

УДК 004.032.2; 004.932

МОДИФИЦИРОВАННЫЙ АЛГОРИТМ ОБНАРУЖЕНИЯ РАЗЛАДКИ СЛУЧАЙНОГО ПРОЦЕССА И ЕГО ПРИМЕНЕНИЕ ПРИ ОБРАБОТКЕ МНОГОСПЕКТРАЛЬНЫХ ДАННЫХ

И. А. Козин,

канд. техн. наук

Г. Н. Мальцев,

доктор техн. наук, профессор

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

Рассмотрено использование алгоритмов обнаружения разладки случайного процесса при обработке данных, полученных системами сбора информации. Исходные данные представляют собой функцию изменения по некоторой координате анализируемого параметра, по которой определяются координаты его существенного (скачкообразного) изменения. Представлено решение задачи обнаружения последовательности скачков анализируемой функции в виде решающей статистики для оценки максимального правдоподобия координат скачков с использованием модифицированного алгоритма кумулятивных сумм с настраиваемым скользящим окном. Приводится пример применения предложенного модифицированного алгоритма обнаружения разладки при обработке данных многоспектрального дистанционного зондирования и определении наиболее информативных спектральных каналов.

Ключевые слова — обработка данных, разладка случайного процесса, максимум правдоподобия, алгоритм кумулятивных сумм, многоспектральное дистанционное зондирование.

Введение

Задача обнаружения разладки (изменения свойств) стохастических сигналов и динамики систем является одной из широко распространенных задач анализа и обработки информации [1–4]. К ней сводятся многие прикладные задачи обработки данных, полученных системами сбора информации, такими как системы дистанционного зондирования, телеметрии, контроля технологических процессов и др. Обнаружение изменения свойств является составной частью анализа сигналов и динамических систем, оно представляет основу алгоритмов распознавания образов, контроля и технической диагностики информационно-управляющих систем, а также дополняет адаптивные процедуры идентификации состояний систем со сложной динамикой.

При формальной постановке задачи обнаружения существенных изменений свойств случайных процессов под разладкой понимают скачкообразное изменение описывающих этот процесс параметров, происходящее в неизвестный момент изменения процесса по той или иной коор-

динате, в большинстве случаев — во времени [1–4]. Задачей обнаружения разладки является установление факта разладки, и если считается, что разладка произошла, — оценивание момента разладки. Исходными для решения задачи обнаружения разладки являются данные о распределениях значений анализируемого процесса до и после возможной разладки, реализация случайного процесса, а в некоторых случаях также априорные распределения на множестве неизвестных параметров, одним из которых может быть координата момента разладки.

В то же время «классический» математический аппарат обнаружения разладки, основанный на алгоритме вычисления кумулятивных сумм [1, 2], позволяет обнаруживать однократную разладку случайного процесса и получать оценку максимального правдоподобия координат скачка. Кроме того, в большинстве прикладных задач обработки данных необходимо обнаружение последовательности скачков анализируемой функции. Для решения таких задач в настоящей работе предложен модифицированный алгоритм обнаружения последовательности скачков ана-

лизируемой функции на основе вычисления кумулятивных сумм с настраиваемым скользящим окном и показано его применение при анализе данных многоспектрального дистанционного зондирования.

Общее решение задачи обнаружения разладки с использованием алгоритма вычисления кумулятивных сумм

Пусть в качестве анализируемой выступает функция $f(x)$ координаты x , заданная в виде совокупности M дискретных составляющих f_m , каждая из которых соответствует координате x_m , $m = 1, \dots, M$. Необходимо определить моменты разладки, в качестве которых выбираются значения координат x_m , соответствующие областям, в которых функция $f(x)$ претерпевает явно выраженные изменения (резкое увеличение или уменьшение, локальные максимальные и минимальные значения). Эти области могут быть найдены методами поиска скачков в анализируемой функции, удовлетворяющих некоторому критерию. В качестве такого критерия при известных допущениях о функции потерь будем использовать критерий максимального правдоподобия, позволяющий получить оценки максимального правдоподобия моментов (координат) изменения свойств анализируемых процессов [1, 2].

Для того чтобы определить на некотором интервале изменения координаты x скачок функции $f(x)$, представленной составляющими f_1, \dots, f_M , необходимо в каждой координате x_m , $m = 1, \dots, M$, последовательно проверить выполнение двух гипотез: H_0 и H_1 . Гипотеза H_0 состоит в том, что вся совокупность составляющих f_1, \dots, f_M соответствует одной и той же модели B_0 . Гипотеза H_1 состоит в том, что совокупность составляющих f_1, \dots, f_k соответствует одной модели B_0 , а совокупность составляющих f_{k+1}, \dots, f_M соответствует другой модели B_1 . Модели B_0 и B_1 в общем случае параметризованы некоторым вектором. Если гипотеза B_1 истина, то координата x_k принимается за момент скачка (существенного изменения) функции $f(x)$. При равномерной дискретизации это эквивалентно определению номера k координаты x_k , в которой обнаружен скачок, $k \in [1, \dots, M]$.

Совокупность составляющих дискретной функции $f(x)$ в области анализа f_1, \dots, f_M может быть представлена как последовательность независимых случайных гауссовых величин с параметрами: дисперсией σ_f^2 и кусочно-постоянным средним μ_f , которое скачкообразно изменяется при переходе от одной дискретной координаты x_m к другой [3]. Обнаружение скачка в точке с координатой x_k эквивалентно принятию для анализируемой функции в этой точке гипотезы H_1 , когда

она проверяется против гипотезы H_0 . Поиск (просмотр) начинается в точке с координатой x_1 и осуществляется в направлении увеличения номера m . Тем самым решается задача обнаружения однократной разладки. После обнаружения первого скачка в точке с координатой x_k процедура поиска скачка повторяется и продолжает осуществляться в направлении увеличения номера m , начинаясь в точке с координатой x_{k+1} . В результате решается задача обнаружения многократной разладки — последовательно обнаруживаются все скачки (значимые изменения) дискретной функции $f(x)$.

В рамках допущения о гауссовой статистике составляющих f_1, \dots, f_k анализируемой функции $f(x)$ с помощью последовательных алгоритмов обнаружения разладки могут быть получены оценки моментов разладки, удовлетворяющие критерию максимального правдоподобия. Логарифм отношения правдоподобия различения гипотез H_0 и H_1 при гауссовом распределении дискретных составляющих функции $f(x)$ для координаты скачка x_k , задаваемой ее номером k , при выполнении условия независимости составляющих принимает вид

$$\Lambda(k) = \frac{\nu_f}{\sigma_f^2} \sum_{m=k}^M \left(f_m - \mu_{f0} - \frac{\nu_f}{2} \right), \quad (1)$$

где $\nu_f = \mu_{f0} - \mu_{f1}$ с учетом знака есть величина скачка функции, представленной составляющими f_1, \dots, f_M , в анализируемой точке, μ_{f0} — среднее до скачка, μ_{f1} — среднее после скачка; σ_f^2 — дисперсия значений f_1, \dots, f_M анализируемого ряда. Полагается, что изменение свойств связано с изменением кусочно-постоянного среднего μ_f , которое изменяет свое значение с μ_{f0} на μ_{f1} , а дисперсия σ_f^2 при переходе через точку скачка не изменяется.

Критерием качества проверки гипотез H_0 и H_1 является стандартный для задач различения сигналов [1, 5] критерий идеального наблюдателя, которому соответствуют и минимум суммы вероятностей ошибок первого и второго рода (соответственно α и β). Соответствующий алгоритм реализуется при сравнении логарифма отношения правдоподобия $\Lambda(k)$ с порогом, выбираемым исходя из условия минимума суммы вероятностей α и β или их заданных значений. Критерием качества оценки координаты скачка является максимум логарифма отношения правдоподобия $\Lambda(k)$.

За неизвестную координату скачка x_k принимается координата \tilde{x}_k , задаваемая оценкой максимального правдоподобия номера точки \tilde{k} , в которой выполняется гипотеза H_1 . С учетом выражения (1) оценка максимального правдоподобия \tilde{k} может быть записана в виде

$$\tilde{k} = \arg \max_k v_f \sum_{m=k}^M \left(f_m - \mu_{f0} - \frac{v_f}{2} \right), \quad (2)$$

а правило принятия решения о наличии скачка в точке \tilde{k} — решающая статистика — определяется следующим образом:

$$g_k = \Lambda(\tilde{k}) = \max_k v_f \sum_{m=k}^M \left(f_m - \mu_{f0} - \frac{v_f}{2} \right) \begin{matrix} > \\ < \end{matrix} \begin{matrix} H_1 \\ H_0 \end{matrix} > \delta, \quad (3)$$

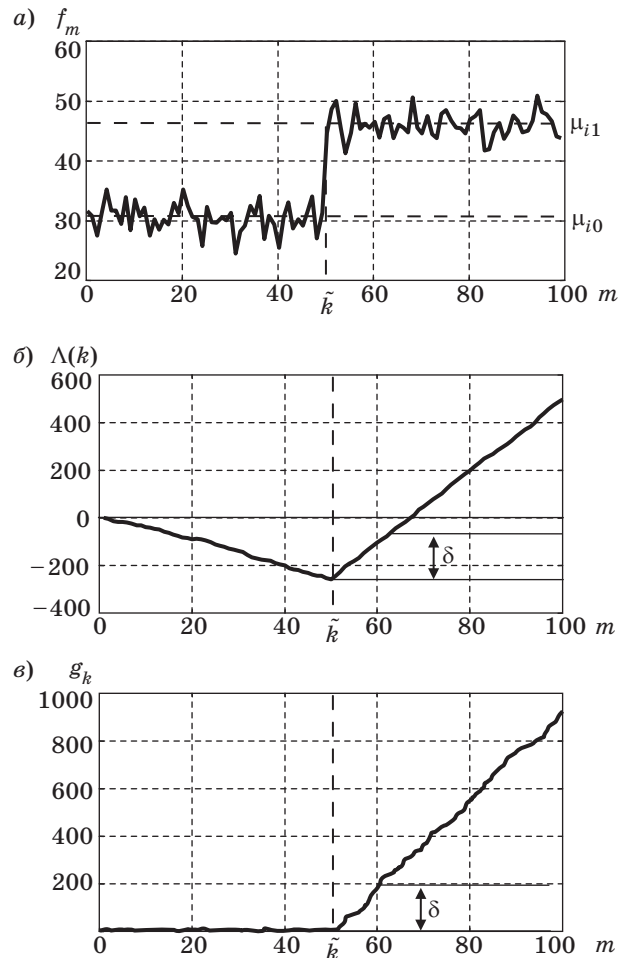
где δ — порог, выбираемый исходя из выбранного критерия; $k \in [1, \dots, M]$. В выражениях (2) и (3) проверка условия максимизации функции правдоподобия $\Lambda(k)$ и ее сравнение с порогом осуществляются с учетом значений дискретной функции $f(x)$ для всех координат x_m , $m = 1, \dots, M$.

Условие $\Lambda(\tilde{k}) > \delta$ определяет величину обнаруживаемого скачка анализируемой функции $v_f = \mu_{f0} - \mu_{f1}$, по сути дела — пороговую чувствительность детектора скачков — устройства или программно-аппаратных средств, анализирующих функцию $f(x)$ по ее дискретным отсчетам f_1, \dots, f_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 $f(x)$ на интервале анализа.

Пример, демонстрирующий принцип обнаружения ярко выраженного скачка анализируемой случайной функции $f(x)$, приведен на рис. 1. При отсутствии увеличения среднего значения функции f_m до скачка логарифм отношения правдоподобия имеет тенденцию убывания (его среднее значение убывает по прямой с тангенсом угла наклона $-v_f^2 / \sigma_f^2$) (рис. 1, а). После скачка среднего значения функции f_m логарифм отношения правдоподобия имеет тенденцию возрастания (его среднее значение возрастает по прямой с тангенсом угла наклона v_f^2 / σ_f^2) (рис. 1, б). Таким же образом в среднем возрастает и решающая статистика g_k , которая до момента скачка была близка к нулю (рис. 1, в).

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

$$\begin{aligned} \tilde{v}_f &= \arg \max_{v_f} v_f \sum_{m=k}^M \left(f_m - \mu_{f0} - \frac{v_f}{2} \right) = \\ &= \frac{1}{M - k + 1} \sum_{m=k}^M (f_m - \mu_{f0}). \end{aligned} \quad (4)$$



■ Рис. 1. Пример реализации функции $f(x)$ (а), поведения соответствующего ей логарифма функции правдоподобия $\Lambda(k)$ (б) и решающей статистики g_k (в)

При оценивании среднего μ_{f0} используется модель анализируемого процесса B_0 , а наличие явного решения для \tilde{v}_f вида (4) 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 [2].

С учетом выражения (4) оценка максимального правдоподобия номера координаты, которой выполняется гипотеза H_1 , определяется выражением

$$\tilde{k} = \arg \max_k \frac{1}{2(M - k + 1)} \left[\sum_{m=k}^M (f_m - \mu_{f0}) \right]^2. \quad (5)$$

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

$$g_k = \Lambda(\tilde{k}, \tilde{v}_f) = \max_k \frac{1}{2(M-k+1)} \left[\sum_{m=k}^M (f_m - \mu_{f0}) \right] \begin{matrix} > \\ < \end{matrix} \begin{matrix} H_1 \\ H_0 \end{matrix} \delta. \quad (6)$$

Выражения (5) и (6) определяют процедуру оптимального обнаружения момента и величины разладки анализируемой функции $f(t)$, задаваемой составляющими f_m , $m = 1, \dots, M$. Реализуемый при этом алгоритм вычислений называется алгоритмом кумулятивных сумм. В работах [1, 2] показано, что данный алгоритм является оптимальным последовательным алгоритмом, 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 (6) также называют алгоритмом кумулятивных сумм с отражающим от нуля экраном [3].

Модифицированный алгоритм вычисления кумулятивных сумм с настраиваемым скользящим окном

На практике при решении прикладных задач обнаружения разладки необходимо обнаруживать не единственный скачок функции $f(x)$, а несколько скачков. Для поиска нескольких скачков необходимо выполнить процедуру поиска первого скачка функции $f_i(x)$, описываемую выражениями (5) и (6), для каждого значения x_m , $m = 1, \dots, M$. При принятии решения о наличии скачка функции f_m в момент x_k это значение фиксируется. Далее процедура оценки координаты скачка повторяется для значений x_m , $m = k + 1, \dots, M$ до обнаружения следующего скачка и так далее до тех пор, пока не будут оценены все M значений функции $f_i(x)$. Количество обнаруживаемых скачков M_1 ограничивается величиной порога δ .

Как показала практика использования алгоритма кумулятивных сумм (6) при последовательном поиске нескольких скачков, анализ функции $f(x)$ по всей выборке $m = 1, \dots, M$ не всегда позволяет обнаружить необходимые скачки. 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 $f_i(x)$ принимается в результате сравнения разницы ее среднего значения μ_{f0} на интервале $m = 1, \dots, k$ и среднего значения μ_{f1} 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 $f(x)$ и в положительную, и отрицательную сторону, то эти изменения дают противоположный по знаку вклад в величину решающей статистики g_k , что затрудняет их обнаружение.

Для устранения этого эффекта предлагается использовать подход к обнаружению последовательности скачков анализируемой функции $f(x)$, основанный на ограничении интервала анализа дискретных отсчетов f_1, \dots, f_M . Многократная проверка с использованием критерия максимума логарифма отношения правдоподобия гипотезы H_0 при альтернативе H_1 осуществляется по ограниченной вспомогательной выборке f_{k+1}, \dots, f_{k+n} , выделенной с помощью скользящего окна, из исходной совокупности составляющих f_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 f_{k+1}, \dots, f_{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 f_{k+n+1} и обрабатывается следующая вспомогательная выборка $f_{k+2}, \dots, f_{k+n+1}$.

В результате имеем следующее правило принятия оптимального решения при оценке координаты скачка \tilde{k} :

$$g_{k+1}^{k+n} = \max_k \frac{v_f}{\sigma_f^2} \left[\sum_{m=k+1}^{k+n} f_m - n\mu_{f0} - n\frac{v_f}{2} \right] \begin{matrix} > \\ < \end{matrix} \begin{matrix} H_1 \\ H_0 \end{matrix} \delta. \quad (7)$$

Алгоритм, определяемый выражением (7), представляет собой модифицированный алгоритм вычисления кумулятивных сумм со скользящим окном. При его использовании необходимо выбрать размер вспомогательной выборки (окна) n , соответствующий интервалу корреляции анализируемой функции $f(x)$. Этот интервал корреляции характеризует поле рассеяния пар значений функции f_m, f_{m+l} , где l — интервал между отсчетами, выраженный в числе точек, и оценивается по выборочной автоковариационной функции $s_x(l)$ анализируемой функции $f(x)$.

Важным свойством автоковариационной функции является способность находить по ним характерные и повторяющиеся участки анализиру-

емого процесса (сигнала) [6]. 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 $x_{\text{кор}}$ анализируемой функции $f(x)$, соответствующему ширине центрального пика выборочного коэффициента автоковариации $r_x(l)$ от максимума до первого пересечения с линией, взятой на некотором уровне, например на уровне $r_x = 0,5$. В последующем полагается, что $n = x_{\text{кор}0,5}/\Delta x$, где $x_{\text{кор}0,5}$ — интервал корреляции спектрального образа, определяемый по уровню $r_x = 0,5$; Δx — расстояние между отсчетами f_1, \dots, f_M функции $f(x)$.

Выборочный коэффициент автоковариации определяется выражением

$$r_x(l) = \frac{1}{Mc_x(0)} \sum_{m=1}^{M-l} \left(f_m - \frac{1}{(M-l)} \sum_{m=1}^M f_m \right) \times \left(f_{m+l} - \frac{1}{(M-l)} \sum_{m=1}^M f_m \right), \quad (8)$$

где $l = 1, 2, 3, \dots, L$; $c_x(0)$ — автоковариация при $l = 0$ (дисперсия выборки). Отметим принципиальное значение использования именно автоковариационной, а не автокорреляционной функции при определении интервала корреляции анализируемой функции $f(x)$, которая в общем случае имеет ненулевое среднее значение — постоянную составляющую. Наличие этой постоянной составляющей приводит к тому, что для стационарного ряда значений анализируемой функции автокорреляционная функция с увеличением x принимает положительные значения и демонстрирует свойство монотонного убывания по абсолютной величине. В результате масштаб автокорреляционной функции выбирается в соответствии с интервалом определения анализируемой функции, а влияние на автокорреляционную функцию изменений анализируемой функции в пределах этого интервала оказывается незначительным. В то же время масштаб автоковариационной функции, которая вычисляется без учета постоянной составляющей анализируемой функции, определяется, прежде всего, масштабом ее изменений на интервале анализа, что позволяет выделить интервал корреляционной связи между отсчетами.

При вычислении значений $r_x(l)$ в соответствии с выражением (8) вычитание постоянной составляющей

осуществляется для каждой $\frac{1}{(M-l)} \sum_{m=1}^M f_m$

дого из слагаемых внешней суммы по m . Поэтому выборочный коэффициент автоковариации $r_x(l)$ обладает отмеченными свойствами и может быть использован для определения (настройки) размера окна при реализации модифицированного алгоритма кумулятивных сумм с настраиваемым скользящим окном.

Пример применения предложенного модифицированного алгоритма обнаружения разладки при обработке многоспектральных данных

Одной из областей применения предложенного алгоритма обнаружения последовательности скачков анализируемой функции является анализ данных многоспектрального дистанционного зондирования Земли (ДЗЗ). В современных системах ДЗЗ широко используются многоспектральные и гиперспектральные датчики с числом спектральных каналов до нескольких тысяч, а корреляционное распознавание наблюдаемых объектов осуществляется по зарегистрированным спектральным образам [7, 8]. В этих условиях для оперативной обработки и передачи данных ДЗЗ необходимо выбирать так называемые наиболее информативные спектральные каналы, соответствующие областям существенных изменений спектральных образов [7].

Рассмотрим пример работы модифицированного алгоритма вычисления кумулятивных сумм с настраиваемым скользящим окном при обнаружении наиболее информативных спектральных каналов многоспектральной системы ДЗЗ. В данном случае в качестве анализируемой функции $f(x)$ выступает спектральный образ наблюдаемого объекта или подстилающей поверхности $I(\lambda)$, где λ — длина волны излучения, в виде дискретных составляющих I_1, \dots, I_M значений коэффициента спектральной яркости, регистрируемой в M спектральных каналах, каждый со средней длиной волны λ_m и спектральным разрешением $\Delta\lambda_m$, $m = 1, \dots, M$. С точки зрения корреляционного распознавания, выбор наиболее информативных спектральных каналов состоит в определении спектральных составляющих, которые при распознавании дают наибольшую корреляцию зарегистрированного спектрального образа с эталонным. Очевидно, что при корреляционном распознавании по спектральным образам максимальный вклад в величину коэффициента корреляции вносят спектральные составляющие, соответствующие участкам наиболее быстрого изменения анализируемого спектрального образа $I(\lambda)$, на которых функция, описывающая спектральный образ, имеет некоторые локальные особенности и существенные изменения в характере поведения.

В рассматриваемом случае анализируемым процессом является спектральный образ $I(\lambda)$, и задача обнаружения изменения свойств анализируемого процесса переносится в частотную область (координатой изменения процесса является длина волны λ). При этом постановка задачи определения участков существенных изменений анализируемых спектральных образов полностью соответствует рассмотренной задаче обнаружения разладки анализируемого процесса, и для нее при принятых допущениях может быть получено строгое решение по критерию максимального правдоподобия.

В общем случае задача обнаружения изменения свойств (разладки) анализируемого процесса включает две последовательно решаемые задачи — установления факта изменения свойств и определения момента (координаты) изменения свойств, если считается (по результатам решения первой задачи), что оно произошло. В случае анализа спектральных образов при многоспектральном ДЗЗ решение двух указанных задач позволяет получить оценки координат λ_k или номеров k наиболее информативных спектральных каналов для распознавания по ограниченному объему многоспектральных данных. При этом следует отметить, что выбор для распознавания наиболее информативных спектральных каналов соответствует гипотезе существования существенной размерности спектральных признаков при распознавании объектов в системах ДЗЗ, что следует из физической природы оптического излучения [9]. В частности, установлено, что существенная размерность многоспектральных данных в диапазоне 0,4 ÷ 15,0 мкм, определяемая для явлений отражения и излучения энергии от поверхности Земли, составляет около шести. Близкую оценку спектральной размерности данных наблюдений дают и исследования вероятности распознавания по спектральным признакам техногенных объектов на фоне природной подстилающей поверхности [9].

При определении областей изменения зарегистрированного спектрального образа его составляющие в области анализа I_1, I_2, \dots, I_M рассматриваются как последовательность независимых случайных гауссовых величин с параметрами: дисперсией σ_I^2 и кусочно-постоянным средним μ_I , которое скачкообразно изменяется при переходе от одной дискретной координаты λ_m к другой. Оценки максимального правдоподобия координат k наиболее значимых изменений анализируемого спектрального образа и правила принятия решения (решающая статистика) с использованием алгоритма вычисления кумулятивных сумм спектральных составляющих со скользящим окном определяются выражениями (5) и (7)

соответственно с заменой составляющих f_1, \dots, f_M на составляющие дискретного спектрального образа в области анализа I_1, \dots, I_M , дисперсии σ_f^2 — на σ_I^2 , среднего значения μ_{f0} — на μ_{I0} и величины скачка v_f — на v_I . Последовательное применение данного алгоритма к выборке значений спектральных составляющих позволяет определить номера наиболее информативных спектральных каналов.

На рис. 2 представлены результаты применения рассмотренного модифицированного алгоритма обнаружения разладки к типовому спектральному образу наблюдаемого объекта, полученному в $M = 200$ спектральных каналах в диапазоне длин 0,35 ÷ 2,5 мкм. Для этого спектрального образа в соответствии с алгоритмом, определяемым выражениями (5) и (7) и запрограммированным в среде программирования MathLab, был осуществлен выбор наиболее информативных спектральных каналов при различных значениях параметров n и δ . На графиках приведен анализируемый спектральный образ, совмещенный с диаграммами, показывающими выбор областей существенного изменения спектрального образа в группах по 10 каналов, соответствующих номерам каналов, приведенным в таблицах под диаграммами для различных значений размера вспомогательной выборки n порога δ . Анализ диаграмм и табличных данных позволяет сравнить результаты выбора наиболее информативных спектральных каналов при различных значениях n и δ и выявить ряд тенденций.

При использовании вспомогательной выборки с изменяющимся объемом $n = M - k$ и значениями порога $\delta = 1, \dots, 0,01$ (рис. 2, а) хорошо обнаруживаются только наиболее крупные скачки функции спектрального образа (спектральные каналы 78–82, 137–139, 169–170). При пороге $\delta < 0,01$ дополнительно обнаруживаются менее крупные скачки функции спектрального образа (спектральные каналы 146, 168–169, 182–189). Таким образом, алгоритм с таким объемом вспомогательной выборки способен обнаруживать малые скачки функции спектрального образа только для больших номеров спектральных каналов, ближайших к M . Это обусловлено тем, что поиск скачков функции спектрального образа в канале с номером k ведется по результату сравнения среднего значения спектральных составляющих на интервале от предыдущего обнаруженного скачка до канала с номером k и среднего на интервале от k до M . В таком случае при малом значении k некоторые скачки функции спектрального образа не вносят существенного изменения в среднее значение спектральных составляющих интервала от k до M . При прибли-

жении значения k к M интервал вычисления среднего уменьшается, и малые скачки становятся более заметными для алгоритма.

При использовании вспомогательной выборки малого объема, например $n = 3$, и при малых значениях порога $\delta < 0,01$ (рис. 2, б) проявляется тенденция к обнаружению большого числа малых скачков. Таким образом, алгоритм становится более чувствительным к многочисленным шумовым составляющим функции спектрального образа, обнаруживая помимо значительных скачков функции спектрального образа все прочие менее значимые скачки.

В двух рассмотренных случаях выбор объема вспомогательной выборки n (размера окна) осуществлялся без учета величины интервала автокорреляции $\lambda_{\text{кор}0,5}$ автоковариационной функции анализируемого спектрального образа. Как показали исследования функций спектральных образов различных наземных объектов и подстилающей поверхности, для рационального выбора значения n может быть рекомендовано его определение по величине интервала корреляции $\lambda_{\text{кор}0,5}$, соответствующего ширине центрального пика автоковариационной функции $c_\lambda(l)$ анализируемой функции $I(\lambda)$. Выборочный коэффициент автоковариации дискретной функции спектрального образа $r_\lambda(l)$ определяется выражением вида (8) с заменой составляющих f_1, \dots, f_M на составляющие I_1, \dots, I_M . Для рассматриваемого в качестве примера спектрального образа определяемая таким путем величина вспомогательной выборки составляет $n = 20$.

Таким образом, в общем случае размер вспомогательной выборки n при настройке алгоритма анализа спектрального образа сверху ограничивается условием $n < M - k$ и выбирается исходя из условия $\lambda_{\text{кор}0,5} / \Delta\lambda \leq n \leq \lambda_{\text{кор}} / \Delta\lambda$. 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.

При использовании вспомогательной выборки объемом $n = 20$ и значениях порога $\delta = 1, \dots, 0,1$ (рис. 2, в) алгоритм позволяет обнаруживать только наиболее крупные скачки функции спектрального образа (спектральные каналы на интервалах 38, 81–83, 138–139, 170–174). При значении порога $\delta < 0,1$ помимо крупных скачков обнаруживаются менее значительные, а крупные скачки представляются обнаружением несколь-

ких скачков в близко расположенных спектральных каналах. Например, при значении порога $\delta = 0,1$ в анализируемой функции спектрального образа обнаруживается пять скачков с соответствующими номерами спектральных каналов 38, 81, 96, 138, 170. При снижении порога до $\delta = 0,005$ скачок функции спектрального образа, обнаруживаемый ранее в спектральном канале 38, также обнаруживается, но представлен уже тремя каналами с номерами 21, 33 и 51. Аналогично скачок в спектральном канале 81, обнаруживаемый при $\delta = 0,1$, при $\delta = 0,005$ представлен четырьмя каналами с номерами 79, 84, 90 и 105. Аналогичная тенденция наблюдается и в других случаях.

Результаты проведенных исследований, в частности диаграммы, приведенные на рис. 2, свидетельствуют, что с изменением порога происходит изменение чувствительности алгоритма с хорошей повторяемостью выбора наиболее информативных спектральных каналов в областях наиболее существенных изменений анализируемых спектральных образов. На практике при реализации алгоритма выбора наиболее информативных спектральных каналов необходимым условием является выбор для любой функции спектрального образа $I(\lambda)$ одного и того же количества M_1 спектральных каналов. При таком условии для каждой реализации функции $I(\lambda)$ необходим подбор порогового значения δ , обеспечивающего выбор заданного количества M_1 областей существенного изменения. Однако вычисление пороговых значений δ , обеспечивающих выбор заданного количества M_1 наиболее информативных спектральных каналов, для всего многообразия реализаций функций $I(\lambda)$ и условий многоспектральной съемки является задачей, не имеющей общего решения. Она должна решаться итерационно для заданных $I(\lambda)$ и M_1 .

Заключение

В работе рассмотрено решение задачи обнаружения координат изменения свойств (разладки) случайного процесса в виде правила принятия решения и оценки максимального правдоподобия координаты скачка анализируемой функции, определяемой с использованием алгоритма вычисления кумулятивных сумм. Для представляющего наибольший практический интерес случая определения последовательности скачков предложен модифицированный алгоритм вычисления кумулятивных сумм со скользящим окном, ограничивающим объем выборки, используемой при расчетах спектральных составляющих. Использование скользящего окна позволяет исключить пропуски скачков, дающих противопо-

ложный по знаку вклад в величину решающей статистики. Основными настраиваемыми параметрами алгоритма являются величина порога и объем вспомогательной выборки, определяющий размер скользящего окна.

Разработанный алгоритм вычисления кумулятивных сумм со скользящим окном программно реализован в виде пакета прикладных программ в среде программирования MathLab и применен для анализа функций спектральных обра-

зов наземных объектов и подстилающей поверхности в целях выбора наиболее информативных спектральных каналов многоспектральных систем ДЗЗ. 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.

Литература

1. Жиглявский А. А., Красковский А. Е. Обнаружение разладки случайных процессов в задачах радиотехники. — Л.: Изд-во ЛГУ, 1988. — 224 с.
2. Бассвиль М. и др. Обнаружение изменения свойств сигналов и динамических систем: пер. с англ. — М.: Мир, 1989. — 280 с.
3. Никифоров И. В. Последовательное обнаружение изменения свойств временных рядов. — М.: Наука, 1983. — 200 с.
4. Ширяев А. Н. Статистический последовательный анализ. Оптимальные правила остановки. — М.: Наука, 1976. — 272 с.
5. Ширяев А. Н. Об условно-экстремальных задачах скорейшего обнаружения непредсказуемых моментов у наблюдаемого броуновского движения // Теория вероятностей и ее применения. 2008. Т. 53. Вып. 4. С. 751–768.
6. Репин В. Г., Таргаковский Г. П. Статистический синтез при априорной неопределенности и адаптация информационных систем. — М.: Сов. радио, 1977. — 432 с.
7. Мальцев Г. Н., Козин И. А., Фатеев В. Ф. Методы выбора наиболее информативных спектральных каналов при дистанционном зондировании Земли с малых КА // Изв. вузов. Приборостроение. 2007. № 6. С. 23–31.
8. Мальцев Г. Н., Козин И. А., Данилкин А. П. Космические системы и технологии многоспектрального дистанционного зондирования Земли // Информатика и космос. 2010. № 1. С. 148–158.
9. Дейвис Ш. М. и др. Дистанционное зондирование: количественный подход: пер. с англ. — М.: Недра, 1983. — 415 с.