

УДК 528.85/.87(15)

© Г. Н. Мальцев, И. А. Козин

Военно-космическая академия им. А. Ф. Можайского, Санкт-Петербург

georgy_maltsev@mail.ru

ОПТИМИЗАЦИЯ ЧИСЛА СПЕКТРАЛЬНЫХ КАНАЛОВ В ЗАДАЧАХ ОБРАБОТКИ И АНАЛИЗА ГИПЕРСПЕКТРАЛЬНЫХ ДАННЫХ ДИСТАНЦИОННОГО ЗОНДИРОВАНИЯ МИРОВОГО ОКЕАНА

Для решения многих задач мониторинга Мирового океана число спектральных каналов современных многоспектральных и гиперспектральных оптико-электронных систем дистанционного зондирования избыточно и требует минимизации. Целью такой оптимизации является формирование выборки снимков, полученных в различных наиболее информативных спектральных каналах, которая при размерности, существенно более низкой, чем число каналов гиперспектральной системы, обеспечивает решение задачи тематической обработки. Математически задача выбора наиболее информативных спектральных каналов гиперспектральной съемки для пикселя гиперспектрального датчика сформулирована как задача обнаружения координат изменения свойств зарегистрированного или эталонного дискретного спектрального образа, описываемого совокупностью спектральных составляющих. Совокупность составляющих дискретного спектрального образа в области анализа рассматривается как последовательность независимых случайных гауссовских величин с дисперсией и кусочно-постоянным средним, которое скачкообразно изменяется при переходе от одной дискретной координаты к другой. Представлен алгоритм решения поставленной задачи с использованием методов статистического оценивания в виде решающей статистики для обнаружения и оценки максимального правдоподобия координаты (спектрального канала) изменения свойств (скачков) анализируемого процесса. Последовательное применение данного алгоритма к выборке значений спектральных составляющих позволяет определить номера наиболее информативных спектральных каналов. Рассмотрен вопрос выбора (настройки) параметров синтезированного алгоритма выбора наиболее информативных спектральных каналов. Основными настраиваемыми параметрами являются величина порога и размер скользящего окна. В соответствии с рассмотренным алгоритмом, программно реализованным в среде программирования MatLab, приведен пример выбора наиболее информативных спектральных каналов для спектрального образа полученного из пикселя гиперспектрального изображения прибрежной территории поросшей водорослями.

Ключевые слова: мониторинг Мирового океана, дистанционное зондирование, обработка гиперспектральных снимков, спектральный канал.

G. N. Maltsev, I. A. Kozinov

A. F. Mozhaiskiy Military Space Academy, Saint-Petersburg, Russia

OPTIMIZING OF THE NUMBER OF SPECTRAL CHANNELS IN PROBLEMS OF PROCESSING AND ANALYSIS OF HYPERSPECTRAL REMOTE SENSING OF THE OCEAN DATA

The number of modern channels of multispectral and hyperspectral modern optoelectronic systems for remote sensing is redundant for many tasks of the ocean monitoring and requires minimization. The aim of this optimization is a creation of a selection of images obtained from different most informative spectral channels. The selection in dimension is significantly lower than the number of channels of hyperspectral system and provides a solution to the problem of thematic processing. Mathematically, the problem of choosing the most informative spectral channels of hyperspectral survey for pixel of sensor is formulated as the problem of detecting changes in the properties of the registered coordinates or reference discrete spectral image described by a set of spectral components. Set of discrete spectral image components in the analysis is considered as a sequence of independent random Gaussian variables with variance and piecewise constant mean, which abruptly changes from one discrete location to another. The algorithm for solving this problem using methods of statistical estimation as a crucial statistic for detection and maximum likelihood estimation coordinates (spectral channel) changes in the properties of the analyzed process is shown. The consistent application of the algorithm to the sample values

of the spectral components allows determination of numbers of the most informative spectral channels. The task of the configuration parameters of the synthesized algorithm to select the most informative spectral channels is considered. The basic parameters are the adjustable threshold value and the size of the sliding window. In accordance with the considered algorithms, implemented in MathLab software programming environment, an example of selecting the most informative spectral channels for spectral image obtained from hyperspectral image pixel of coastal area overgrown with algae is presented.

Key words: monitoring of the Oceans, remote sensing, processing hyperspectral imagery, spectral channel.

В настоящее время в связи с активной антропогенной деятельностью человечества остро встает проблема оперативного дистанционного контроля загрязнений атмосферы и гидросферы. По имеющимся оценкам аэрозоли антропогенного происхождения составляют 10—15 % в общем составе аэрозолей земной атмосферы, и их концентрация продолжает возрастать [1]. Мировой океан и, особенно, его шельфовая зона, занимающая 7.4 % акватории, так же как и атмосфера, находится под сильным влиянием антропогенных факторов. Индустриальное воздействие на природную среду настолько серьезно, что требует постоянного контроля экологической обстановки для своевременного принятия неотложных мер по сохранению в допустимых пределах уровня загрязнений атмосферы и гидросферы.

Наиболее оперативно контроль экологической обстановки осуществляется аэрокосмическими средствами дистанционного зондирования Земли (ДЗЗ) и атмосферы в оптическом и радио-диапазонах спектра. Современный этап развития методов ДЗЗ характеризуется активным использованием космических аппаратов (КА), оснащенных различными по характеристикам оптико-электронными системами наблюдения [2—4]. Оптико-электронная аппаратура ДЗЗ, устанавливаемая на КА, позволяет проводить периодическую съемку земной поверхности и таким образом наблюдать за изменениями окружающей среды и влиянием на нее человека. Современные космические системы дистанционного зондирования отличаются большим разнообразием, а их технические характеристики позволяют решать широкий круг задач в области мониторинга состояния природной среды. В настоящее время на КА используются оптико-электронные системы ДЗЗ, которые ведут съемку в десятках различных диапазонов электромагнитного спектра с разрешением от одного километра до одного метра или даже нескольких десятков сантиметров. Получают распространение также многоспектральные и гиперспектральные системы ДЗЗ, охватывающие до сотни спектральных зон шириной около 10 нм каждая.

Приведем перечень областей применения современных космических систем ДЗЗ [5]: мониторинг и оценка состояния окружающей среды; наблюдение за глобальными изменениями; сельское хозяйство; добыча полезных ископаемых; наблюдение за возобновляемыми природными ресурсами; метеорология; картография; военная область; средства массовой информации.

В настоящее время одним из направлений, в котором активно применяются космические системы ДЗЗ, является изучение Мирового океана. Наибольший интерес представляет исследование прибрежных зон Мирового океана [6]. Это обусловлено тем, что эти области, с одной стороны, являются перспективными зонами разведки природных ресурсов, а с другой стороны — они подвергаются интенсивным антропогенным воздействиям, наиболее существенными из которых являются глубинные стоки в океан. Динамика различных процессов в прибрежной зоне Мирового океана является предметом исследований, проводимых различными методами. При этом методы контактных измерений, хотя и обладают высокой точностью, весьма дороги и не позволяют получить полную картину происходящих процессов, например, распределения шлейфов загрязнений. Поэтому задача оценки физико-химических характеристик интересующих участков Мирового океана и обнаружения областей загрязнений дистанционными методами является весьма актуальной.

Постановка задачи оптимизации числа спектральных каналов. Для решения задач космического мониторинга Мирового океана могут быть эффективно использованы многоспектральные и гиперспектральные методы. Их главное преимущество заключается в использовании снимков, полученных в большом числе спектральных каналов, охватывающих практически непрерывный спектр в определенном диапазоне длин волн. Условно к многоспектральным относят системы ДЗЗ с десятками спектральных каналов, а к гиперспектральным — с сотнями спектральных каналов [4]. Использование при проведении океанографических исследований возможностей многоспектральных и гиперспектральных систем ДЗЗ может иметь ключевое значение для получения информации, необходимой при принятии решений. Данные, получаемые с помощью таких систем, позволяют осуществлять

дистанционную классификацию мелкомасштабных особенностей водной среды (например, тип и характеристики дна, цветение фитопланктона), выявление зависящих от глубины внутренних оптических свойств воды и регистрацию проявлений различных антропогенных воздействий [7, 8].

Многоспектральные и гиперспектральные оптико-электронные системы дистанционного зондирования обычно имеют среднее и низкое пространственное разрешение (десятки-сотни метров) и высокое спектральное разрешение (до единиц и десятков нанометров) [3, 4]. Так, спектральное разрешение данных современных гиперспектрометров составляет около 1 нм для длин волн видимой и ближней инфракрасной области спектра. При таком высоком спектральном разрешении в исходных данных проявляются отдельные спектральные линии различных минералов, природных образований, искусственных материалов и атмосферной среды, а также целые полосы таких высокомолекулярных соединений, как хлорофилл — основной пигмент фотозлементов. Это создает возможность распознавания наблюдаемых объектов и количественной оценки их состояния по спектральным признакам (образам). При этом наибольшее распространение получает корреляционное распознавание, основанное на определении выбранной меры близости векторов анализируемого и эталонных спектральных образов.

Экспериментальное изучение информационного содержания различных спутниковых снимков показывает, что если в многоспектральных снимках информация об объектах дистанционного зондирования обычно достаточно равномерно распределена внутри всего рабочего спектрального диапазона, то в случае гиперспектральных снимков наиболее ценная для интерпретации информация сосредоточена, как правило, в небольшом числе каналов [9, 10]. Данное обстоятельство позволяет проводить оптимизацию числа спектральных каналов в задачах обработки и анализа гиперспектральных снимков. Целью такой оптимизации является формирование выборки снимков, полученных в различных спектральных каналах, которая при размерности, существенно более низкой, чем число каналов гиперспектральной системы, обеспечивает решение поставленной тематической задачи (например, задачи распознавания объектов разных классов). Таким образом, при решении задач космического мониторинга Мирового океана и его прибрежных зон гиперспектральными съемочными системами ДЗЗ, встает задача определения наиболее информативных спектральных каналов, обеспечивающих решение задачи распознавания объектов с заданным качеством.

Под наиболее информативными спектральными каналами будем понимать спектральные каналы, которые при корреляционном распознавании объектов по спектральным образам позволяют обеспечить вероятность распознавания, близкую к вероятности распознавания по полному объему данных. Исходя из физических принципов корреляционной обработки, лежащей в основе автоматизированного распознавания по спектральным данным, для выбора наиболее информативных спектральных каналов при регистрации каждого гиперспектрального изображения необходимо выявлять некоррелированные спектральные каналы, соответствующие областям наиболее быстрого изменения спектрального образа наблюдаемого объекта [9—11]. Тем самым при проведении каждой съемки число спектральных каналов выбирается соответствующим так называемой существенной спектральной размерности данных наблюдений, что гарантирует наименьшие потери в достоверности распознавания. При этом обеспечивается выигрыш в достоверности распознавания по сравнению с использованием того же числа спектральных каналов, выбранных в рабочем диапазоне длин волн с равномерным «прореживанием».

Математическая постановка задачи выбора наиболее информативных спектральных каналов гиперспектральной оптико-электронной съемки может быть представлена как задача обнаружения координат изменения свойств зарегистрированного или эталонного дискретного спектрального образа, описываемого совокупностью M спектральных составляющих I_1, I_2, \dots, I_M . При принятых предположениях наиболее информативными являются спектральные каналы, соответствующие участкам существенного изменения свойств (скачков) анализируемого спектрального образа и дающие при корреляционном распознавании наибольший вклад в величину коэффициента корреляции. Решение поставленной задачи ищется с использованием методов статистического оценивания в виде решающей статистики для обнаружения и оценки максимального правдоподобия координаты (спектрального канала) изменения свойств (скачков) анализируемого процесса [12—14].

Решение задачи оптимизации числа спектральных каналов. Совокупность составляющих дискретного спектрального образа в области анализа I_1, I_2, \dots, I_M рассматривается как последовательность независимых случайных гауссовских величин [9, 10], характеризуемых дисперсией σ_j^2 и кусочно-постоянным средним, которое скачкообразно изменяется при переходе от одной дискретной координаты λ_m к другой. Координата скачка определяется на основании сравнения двух гипотез. Первая гипотеза H_0 состоит в том, что вся совокупность составляющих I_1, I_2, \dots, I_M соответствует

одной и той же модели процесса B_0 , гипотеза H_1 состоит в том, что в точке с координатой λ_k происходит изменение модели процесса B_0 на модель процесса B_1 . Отношение правдоподобия $W(k)$ гипотез H_0 и H_1 для координаты скачка λ_k , задаваемой ее номером k , при выполнении условия независимости спектральных составляющих I_m определяется выражением

$$W(k) = \prod_{m=1}^k \exp \left[-\frac{(I_m - \mu_{I0})^2}{2\sigma_I^2} \right] \cdot \prod_{m=k+1}^M \exp \left[-\frac{(I_m - \mu_{I1})^2}{2\sigma_I^2} \right] / \prod_{m=1}^M \exp \left[-\frac{(I_m - \mu_{I0})^2}{2\sigma_I^2} \right], \quad (1)$$

где μ_{I0} — среднее до скачка; μ_{I1} — среднее после скачка; σ_I^2 — дисперсия значений I_1, I_2, \dots, I_M анализируемой функции.

Строго говоря, дискретные составляющие спектрального образа $I_m, m = 1, \dots, M$, являются коррелированными, и учет этой корреляции лежит в основе устранения избыточности гиперспектральных данных при выборе наиболее информативных спектральных каналов. Однако представление анализируемых составляющих I_1, I_2, \dots, I_M дискретного спектрального образа $I^c(\lambda)$ как случайных величин, позволяет рассматривать их как независимые, и определять по координатам скачка точки разладки, в которых происходят существенные изменения анализируемой функции — спектрального образа.

При синтезе оптимальных алгоритмов обработки удобно рассматривать не отношение правдоподобия $W(k)$, а его логарифм $\Lambda(k)$. Это не изменяет результата, так как логарифмическая функция возрастает монотонно вместе со своим аргументом. Расчет для нормального и некоторых других распределений при использовании логарифма отношения правдоподобия оказывается несколько проще. Переходя к логарифму отношения правдоподобия (1), получаем:

$$\Lambda(k) = \frac{\nu_I}{\sigma_I^2} \sum_{m=k}^M \left(I_m - \mu_{I0} - \frac{\nu_I}{2} \right), \quad (2)$$

где введенная величина $\nu_I = \mu_{I1} - \mu_{I0}$ с учетом знака есть величина скачка функции, представленной составляющими I_1, I_2, \dots, I_M в анализируемой точке.

За неизвестную координату скачка λ_k принимается координата $\tilde{\lambda}_k$, задаваемая оценкой максимального правдоподобия номера точки \tilde{k} , в которой выполняется гипотеза H_1 . С учетом выражения (2) для логарифма отношения правдоподобия $\Lambda(k)$ оценка максимального правдоподобия может быть записана в виде

$$\tilde{k} = \arg \max_k \nu_I \sum_{m=k}^M \left(I_m - \mu_{I0} - \frac{\nu_I}{2} \right), \quad (3)$$

а правило принятия решения о наличии скачка в точке — решающая статистика — определяется следующим образом:

$$r : g_k = \Lambda(\tilde{k}) = \max_k \nu_I \sum_{m=k}^M \left(I_m - \mu_{I0} - \frac{\nu_I}{2} \right) \begin{matrix} > \\ < \end{matrix} \delta, \quad (4)$$

где δ — порог, выбираемый исходя из выбранного критерия. В выражениях (3) и (4) $k \in [1, \dots, M]$ — проверка условия максимизации функции правдоподобия и ее сравнение с порогом осуществляется для всех координат λ_m , охватывающих частотную область, в которой регистрируются дискретные составляющие спектрального образа $I_m, m = 1, \dots, M$.

Условие $\Lambda(\tilde{k}) > \delta$ определяет величину обнаруживаемого скачка анализируемой функции $\nu_I = \mu_{I1} - \mu_{I0}$, по сути дела — пороговую чувствительность детектора скачков — устройства или программно-аппаратных средств, анализирующих функцию $I(\lambda)$ по ее дискретным отсчетам I_1, \dots, I_M . На выбор величины порога δ оказывают влияние, с одной стороны, допустимый уровень ложных тревог (значимости решений) при обнаружении скачков, с другой стороны — требуемое или ожидаемое количество скачков (существенных изменений) функции $I(\lambda)$ на интервале анализа. При принятии представлений процедуры выбора наиболее информативных спектральных каналов как обнаружения изменений в анализируемом спектральном образе настройку величины порога δ можно рассматривать как адаптацию при автоматическом выборе рабочих спектральных каналов. При этом величина δ тем больше, чем меньше требуемое количество обнаруживаемых изменений, соответствующее числу спектральных каналов M_1 , запланированных для обработки.

На рис. 1 приведен пример реализации функции $I(\lambda)$ с одним обнаруживаемым скачком, представленной спектральным образом $I_m, m = 1, \dots, 100$, и поведения соответствующих ей логарифма функции правдоподобия $\Lambda(k)$ и решающей статистики g_k . При отсутствии увеличения среднего

значения функции $I(\lambda)$ логарифм отношения правдоподобия имеет тенденцию убывания (его среднее значение убывает по прямой с тангенсом угла наклона $-$). При наличии скачка среднего значения функции $I(\lambda)$ логарифм отношения правдоподобия $\Lambda(k)$ имеет тенденцию возрастания (его среднее значение возрастает почти линейно с тангенсом угла наклона $v_I^2/2\sigma_I^2$). Таким же образом в среднем возрастает и решающая статистика g_k , которая до момента скачка была близка к нулю.

В условиях априорной неопределенности, когда величина обнаруживаемого скачка v_I неизвестна, она является неизвестным параметром функции правдоподобия $\Lambda(k, v_I)$ и также подлежит оцениванию по критерию максимального правдоподобия. При известной координате точки скачка $\tilde{\lambda}_k$ (фиксированном k) оценка максимального правдоподобия величины скачка имеет вид:

$$\tilde{v}_I = \arg \max_{v_I} v_I \sum_{m=k}^M \left(I_m - \mu_{I0} - \frac{v_I}{2} \right) = \frac{1}{M - k + 1} \sum_{m=k}^M (I_m - \mu_{I0}). \quad (5)$$

При оценивании среднего μ_{I0} используется модель анализируемого процесса B_0 , а наличие явного решения для \tilde{v}_I вида (5) может быть присуще любой линейной модели анализируемого процесса B_1 . Однако если координата точки скачка неизвестна, то оценивание координаты скачка и его величины в каждой возможной координате составляет задачу двухпараметрической максимизации отношения правдоподобия.

С учетом выражения (5) оценка максимального правдоподобия номера точки, в которой выполняется гипотеза H_1 определяется выражением

$$\tilde{k} = \arg \max_k \frac{1}{2(M - k + 1)} \left[\sum_{m=k}^M (I_m - \mu_{I0}) \right]^2. \quad (6)$$

Соответствующее правило принятия оптимального решения при оценке координаты скачка \tilde{k} имеет вид:

$$r : g_k = \Lambda(\tilde{k}, \tilde{v}_I) = \max_k \frac{1}{2(M - k + 1)} \left[\sum_{m=k}^M (I_m - \mu_{I0}) \right]^2 \begin{matrix} > \delta \\ < \delta \end{matrix} \begin{matrix} H_1 \\ H_0 \end{matrix}. \quad (7)$$

Выражения (6) и (7) определяют процедуру оптимального выбора наиболее информативных спектральных каналов для спектрального образа, задаваемого составляющими $I_m, m = 1, \dots, M$. Реализуемый при этом алгоритм вычислений называется алгоритмом кумулятивных сумм. В работе [13] показано, что данный алгоритм является оптимальным последовательным алгоритмом,

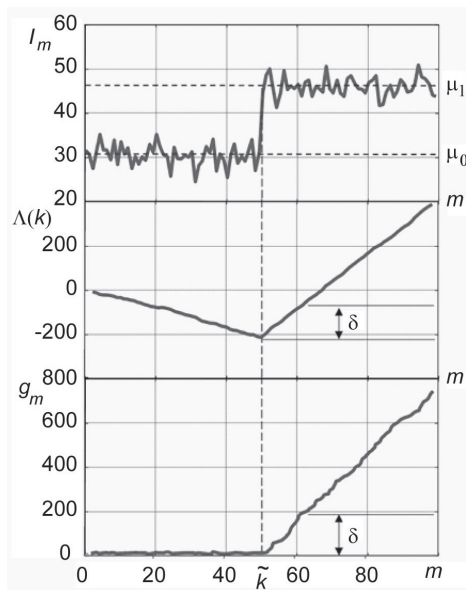


Рис. 1. Пример реализации функции спектрального образа I_m , соответствующих ей логарифма функции правдоподобия $\Lambda(k)$ и решающей статистики g_k .

который при фиксированной вероятности ошибки первого рода α минимизирует среднее число измерений, производимых от начала разладки до ее обнаружения. Получаемая при этом оценка \tilde{k} совпадает с моментом последнего отрыва от нуля статистики g_k . Алгоритм кумулятивных сумм (7) также называют алгоритмом кумулятивных сумм с отражающим от нуля экраном [14].

На практике необходимо обнаруживать не один единственный скачок в функции $I(\lambda)$, а несколько скачков. Общее число скачков, с одной стороны, определяется свойствами анализируемой функции и выбранным порогом. С другой стороны, на практике может задаваться необходимое количество обнаруживаемых скачков, которое определяется числом M_1 выбираемых наиболее информативных спектральных каналов. Для поиска нескольких скачков необходимо выполнить процедуру поиска первого скачка функции $I_i(\lambda)$, описываемую выражениями (6) и (7), для каждого значения $\lambda_m, m = 1, \dots, M$. При принятии решения о наличии скачка функции I_m в спектральном канале с номером k это значение фиксируется. Далее процедура оценки координаты скачка повторяется для значений $\lambda_m, m = k + 1, \dots, M$ до обнаружения следующего скачка и так далее до тех пор, пока не будут оценены все M значений функции $I_i(\lambda)$. Количество

выбираемых спектральных каналов M_1 ограничивается величиной порога δ .

На рис. 2 приведен пример выбора наиболее информативных спектральных каналов для спектрального образа I_m , $m = 1, \dots, 200$, по решающей статистике g_k в соответствии с рассмотренным алгоритмом. Спектральный образ получен из пикселя гиперспектрального изображения прибрежной территории, поросшей водорослями полученного авиационным сканирующим гиперспектрометром AVIRIS (производит съемку Земной поверхности в 224 спектральных каналах шириной по 10 нм в диапазоне 0.4—2.4 мкм). Рис. 2, *a* соответствует выбору $M_1 = 7$ спектральных каналов при пороговом значении $\delta = 0.1$, рис. 2, *б* — выбору $M_1 = 11$ каналов при $\delta = 0.01$. Выбирались и полагались наиболее информативными те спектральные каналы, для номеров которых решающая статистика превышает пороговое значение δ .

Как показали исследования, при последовательном поиске нескольких скачков анализ функции $I(\lambda)$ по всей выборке $m = 1, \dots, M$ не всегда позволяет обнаружить некоторые наиболее информативные спектральные каналы. Это является следствием распространения алгоритма обнаружения одиночной разладки на задачу обнаружения многократной разладки и связано с тем, что решение о наличии разладки (скачка) анализируемой функции $I_i(\lambda)$ принимается в результате сравнения разницы ее среднего значения μ_{i1} на интервале $m = 1, \dots, k$ и среднего значения на интервале $m = k + 1, \dots, M$. При этом на первом интервале ($m = 1, \dots, k$) до момента обнаружения разладки в силу выполнения гипотезы H_0 других разладок нет, а на втором интервале ($m = k + 1, \dots, M$) после момента обнаружения разладки, соответствующей выполнению гипотезы H_1 , могут иметь место и другие разладки. Если в моменты этих разладок на втором интервале имеют место скачкообразные изменения функции $I(\lambda)$ и в положительную, и отрицательную сторону, то эти изменения дают противоположный по знаку вклад в величину решающей статистики g_k и могут быть не обнаружены.

Для разрешения этой проблемы может быть использован подход, основанный на ограничении интервала анализа функции $I(\lambda)$, представленной дискретными отсчетами I_1, \dots, I_M . Многократная проверка с использованием критерия максимума логарифма отношения правдоподобия гипотезы H_0 при альтернативе H_1 осуществляется по ограниченной вспомогательной выборке I_{k+1}, \dots, I_{k+n} , выделенной с помощью скользящего окна, из исходной совокупности составляющих I_m , $m = 1, \dots, M$. Число элементов n во вспомогательной выборке, определяемое размером окна, выбирается таким образом, чтобы кумулятивная сумма формировалась из составляющих анализируемой функции, не содержащих противоположных скачков. Для выборки I_{k+1}, \dots, I_{k+n} вычисляется решающая статистика g_{k+1}^{k+n} и сравнивается с порогом δ . При $g_{k+1}^{k+n} \geq \delta$ выдается сигнал о наличии скачка, а при $g_{k+1}^{k+n} < \delta$ рассматривается следующее значение I_{k+n+1} и обрабатывается следующая вспомогательная выборка $I_{k+2}, \dots, I_{k+n+1}$.

В рассматриваемом случае имеем следующее правило принятия оптимального решения при оценке координаты скачка k :

$$r : g_k = \max_k \sum_{m=k+1}^{k+n} I_m \begin{matrix} H_1 > \frac{\sigma_I^2 \delta}{\nu_I} + n\mu_{I0} + n \frac{\nu_I}{2} \\ H_0 < \nu_I \end{matrix} \quad (8)$$

Алгоритм, определяемый выражением (8), представляет собой модифицированный алгоритм вычисления кумулятивных сумм со скользящим окном — частный случай алгоритма кумулятивной суммы со скользящим геометрическим средним [14]. При его использовании необходимо выбрать размер вспомогательной выборки (окна) n , соответствующий интервалу корреляции анализируемого спектрального образа $I(\lambda)$. Этот интервал корреляции характеризует поле рассеяния пар значений спектральных составляющих I_m, I_{m+t} , где t — интервал между отсчетами, выраженный в числе точек, и оценивается по автоковариационной функцией анализируемого спектрального образа $I(\lambda)$.

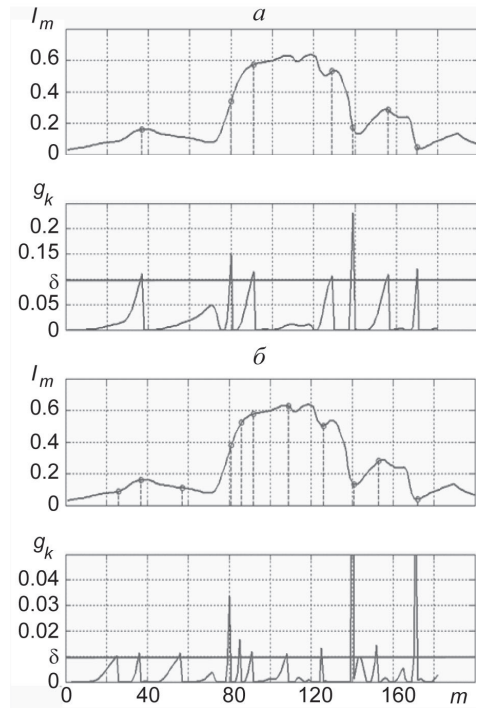


Рис. 2. Примеры выбора наиболее информативных спектральных каналов для спектрального образа I_1 ($M = 200$).

Важным свойством автоковариационной функции, используемым при обработке сигналов, является способность находить по ним характерные и повторяющиеся участки анализируемой функции [15]. В рассматриваемом случае анализ автоковариационной функции позволяет обнаруживать масштаб изменений анализируемой функции, соответствующий участкам изменения, которые дают противоположный по знаку вклад в величину решающей статистики g_k и могут быть не обнаружены. Объем вспомогательной выборки n можно определить по интервалу корреляции $\lambda_{\text{кор}}$ функции спектрального образа, соответствующему ширине центрального пика выборочного коэффициента автоковариации $r_\lambda(t)$ от максимума до первого пересечения с линией, взятой на некотором уровне, например, на уровне $r_\lambda = 0.5$. В дальнейшем полагается, что $n = \lambda_{\text{кор}0.5} / \Delta\lambda$, где $\lambda_{\text{кор}0.5}$ — интервал корреляции спектрального образа, определяемый по уровню $r_\lambda = 0.5$; $\Delta\lambda$ — расстояние между спектральными каналами.

При практической реализации рассмотренного алгоритма выбора наиболее информативных спектральных каналов для выбора параметров алгоритма могут быть использованы стандартные для математической статистики критерии качества процедур проверки гипотез — вероятности ошибок первого и второго рода (соответственно α и β) [16]. При этом гипотеза о наличии существенного изменения функции (скачка) H_1 является сложной, и вероятность ошибки второго рода β зависит от k — неизвестного номера спектрального канала, в котором имеет место обнаруживаемый скачок анализируемого спектрального образа.

Алгоритм можно настроить на определенные значения вероятности ошибки первого рода α и интервал запаздывания R в обнаружении скачка за счет выбора величины порога δ . Однако настройка алгоритма по этим параметрам позволяет качественно обнаруживать лишь некоторые скачки определенной амплитуды. На практике необходимо выбирать заданное число M_1 наиболее информативных спектральных каналов. Это в свою очередь приводит к необходимости установить некоторое конкретное значение порога δ , позволяющее обнаруживать M_1 скачков функции спектрального образа. Выбранное таким образом значение порога δ для различных скачков будет давать различное запаздывание в обнаружении, и обнаружение будет производиться с различным значением вероятности ошибки первого рода α .

Помимо порога δ на работу алгоритма влияет объем n анализируемой вспомогательной выборки I_k, \dots, I_{k+n} . Параметр n , определяющий размер скользящего окна, может принимать значения в диапазоне $[1, \dots, M-k]$. Настройка этого параметра позволяет ускорить процесс поиска скачков функции спектрального образа и обнаруживать скачки, не обнаруживаемые без использования скользящего окна (в частности, скачкообразные изменения функции $I_i(\lambda)$ в положительную и отрицательную сторону), и дополнительно скорректировать обнаружительные свойства алгоритма.

Объем вспомогательной выборки n можно определить по результатам анализа выборочной автокорреляционной функции анализируемого спектрального образа. В математической статистике используется несколько способов оценки интервала корреляции анализируемого процесса, но наиболее удовлетворительной оценкой автокорреляции на интервале t является оценка, основанная на выборочном коэффициенте автоковариации [15]

$$r_\lambda(t) = \frac{c_\lambda(t)}{c_0}, \quad (9)$$

где $c_\lambda(t) = \frac{1}{M} \sum_{m=1}^{M-t} \left(I_m - \frac{1}{(M-t)} \sum_{m=1}^M I_m \right) \left(I_{m+t} - \frac{1}{(M-t)} \sum_{m=1}^M I_m \right)$ — выборочная оценка автоковариации; $t = 1, 2, 3, \dots, T$; — дисперсия процесса (автоковариация при $t = 0$).

Отметим принципиальное значение использования именно автоковариационной, а не автокорреляционной функции при определении интервала корреляции анализируемого спектрального образа $I(\lambda)$. Спектральный образ, как зарегистрированная или эталонная функция яркости наблюдаемого объекта, является неотрицательной функцией и всегда имеет ненулевое среднее значение — постоянную составляющую. Наличие этой постоянной составляющей приводит к тому, что для стационарного ряда значений анализируемой функции автокорреляционная функция с увеличением t принимает положительные значения и демонстрирует свойство монотонного убывания по абсолютной величине. В результате масштаб автокорреляционной функции определяется интервалом определения анализируемой функции, а влияние на автокорреляционную функцию изменений анализируемой функции в пределах этого интервала оказывается незначительно. В то же время масштаб автоковариационной функции, которая вычисляется без учета постоянной составляющей анализируемой функции, определяется, прежде всего, масштабом изменений анализируемой

функции в пределах интервала ее определения. В общем случае с увеличением t автокорреляционная функция характеризуется знакопеременным изменением с тенденцией к затуханию, что позволяет выделить интервал корреляционной связи между отсчетами анализируемой функции $I(\lambda)$.

Отмеченными свойствами обладает и выборочный коэффициент автоковариации $r_\lambda(t)$, вычисляемый по формуле (9). Вычитание постоянной составляющей функции спектрального образа $\frac{1}{(M-t)} \sum_{m=1}^M I_m$ осуществляется для каждого из слугаемых внешней суммы по m при вычислении выборочной оценки автоковариации процесса $c_\lambda(t)$ и его дисперсии c_0 .

На рис. 3 представлен спектральный образ I_m , $m = 1, \dots, 200$, и его выборочная автоковариационная функция $r_\lambda(t)$. Для рассматриваемого случая интервал корреляции $\lambda_{кор}$ функции спектрального образа, определяемый как величина ширины центрального пика автокорреляционной функции от максимума до первого пересечения нулевой линии составляет $\lambda_{кор} = 40\Delta\lambda$. Анализ характера поведения функции анализируемого спектрального образа I_m , а также многих других функций спектральных образов показывает, что интервал корреляции $\lambda_{кор0.5}$, соответствующий уровню $r_\lambda = 0.5$, практически во всех случаях позволяет определить объем вспомогательной выборки n исключающей при использовании правила принятия решения (8) совместное влияние скачкообразных изменений функции $I(\lambda)$ в положительную и отрицательную сторону, которые не позволяют себя обнаружить при использовании правила принятия решения (7). Для анализируемого спектрального образа I_m объем вспомогательной выборки составляет $n = \lambda_{кор0.5} / \Delta\lambda = 20$. Исследования многочисленных функций спектральных образов типовых наземных объектов и подстилающей поверхности показывают, что значения n в большинстве случаев лежат в пределах от 10 до 50 [17].

В общем случае размер вспомогательной выборки n при настройке алгоритма анализа спектрального образа сверху ограничивается условием $n < M - k$ и выбирается исходя из условия $\lambda_{кор0.5} / \Delta\lambda \leq n \leq \lambda_{кор} / \Delta\lambda$. При выборе параметра n , результаты работы алгоритма становятся зависимыми только от выбора порога δ . Увеличение порога δ приводит к уменьшению количества выбираемых наиболее информативных спектральных каналов, а уменьшение δ — соответственно к увеличению их количества. Подбирая значения порога δ , можно осуществлять управление алгоритмом обработки спектрального образа по параметру M_1 — числу выбираемых наиболее информативных спектральных каналов.

Выбор уровня порога δ выбирается исходя из числа наиболее информативных спектральных каналов M_1 , которые необходимо обнаружить. Для каждой конкретной реализации функции спектрального образа значение порога δ для выбора заданного числа M_1 будет различным. Поэтому при необходимости выбора для всех анализируемых пикселей многоспектрального изображения одинакового количества M_1 наиболее информативных спектральных каналов необходимо для каждого пикселя подбирать свое значение порога δ . А при анализе всех пикселей гиперспектрального изображения с заданным порогом δ количество обнаруженных наиболее информативных спектральных каналов M_{i1} для каждого i -го пикселя в общем случае будет различным. При этом ограничение на число выбираемых наиболее информативных спектральных каналов M_1 может накладываться заблаговременно на этапе планирования гиперспектральной съемки.

Если существует условие выбора в любом i -ом пикселе гиперспектрального изображения для любой реализации функции $I(\lambda) = I_i(\lambda)$ одного и того же количества M_1 наиболее информативных спектральных каналов, то для каждой реализации функции $I(\lambda)$ необходим подбор порогового значения δ обеспечивающего выбор заданного количества M_1 спектральных каналов. Однако вычисление оптимального порогового значения δ , обеспечивающего выбор заданного количества M_1 наиболее информативных спектральных каналов для всего многообразия реализаций функций $I(\lambda)$, является задачей, не имеющей общего решения ввиду многообразия спектральных образов и условий многоспектральной съемки. Поэтому такая задача должна решаться итерационно для заданных $I(\lambda)$ и M_1 . Устанавливается некоторое значение порога δ и производится отбор наиболее информативных спектральных каналов для i -го пикселя в количестве M_{i1} . В случае если $M_{i1} < M_1$, необходимо пороговое значение δ уменьшить на некоторую величину $\Delta\delta$ и произвести повторный расчет, а если $M_{i1} > M_1$ необходимо пороговое значение δ увеличить на величину $\Delta\delta$ и также повторно произвести

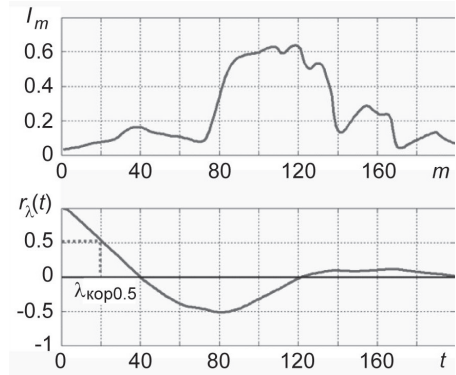


Рис. 3. Спектральный образ I_m и его выборочная автоковариационная функция $r_\lambda(t)$.

отбор наиболее информативных спектральных каналов. В результате искомое пороговое значение δ подбирается до тех пор, пока не выполнится условие $M_{i_i} = M_1$.

Таким образом, получено решение задачи оптимизации числа спектральных каналов гиперспектральных данных дистанционного зондирования, представляющее несомненный практический интерес для обработки и анализа гиперспектральных аэрокосмических снимков Мирового океана и прибрежной зоны. Задача выбора наиболее информативных спектральных каналов сформулирована как задача обнаружения координат изменения свойств (разладки) случайного процесса в виде правила принятия решения и оценки максимального правдоподобия координаты скачка анализируемой функции, определяемой с использованием алгоритма кумулятивных сумм. Для представляющего наибольший практический интерес случая определения последовательности скачков предложен модифицированный алгоритм вычисления кумулятивных сумм со скользящим окном, ограничивающий объем выборки, используемой при расчетах спектральных составляющих. Использование скользящего окна позволяет исключить пропуски скачков, дающих противоположный по знаку вклад в величину решающей статистики. Основными настраиваемыми параметрами алгоритма являются величина порога и объем вспомогательной выборки, определяющей размер скользящего окна.

Разработанный алгоритм вычисления кумулятивных сумм со скользящим окном реализован в виде прикладной программы и применен для анализа функции спектральных образов различных объектов океана и прибрежной зоны в целях выбора наиболее информативных спектральных каналов гиперспектральных систем ДЗЗ. Прикладная программа разработана в среде программирования Matlab и позволяет осуществлять попиксельный анализ гиперспектральных изображений с выбором наиболее информативных каналов и классификацией объектов на изображении корреляционным методом. Анализ полученных результатов демонстрирует работоспособность алгоритма и возможности его настройки с хорошей повторяемостью выбора наиболее информативных спектральных каналов в областях наиболее существенных изменений анализируемых типовых спектральных образов. Результаты проведенных исследований свидетельствуют, что с изменением настраиваемого порога происходит изменение чувствительности алгоритма с хорошей повторяемостью выбора наиболее информативных спектральных каналов в областях наиболее существенных изменений анализируемых спектральных образов.

Литература

1. Козинцев В. И., Орлов В. М., Белов М. Л. Оптико-электронные системы экологического мониторинга природной среды. М.: МГТУ имени Н. Э. Баумана, 2002. 528 с.
2. Еремеев В. А., Мордвинцев И. Н., Платонов Н. Г. Современные гиперспектральные сенсоры и методы обработки гиперспектральных данных // Исследование Земли из космоса. 2003. № 6. С. 80—90.
3. Рябова Н. В., Еськов Д. Н., Данилкин А. П. Малые искусственные спутники с оптико-электронной аппаратурой в программах дистанционного зондирования Земли // Оптический журнал. 1996. № 1. С. 4—19.
4. Мальцев Г. Н., Козинев И. А., Данилкин А. П. Космические системы и технологии многоспектрального дистанционного зондирования Земли // Информатика и космос. 2010. № 1. С. 148—158.
5. Шовенгердт Р. А. Дистанционное зондирование. Модели и методы обработки изображений. М.: Техносфера, 2010. 560 с.
6. Бондур В. Г. Аэрокосмические методы в современной океанологии // Новые идеи в океанологии. Т. 1. Физика. Химия. Биология. М.: Наука, 2004. С. 55—117.
7. Bondur V. G. Complex Satellite Monitoring of Coastal Water Areas // Proc. of 31 Int. Symp. on Remote Sensing of Environment. St.-Petersburg. 2005. P. 32—35.
8. Grace Chang et al. The new age of hyperspectral oceanography // Oceanography. 2004. P. 22—29.
9. Тарасов В. В., Якушенков Ю. Г. Многоспектральные оптико-электронные системы // Специальная техника. 2002. № 4. С. 56—62.
10. Дейвис Ш. М. и др. Дистанционное зондирование: количественный подход. М.: Недра, 1983. 415 с.
11. Мальцев Г. Н., Козинев И. А., Фатеев В. Ф. Методы выбора наиболее информативных спектральных каналов при дистанционном зондировании Земли с малых космических аппаратов // Известия вузов. Приборостроение. 2007. № 6. С. 23—31.
12. Жигляевский А. А., Красковский А. Е. Обнаружение разладки случайных процессов в задачах радиотехники. Л.: ЛГУ, 1988. 224 с.
13. Никифоров И. В. Последовательное обнаружение изменения свойств временных рядов. М.: Наука, 1983. 200 с.
14. Бассвиль М., Вилски А., Банвенист А. Обнаружение изменения свойств сигналов и динамических систем. М.: Мир, 1989. 280 с.
15. Бокс Дж., Дженкинс Г. Анализ временных рядов: прогноз и управление. М.: Мир, 1974. 408 с.
16. Крамер Г. Математические методы статистики: пер с англ. М.: Мир, 1975. 648 с.
17. Козинев И. А., Мальцев Г. Н. Модифицированный алгоритм обнаружения разладки случайного процесса и его изменение при обработке многоспектральных данных // Информационно-управляющие системы. 2012. № 3. С. 9—17.

Статья поступила в редакцию 10.11.2014 г.