных признаков) ниже, чем при кластеризации гиперспектров по их форме;

- яркостная картина объектов и фонов на местности локальна по пространству, что позволяет проводить кластеризацию спектральных слоев по частям независимо друг от друга и, следовательно, возможно распараллеливание вычислений с целью обеспечения их выполнения в масштабе реального масштаба времени независимо от размерности ГСИ (за счет увеличения числа вычислительных кластеров многопроцессорных систем обработки).

Как было сказано выше, использование ГСИ расширяет возможности по получению информации об объектах за счет различий в структуре спектральных признаков объектов и фонов. Поэтому алгоритм кластеризации должен в первую очередь выявлять такие различия, в том числе малозначимые. Для обеспечения такой возможности были разработаны алгоритм и реализующая его программа автоматической кластеризации, использующая специальную меру близости, предложенную в [13]:

$$\forall \mathbf{x} \in \mathbf{R}_+^n, \forall \mathbf{y} \in \mathbf{R}_+^n, \rho(\mathbf{x}, \mathbf{y}) = \log_2 \left(\sum_{i=1}^n \frac{x_i}{y_i} \sum_{i=1}^n \frac{y_i}{x_i} \right) \quad (1)$$

где \mathbf{x} и \mathbf{y} – сравниваемые массивы данных, $\mathbf{x}=(x_1, x_2, \dots, x_i, \dots, x_n)^T$, $\mathbf{x} \in \mathbf{R}_+^n \in \{\mathbf{y}=(y_1, y_2, \dots, y_i, \dots, y_n)^T \mid y_i > 0 \forall i = \overline{1, n}\}$, x_i, y_i – элементы массивов \mathbf{x} и \mathbf{y} (в нашем случае – это значения спектральных признаков в сравниваемых спектральных слоях гиперкуба).

Свойства (1) описаны в [14] и одной из ее основных особенностей является высокая чувствительность к малым изменениям структуры значений элементов массивов данных. Разработанный алгоритм реализует известный метод динамических сгущений и обеспечивает автоматическое определение числа кластеров в соответствии с особенностями структуры значений элементов массивов кластеризуемых данных [14].

Исходными данными для проверки работоспособности алгоритма были использованы результаты наземной гиперспектральной съемки 4-х человек на опушке леса, на которых надета защитная одежда, предназначенная для снижения их заметности на фоне растительности. Ширина диапазона регистрируемого излучения составляла примерно 400...1000 нм. При съемке для контроля результатов измерений использовался образец сравнения (эталон) диффузного отражения, спектральная характеристика которого имеет примерно постоянное значение во всем диапазоне длин волн отраженного излучения.

Указанные пять объектов рассматривались как целевые при анализе возможности их выделения на окружающем растительном фоне. Размеры гиперкуба составляли: $n = 157, m = 236, r = 332$.

В результате разбиения спектральных слоев гиперкуба были получены 7 кластеров. Номера спектральных каналов, попавших в эти кластеры, представлены в таблице 1. Как следует из таблицы, результаты разбиения спектральных слоев ГСИ на кластеры оказались в жесткой связи с последовательностью спектральных полос, формирующих диапазон видимого и ближнего инфракрасного излучения.

Таблица 1 - Номера кластеров и соответствующие им номера спектральных каналов ГСИ

	Номера кластеров						
	1	2	3	4	5	6	7
Номера спектральных каналов ГСИ	1-83	84-141	142- 171	172- 197	198- 224	225- 277	278- 322
Соответствующие спектральным каналам диапазоны длин волн излучения, нм	494...610	612...692	692...710	711...726	726...743	744...802	803...917
	8	9	10	11	12	13	14
Номера типовых спектральных каналов ГСИ	35, 40, 46	107, 112, 117	152, 155, 166	181, 187, 193	206, 211, 218	239, 251, 260	290, 300, 311

Рисунки хорошо иллюстрируют изменения структурных отличий яркости целевых объектов и окружающего растительного фона в различных диапазонах спектра отраженного излучения. В первую очередь это относится к первой фигуре слева, одетой в специальную малозаметную накидку, которая имеет малые отличия от фона на рисунках 1 и 2, и явно выделяется на окружающем фоне на всех остальных рисунках.

Поскольку в кластерах группируются близкие по форме спектральные слои ГСИ, то возникает возможность использования для проведения экспресс-анализа ГСИ не всей совокупности его спектральных слоев, а типовых представителей кластеров. Это позволит существенно снизить объем данных (как минимум на порядок), которые возможно будет передавать с борта носителей с минимальными затратами времени, чтобы обеспечить реальный масштаб времени их доведения для обработки на наземных пунктах.

Таблица 2 – Виды кластеров (все элементы спектральных каналов)



Рис.1 Вид кластера 1 (элементы 35, 40, 46)



Рис.2 Вид кластера 2 (элементы 107, 112, 117)



Рис.3 Вид кластера 3 (элементы 152, 155, 166)



Рис.4 Вид кластера 4 (элементы 181, 187, 193)



Рис.5 Вид кластера 5 (элементы 206, 211, 218)



Рис. 6 Вид кластера 6 (элементы 239, 251, 260)



Рис. 7 Вид кластера 7 (элементы 290, 300, 311)

Таблица 3 – Виды кластеров (результаты усреднения выбранных элементов спектральных каналов)



Рис.8 Вид кластера 1 (спектр. комп. 1-83)



Рис.9 Вид кластера 2 (спектр. комп. 84-142)



Рис.10 Вид кластера 3 (спектр. комп. 143-171)



Рис.11 Вид кластера 4 (спектр. комп.172-197)



Рис.12 Вид кластера 5 (спектр. комп 198-224)



Рис. 13 Вид кластера 6 (спектр. комп 224-277)



Рис. 14 Вид кластера 7 (спектр. комп 278-332)

Для оценки возможностей использования типовых элементов кластеров ГСИ для проведения экспресс-анализа из каждого кластера были выбраны по три элемента (всего 21 спектральный слой). Изображения результатов усреднения выбранных элементов каждого кластера представлены на рис. 8-14. Они показывают, что эти изображения практически ничем не отличаются от изображений на рис. 1-7 и их точно также можно использовать для проведения экспресс-анализа операторами на предмет выделения целевых объектов.

Таким образом, предложен новый метод быстрой обработки гиперспектральных данных для проведения их качественного экспресс-анализа оператором на предмет выявления целевых объектов и выбора состава спектральных каналов, которые можно использовать для этой цели. В основе метода лежит кластеризация гиперспектрального изображения по форме картины его спектральных слоев (яркости пикселей изображений), выбор типовых элементов каждого кластера, построение изображений усредненных выбранных элементов каждого кластера и представление их оператору. При реализации в бортовой аппаратуре обработки информации метод позволяет существенно снизить объем данных и сократить затраты времени на их передачу на наземные пункты обработки с целью обеспечения функционирования работы в масштабе реального времени с минимальным временем нахождения в эфире.

Список литературы

1. Родионов А.И., Зубков Б.В. и др. Развитие методов многомерных измерений при оптических исследованиях поверхности// Журнал технической физики. 2002. Т. 72, вып. 10. С.37-51.
2. Козодеров В.В., Кондранин Т.В., и др. Инновационная технология обработки многоспектральных космических изображений земной поверхности // Исследование Земли из космоса. 2008. №1. С.56-72.
3. Балтер Б.М., Егоров В.В. и др. Целевое выделение растительных сообществ по данным авиационной гиперспектральной съемки и многоспектрального сенсора ИСЗ Quickbird // Исследование Земли из космоса. № 6, 2008. С.14-42.

4. Козодеров В. В., Кондранин Т. В. и др. Обработка и интерпретация данных гиперспектральных аэрокосмических измерений для дистанционной диагностики природно-техногенных объектов // Исследования Земли из космоса. 2009, №2. С.36-54.

5. Овчинников А.М., Ролдугин Д.С., Овчинников М.Ю. Аппаратно-программный комплекс для обработки спектральной информации //Препринт ИПМ им. М.В. Келдыша РАН. М.- 26 с.

6. Обработка гиперспектральных изображений в программном комплексе ENVI. Сайт компании «Совзонд»

<http://www.sovzond.ru/dzz/publications/542/2345.html>

7. Андрианов Владимир. Искусство узнавания //Компьютера. 2000. №25. С. 14-15.

8. Кочнова И.В. Разработка технологии высокопроизводительной обработки аэрокосмических изображений методом параллельных вычислений: дисс. ... канд. техн. наук. – М., 2004 - 170 с.

9. Попов М.А., Станкевич С.А. Методы оптимизации числа спектральных каналов в задачах обработки и анализа данных дистанционного зондирования Земли // Современные проблемы дистанционного зондирования земли из космоса. М.: ИКИ РАН. 2006. Том 2, №1. С. 61-63.

10. Претт У. Цифровая обработка изображений. Кн. 1. – М.: Мир, 1982.-382 с.

11. Грузман И.С., Киричук В.С. и др. Цифровая обработка изображений в информационных системах. Учебное пособие. Новосибирск: НГТК. 2002. – 352с.

12. Айвазян С.А., Енюков И.С., Мешалкин Л. Д. Прикладная статистика. Классификация и снижение размерности. М.: Финансы и статистика. 1989.-608 с.

13. Куренков Н.И., Лебедев Б.Д. Энтропийный анализ многомерных данных. //Современные проблемы механики гетерогенных сред. Сб. ИПМ РАН к 10-летию его основания. М.: РАН, 2000. С. 27-29.

14. Куренков Н.И., Ананьев С.Н. Энтропийный подход к решению задач классификации многомерных данных// Информационные технологии. 2006. №8. С. 44-47.