

COMPLEX OF SOFTWARE PLANETAMONITORING IN APPLIED REMOTE SENSING PROBLEMS

*Vasiliy V. Asmus¹, Alexey A. Buchnev², Vladimir A. Krovotyntsev¹,
Valeriy P. Pyatkin², Gennady I. Salov²*

¹State Research Center «Planeta», Moscow, Russia

²Institute of Computational Mathematics and Mathematical Geophysics SB RAS,
Novosibirsk, Russia

Abstract

Some technologies for processing satellite data in the PlanetaMonitoring software complex, developed jointly by State Research Center «Planeta» and Institute of Computational Mathematics and Mathematical Geophysics SB RAS, are presented. These technologies were used in solving a number of applied problems of the Earth remote sensing.

Keywords: remote sensing, satellite data, software technologies, preprocessing, supervised classification, clustering, statistical detection of linear and circular structures, spatial drifts of nature objects

КОМПЛЕКС ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ PLANETAMONITORING В ПРИКЛАДНЫХ ЗАДАЧАХ ДИСТАНЦИОННОГО ЗОНДИРОВАНИЯ

Асмус В.В.⁽¹⁾, Бучнев А.А.⁽²⁾, Кровотынцев В.А.⁽¹⁾, Пяткин В.П.⁽²⁾, Салов Г.И.⁽²⁾

¹ ФГБУ «НИЦ «Планета», Москва.

² Институт вычислительной математики и математической геофизики СО РАН, Новосибирск

Представлены некоторые технологии обработки спутниковых данных в программном комплексе Planetamonitoring, разрабатываемом совместно ФГБУ «НИЦ «Планета» и ИВМиМГ СО РАН. Эти технологии были использованы при решении ряда прикладных задач дистанционного зондирования Земли.

Ключевые слова: дистанционное зондирование, спутниковые данные, программные технологии, предварительная обработка, классификация с обучением, кластерный анализ, статистическое выделение линейных и кольцевых структур, пространственные перемещения природных объектов.

Введение. Программный комплекс Planetamonitoring, в течение длительного времени разрабатываемый совместными усилиями ФГБУ «НИЦ «Планета» и ИВМиМГ СО РАН, является представительным набором программных технологий, позволяющих решать различные задачи обработки данных дистанционного зондирования Земли (ДЗЗ) [1,2]. Программный комплекс Planetamonitoring реализует технологии предварительной и тематической обработки многоспектральной спутниковой информации оптического, инфракрасного и микроволнового диапазонов. В процессе предварительной обработки спутниковых данных осуществляются яркостные и геометрические преобразования, геокодирование, составление обзорных монтажей и другие. Тематическая обработка многоспектральных спутниковых данных включает технологии распознавания объектов (без обучения и с обучением), выделения и картирования линейных и кольцевых структур, а также пространственного перемещения природных объектов (ледяных полей, водных масс, облачных образований в атмосфере).

1. Предварительная обработка спутниковых данных. Одна из программных технологий традиционна – технология предварительной обработки данных дистанционного зондирования Земли (ДЗЗ), которая включает в себя набор общепринятых операций по яркостным и контрастным преобразованиям изображений. В группу геометрических преобразований включено масштабирование изображений и отображение космических снимков на растровые карты (геокодирование). Трансформирование спутниковых изображений в картографическую основу является геометрическим преобразованием, отображающим весь снимок либо его часть на заранее подготовленную растровую географическую карту соответствующей территории. Карта строится в выбранной географической проекции с использованием различных баз данных (в том числе шейп-файлов ГИС ArcView) для нанесения на лист карты контурных элементов пространственных объектов. Отображение снимка на карту основано на использовании идентичных опорных точек снимка и карты. Возможно использование одного из двух типов отображений: на основе применения семейства кусочно-аффинных преобразований, которые строятся на множестве треугольников снимка и карты, получаемых в результате триангуляции выпуклой оболочки множества опорных точек снимка, и на основе отображающего полинома заданной степени (первой, второй или третьей). Мозаики спутниковых изображений формируются путем объединения трансформированных на единую картографическую основу изображений. Пример использования этой технологии приведен на рис. 1.

Построена первая в мире радиолокационная карта Антарктиды на основе откалиброванных радиолокационных изображений, полученных с ИСЗ «Метеор-М» №1 (БРЛК «Северянин») [3].

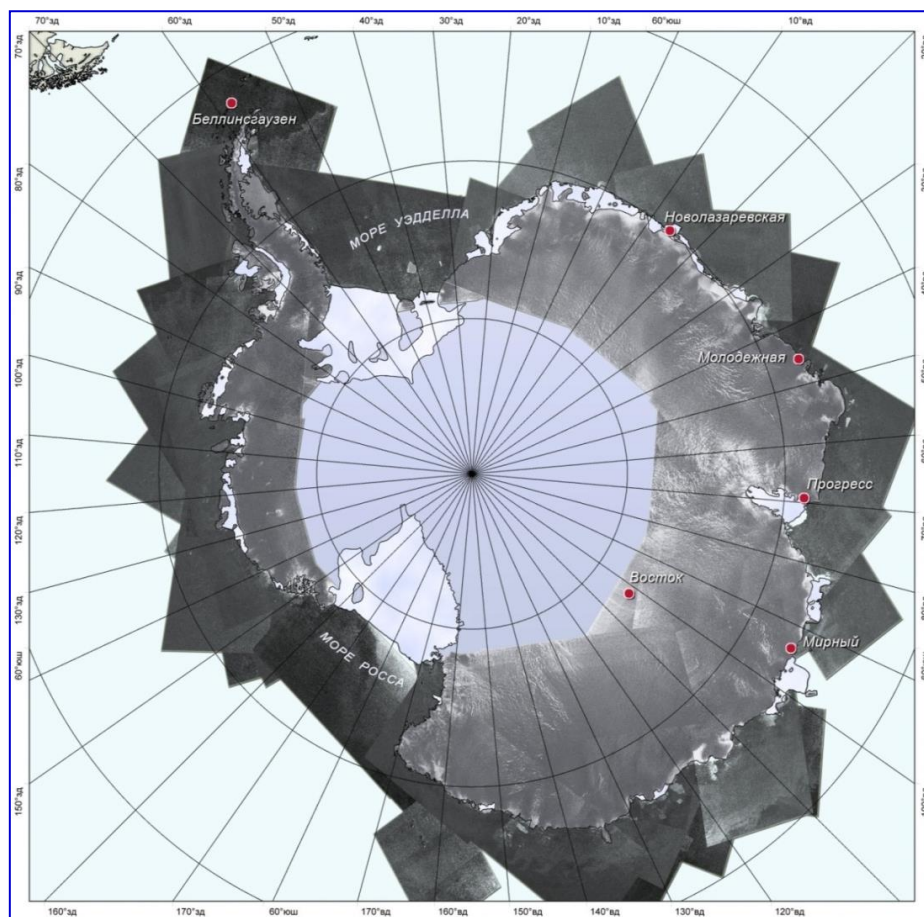


Рис. 1. Цифровая радиолокационная карта Антарктиды.

Программный комплекс включает ряд основанных на видоизменении гистограмм контрастных преобразований. В частности, реализован механизм референтного выравнивания [1], при котором яркостные и контрастные характеристики изображений приводятся либо к характеристикам некоторого изображения, выбранного в качестве референтного (базового), либо к характеристикам гауссово распределения с заданными параметрами. Этот этап является необходимым в тех случаях, когда интересующий исследователя участок земной территории покрывается несколькими разновременными космическими снимками, полученными с разными условиями съемки. Такая ситуация имеет место в задаче обнаружения кольцевых структур на космических снимках, представляющих импактные кратеры на поверхности Земли. На рис. 2 представлен результат выделения импактного кратера Курай (район Горного Алтая), иллюстрирующий выполнение операций программной технологии предварительной обработки данных ДЗЗ. Были использованы разновременные снимки импактного кратера Курай, полученные со спутника SPOT-4 (разрешение 20 м) 08.10.2007 и 14.05.2012 соответственно. Была синтезирована мозаика выровненных по яркости изображений (снимок от 14.05.2012 принят за базовый) и применен непараметрический статистический алгоритм выделения кольцевых структур на космических изображениях (см. ниже).

2. Непараметрический статистический подход к выделению линейных и кольцевых структур. В анализе космических изображений при решении прикладных задач ДЗЗ, в мониторинге природной среды возникает проблема обнаружения протяженных объектов заданной формы, например, линейных и кольцевых структур на случайном фоне. При обработке космических снимков с целью обнаружения на них объектов, представляющих интерес, в силу целого ряда причин предпочтение отдается статистическому подходу [4-6]. Основная причина состоит в том, что вследствие случайного характера природных процессов данные дистанционных измерений (спектросовмещенные изображения) содержат много случайных вариаций,

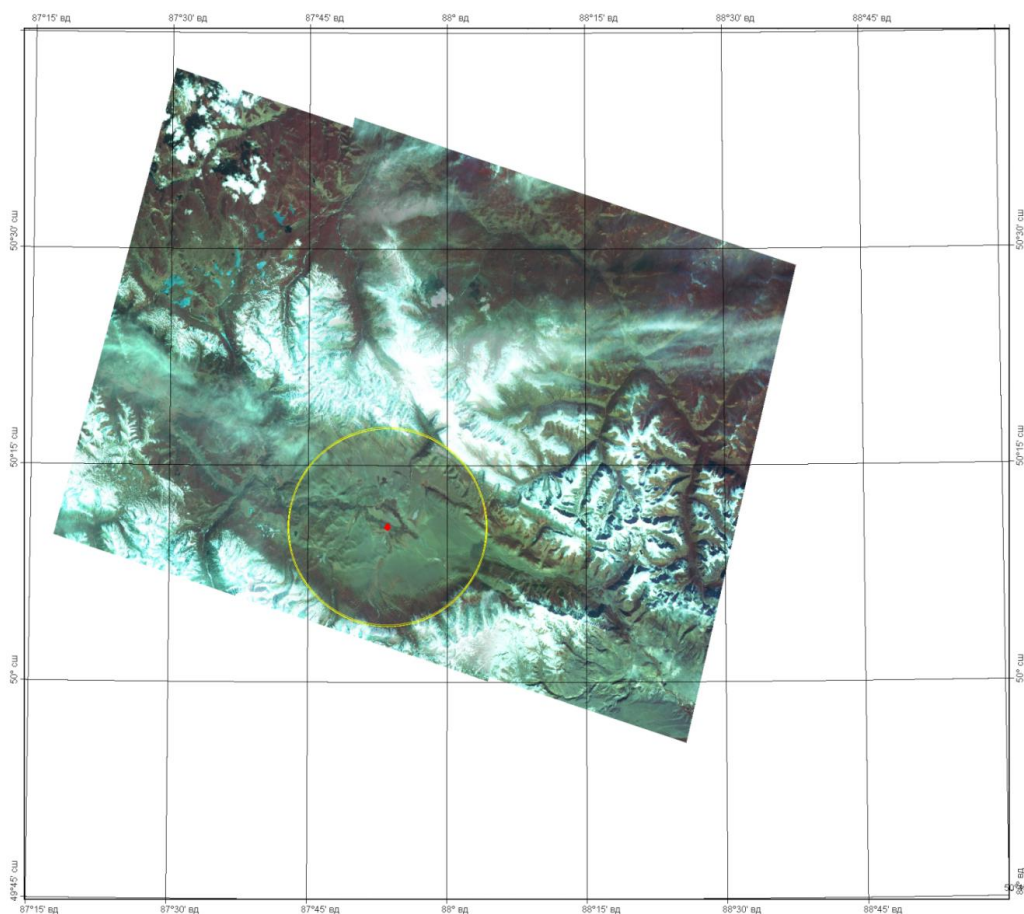


Рис. 2. Результат выделения импактного кратера Курай.

маскирующих различия значений яркости изображения в точках области объекта и в точках области фона. Задачу обнаружения протяженных объектов на зашумленном изображении для большей ясности рассмотрим на примере обнаружения полосы на одном из возможных (проверяемых) положений ее на изображении. Полоса может быть замкнутой, образующей, например, кольцо. В любом случае измерения значений сигнала (наблюдения) удобно выполнять в точках, расположенных на последовательности перпендикуляров к средней линии возможного (проверяемого) положения полосы. Начнем с простейшего случая. Пусть X_i – величина сигнала, наблюдаемая в точке пересечения i -го перпендикуляра и средней линии проверяемой области полосы, а Y_i и Z_i – величины, наблюдаемые также на i -ом перпендикуляре, но в точках области сложного (случайного) фона, расположенных по разные стороны симметрично относительно средней линии проверяемой области полосы, $i = 1, 2, \dots, N$. Если в течение всего периода наблюдений важный объект отсутствовал на проверяемом положении, то при выполнении известных условий, в частности, достаточной удаленности точек наблюдений друг от друга, X_i, Y_i, Z_i можно рассматривать как стохастически независимые случайные величины с одним тем же непрерывным вероятностным законом распределения значений, скажем, $F_i = F_{iX} = F_{iY} = F_{iZ}$, как правило, неизвестным наблюдателю и, быть может, зависящим от i , т.е. возможна неоднородность по длине полосы. Напротив, если в течение всего периода наблюдений объект находился в поле зрения, то величина X_i будет стохастически больше или меньше величин Y_i и Z_i . Без потери общности будем считать, что именно стохастически больше. Таким образом, задача состоит в том, чтобы по наблюдениям X_i, Y_i, Z_i , $i = 1, 2, \dots, N$, распознать (обнаружить) тот случай, где на проверяемом положении важный

объект присутствует. Требуется указать правило (тест), приводящее к верному решению с максимально возможной (или близкой к ней) вероятностью. С точки зрения математической статистики, чтобы свести к минимуму риск принять неверное решение, следует проверить статистическую гипотезу H_0 об однородности величин X_i, Y_i, Z_i при каждом $i = 1, 2, \dots, N$, означающую отсутствие объекта, против альтернативной гипотезы H_1 , о том, что величины X_i стохастически больше величин Y_i и Z_i . Учитывая возможную неоднородность вдоль полосы, введем следующие статистики:

$$\nu^+ = \sum_{i=1}^N I\{X_i > \max(Y_i, Z_i)\}, \quad \nu^- = \sum_{i=1}^N I\{X_i < \min(Y_i, Z_i)\},$$

$$\nu^0 = N - \nu^+ - \nu^-,$$

здесь и далее $I\{A\}$ – функция-индикатор события A , равная 1, если событие A произошло, и 0 в противном случае. Среди всех критериев, основанных на этих статистиках, существует равномерно наиболее мощный, отклоняющий гипотезу H_0 с наибольшей вероятностью, когда верна гипотеза H_1 , имеющий следующий вид

$$\nu^+ > h_0\{\nu_0\}, \quad (1)$$

где критическое значение $h_0 = h_0(z)$ – наименьшее число h такое, что

$$\sum_{i=h+1}^{N-z} \binom{N-z}{i} 2^{-(N-z)} \leq \alpha,$$

здесь α – заранее заданный уровень значимости критерия.

Рассмотрим теперь случай однородности по длине полосы. В этом случае существенно более эффективным будет критерий, основанный на других статистиках:

$$S^+ = \sum_{i=1}^N \sum_{j=1}^N I\{X_i > \max(Y_j, Z_j)\}, \quad S^- = \sum_{i=1}^N \sum_{j=1}^N I\{X_i < \min(Y_j, Z_j)\},$$

$$S^0 = N^2 - S^+ - S^-.$$

Односторонний непараметрический статистический критерий, основанный на этих статистиках, имеет вид

$$S^+ > h_1(S^0) \quad \text{или, что то же,} \quad S^+ - S^- > 2h_1(S^0) + S^0 - N^2. \quad (2)$$

Если положить $h_1(S^0) = (C - S^0)/2$, где C – число, входящее в определение непараметрического статистического критерия Манна-Уитни с двумя выборками, объемов $m = N$ и $n = 2N$, основанного на считающей статике U и имеющего в рассматриваемом случае вид

$$U = \sum_{i=1}^N \sum_{j=1}^N I\{X_i > Y_j\} + \sum_{i=1}^N \sum_{j=1}^N I\{X_i > Z_j\} > C, \quad (3)$$

число C вычисляется по заданному уровню значимости α (но существуют и таблицы), то критерий (2) будет эквивалентен критерию Манна-Уитни. Это следует из равенства (эквивалентности) событий

$$\{S^+ > (C - S^0)/2\} = \{2S^+ + S^0 > C\} = \{U > C\}.$$

С другой стороны, часто может быть получена такая функция h_1 , что критерий (2) будет значимо более мощным, чем критерий Манна-Уитни.

Для дальнейшего увеличения вероятности обнаружения протяженного объекта необходимо увеличивать число наблюдений на каждом перпендикуляре. Достаточно рассмотреть случай, где имеет место неоднородность по длине объекта. Рассмотрим конкретный пример.

Пусть X_{i1}, X_{i2}, X_{i3} – величины, наблюдаемые на i -ом перпендикуляре в точках проверяемой

области объекта, а Y_{i1}, Y_{i2}, Y_{i3} и Z_{i1}, Z_{i2}, Z_{i3} – наблюдаемые на i -ом перпендикуляре в точках проверяемой области фона по разные стороны относительно проверяемой области объекта (полосы). Введем статистики для каждого перпендикуляра

$$S_i^+ = \sum_{u=1}^3 \sum_{v=1}^3 I\{X_{iu} > \max(Y_{iv}, Z_{iv})\}, \quad S_i^- = \sum_{u=1}^3 \sum_{v=1}^3 I\{X_{iu} < \min(Y_{iv}, Z_{iv})\},$$

$$S_i^0 = 9 - S_i^+ - S_i^-$$

и суммарную статистику $\nu = \sum_{i=1}^N I\{S_i^+ > h(S_i^0)\}$, где $h(0) = h(1) = 6$, $h(2) = 5$, $h(3) = h(4) = 4$, $h(5) = 3$, $h(6) = h(7) = 2$, $h(8) = 1$, $h(9) = 0$. Для вероятности при гипотезе H_0 имеет место равенство $p_0 = P\{S_i^+ > h(S_i^0) | H_0\} = 0,190476$ – бесконечная периодическая десятичная дробь. Отсюда гипотезу H_0 следует отклонить, если ν превышает достаточно большое число C , которое следует выбирать так, чтобы

$$\sum_{i=C+1}^N \binom{N}{i} p_0^i (1-p_0)^{N-i} \leq \alpha.$$

Результат использования этого алгоритма представлен на рис. 2.

3. Распознавание и классификация объектов окружающей среды. Центральные вопросы тематической обработки (интерпретации) данных ДЗЗ – вопросы повышения качества дешифрирования – непосредственно связаны с проблемой выбора адекватных алгоритмов распознавания [2,7,8]. Современный опыт автоматизированного распознавания данных ДЗЗ показывает, что заранее практически невозможно установить, какой алгоритм будет лучше с точки зрения точности классификации. Поэтому в распознающую систему целесообразно закладывать несколько алгоритмов и выбор оптимального алгоритма проводить эмпирически. В классификации данных ДЗЗ чаще других используются методы, которые можно разбить на две группы: контролируемая классификация (классификация с обучением) и кластеризация (классификация без обучения).

3.1. Контролируемая классификация (классификация с обучением). В классификации с обучением для распределения векторов признаков по представляющим интерес классам используется процесс обучения классификатора способности различать эти классы на основе репрезентативных выборок представителей классов. Эти выборки называются обучающими. Процесс обучения фактически является процессом построения функций правдоподобия (или решающих функций) для классов. Эти функции будут затем использоваться для классификации каждого вектора признаков как принадлежащему только одному классу (жесткая классификация) либо нескольким классам (нечеткая классификация). Обучение классификатора может быть проконтролировано путем определения вероятности правильной классификации (для этого часть из обучающих выборок переводится в разряд контрольных). Контролируемая классификация в программном комплексе PlanetaMonitoring основана на использовании байесовской стратегии максимального правдоподобия для нормально распределенных векторов признаков. Пусть x – N -мерный вектор признаков $x = (x_1, \dots, x_N)^T$, где N – число спектральных диапазонов. Предполагается, что векторы X имеют в классе ω_i нормальное распределение $N(m_i, B_i)$ со средним m_i и ковариационной матрицей B_i . В этом случае байесовская стратегия максимального правдоподобия для поэлементного классификатора формулируется следующим образом [2].

Пусть $\Omega = (\omega_1, \dots, \omega_m)$ – конечное множество классов, $p(\omega_i)$ – априорная вероятность класса ω_i . Тогда решающая функция класса ω_i имеет вид

$$g_i(x) = \ln(p(\omega_i)) - 0.5 \ln(|B_i|) - 0.5(x - m_i)^T B_i^{-1}(x - m_i).$$

Классическое решающее правило для жесткой классификации принимает следующий вид: вектор X заносится в класс ω_i , если $g_i(x) > g_j(x)$ для всех $j \neq i$.

Поскольку физические размеры реально сканируемых пространственных объектов, как правило, больше разрешения съемочных систем, между векторами признаков существуют взаимосвязи [2]. Использование информации подобного рода дает возможность повысить точность классификации, если пытаться распознавать одновременно блок смежных векторов квадратной или крестообразной формы. Будем называть такой блок векторов объектом. Решение об отнесении центрального элемента объекта тому или иному классу принимается на основе результата классификации всего объекта. Такой подход порождает целое семейство решающих правил. Во-первых, это использование принципа голосования, т.е. независимая классификация элементов объекта и отнесение центрального элемента к тому классу, которому было отнесено большинство элементов объекта. Во-вторых, это применение текстурных операторов (простейший пример – описание объекта X через вектор средних составляющих его элементов) с последующим отнесением центрального элемента классу, к которому был отнесен параметр, характеризующий X . В-третьих, описание объекта X случайным марковским полем.

Проведена адаптация алгоритма к гиперспектральным данным ДЗЗ. Выполнена реализация алгоритма спектрально-углового картирования [9]. В качестве ссылочных векторов (представителей классов) при этом используются векторы средних классов.

Необходимые для построения решающих функций классов оценки статистических характеристик – векторов средних и ковариационных матриц, коэффициентов пространственной корреляции между значениями координат соседних векторов в горизонтальном и вертикальном направлениях – определяются на основе векторов из обучающих выборок (полей). Априорные вероятности классов могут быть взяты из результатов кластеризации.

3.2. Неконтролируемая классификация (кластеризация). В состав программного комплекса PlanetaMonitoring входит реализация классического алгоритма жесткой кластеризации – алгоритма K -средних, широко используемого для разбиения на кластеры больших объемов многомерных данных [10]. Алгоритм K -средних может быть отнесен к классу параметрических, т.к. он неявным образом предполагает природу плотности вероятности: кластеры стремятся иметь конкретную геометрическую форму, зависящую от выбранной метрики. Используются следующие метрики: Евклидова, Махаланобиса, Чебышева, city-block расстояние. Известно также, что результат кластеризации методом K -средних зависит от задания начальных центров кластеров. Предоставляется выбор одного из трех вариантов, два из которых определяются на основе статистических характеристик набора данных и один основан на случайной выборке. В последнее время нами выполнена адаптация алгоритма K -средних к данным гиперспектральных сканеров.

Альтернативой жесткой разделяющей кластеризации является мягкая или нечеткая кластеризация, разрешающая векторам принадлежать всем кластерам с коэффициентом членства $u_{ij} \in [0,1]$, определяющим степень принадлежности j -го вектора i -му кластеру. В состав системы кластеризации программного комплекса нами была включена реализация следующих алгоритмов нечеткой кластеризации: метод S -средних [11], метод S -средних с регуляризацией, расширенных алгоритмов S -средних и Густафсона-Кесселя.

Все упомянутые алгоритмы классификации (как с обучением, так и автоматической) реализованы в системе параллельного программирования OpenMP под управлением ОС Windows.

4. Определение пространственных перемещений объектов по разновременным спутниковым изображениям. Одним из способов определения пространственных перемещений природных объектов по разновременным спутниковым изображениям является метод, ос-

нованный на нахождении максимумов коэффициента взаимной корреляции [12]. В [13] аналогичный подход рассматривается в качестве метода распознавания образов, известного как корреляционное сопоставление. В обоих случаях корреляция используется как средство поиска эквивалентов объекта-эталона, представленного в виде изображения $w(x, y)$ размерами $J \times K$, на изображении $f(x, y)$ размерами $M \times N$; предполагается, что $J \leq M$ и $K \leq N$. Коэффициент взаимной корреляции

$$\gamma(x, y) = \frac{\sum_s \sum_t [f(x+s, y+t) - f_m(x, y)][w(s, t) - w_m]}{\sigma_w \sigma_f}. \quad (4)$$

Здесь w_m – среднее значение пикселей в эталоне w ; f_m – среднее значение элементов изображения f в области, покрываемой эталоном. Знаменатель в (4) является произведением стандартного отклонения σ_w пикселей эталона w на стандартное отклонение σ_f пикселей изображения f в области, покрываемой эталоном. Согласно [14] поиск позиций найденных эталонов (определение смещений) на следующем изображении серии может быть реализован одним из трех методов: определением максимума коэффициента взаимной корреляции в пространственной области, определением максимума коэффициента взаимной корреляции в частотной области на основе быстрого преобразования Фурье и нахождением минимума суммы квадратов расстояний. При этом не предполагается при поиске смещений каких-либо преобразований эталона за исключением преобразования переноса. В представляемой работе определение смещений эталонов производится на основе определения максимума коэффициента взаимной корреляции в пространственной области в соответствии с формулой (4). При этом эталон может подвергаться преобразованию, состоящему из масштабирования, поворота и переноса. Реализован соответствующий алгоритм и получены результаты вычислительных экспериментов на изображениях, полученных с КА «МЕТЕОСАТ-8» [15]. Эти результаты свидетельствуют как о необходимости учета масштабирования и поворота эталона, так и о приемлемом времени соответствующих вычислений. Решение задачи состоит из следующих основных шагов: 1) поиск на текущем изображении квадратных объектов-эталонов, основанный на достижении максимума контраста либо максимума энтропии; 2) поиск позиций найденных эталонов на последующем изображении, основанный на достижении максимального значения коэффициента взаимной корреляции; 3) построение векторных полей пространственных перемещений объектов в соответствии с найденными позициями. На рис.3 представлено поле перемещений ледяных полей в Азовском море в период с 5 по 6 февраля 2012 года с использованием предложенной технологии определения перемещений природных объектов по разновременным многоспектральным спутниковым изображениям.

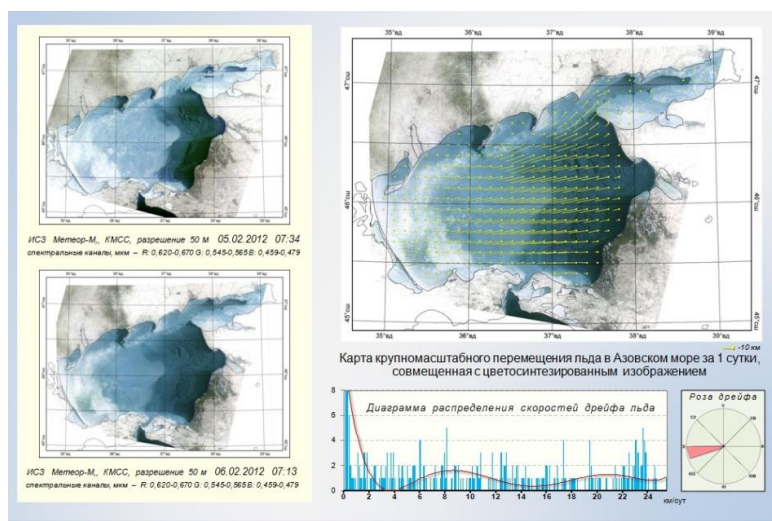


Рис. 3. Дрейф льда в Азовском море (по данным ИСЗ «Метеор-М» №1, КМСС).

Заключение. Многолетний успешный опыт использования описанного выше программного комплекса обработки спутниковых данных в различных прикладных задачах ДЗЗ подтверждает высокую эффективность реализованных в комплексе алгоритмов обработки данных ДЗЗ.

Работа выполнена частично при финансовой поддержке Российского фонда фундаментальных исследований (проект № 16-07-00066) и Программы I.33П фундаментальных исследований Президиума РАН (проект № 0315-2015-0012).

ЛИТЕРАТУРА

- [1] Шовенгердт Р.А. Дистанционное зондирование. Модели и методы обработки изображений. М.-Техносфера, 2010. – 560 с.
- [2] Асмус В.В. Программно-аппаратный комплекс обработки спутниковых данных и его применение для задач гидрометеорологии и мониторинга природной среды. Диссертация в виде научного доклада на соискание ученой степени доктора физико-математических наук. На правах рукописи. Москва – 2002. – 75 с.
- [3] Кровотынцев В.А., Тренина И.С., Волгутов Р.В., Максимов А.А., Маслова Н.А. Информационная продукция спутникового мониторинга полярных акваторий Земли и замерзающих морей России // Метеоспектр. 2014. № 2, С. 89-98.
- [4] Салов Г.И. О мощности непараметрических критериев для обнаружения протяженных объектов на случайном фоне // Автометрия. 1997. № 3, С. 60-75.
- [5] Салов Г.И. Новый статистический критерий для задач с двумя и тремя выборками, более мощный, чем критерии Вилкоксона и Уитни // Автометрия. 2011. № 4, С. 58-70.
- [6] Asmus V.V., Buchnev A.A., Pyatkin V.P., Salov G.I. Software System for Satellite Data Processing of Applied Tasks in Remote Sensing of the Earth // Patt. Recogn. and Image Anal. 2009. Vol. 19, N 3, P. 69-74.
- [7] Асмус В.В., Бучнев А.А., Пяткин В.П. Контролируемая классификация данных дистанционного зондирования Земли // Автометрия. 2008. № 4, С. 60-67.
- [8] Асмус В.В., Бучнев А.А., Пяткин В.П. Кластерный анализ данных дистанционного зондирования Земли // Автометрия. 2010. № 2, С. 58-66.
- [9] Kruse, F.A., Lefkoff A.B., Boardman J.W. et al. The Spectral Image Processing System (SIPS) – interactive visualization and analysis of imaging spectrometer data // Remote Sensing of Environment. 1993. N 44. P. 145-163.
- [10] Jain A.K. Data clustering: 50 years beyond K-means // Patt. Recogn. Lett. 2010. Vol. 31, P. 651-666.
- [11] Bezdek J.C., Ehrlich R., Full W. FCM: The fuzzy c-means clustering algorithm // Computers & Geosciences. 1984. Vol. 10, N 2, P. 191-203.
- [12] Jähne B. Digital Image Processing. Berlin, Heidelberg: Springer-Verlag, 2005. 607 p.
- [13] Гонсалес Р., Вудс Р. Цифровая обработка изображений. Пер. с англ. под ред. П.А. Чочиа. М.: Техносфера, 2005. 1072 с.
- [14] MSG Meteorological Products Extraction Facility. Algorithm Specification Document. Doc. N EUM/MSG/SPE/022. Is. 2.6. 1 June 2004.
- [15] Бучнев А.А., Пяткин В.П. Мониторинг облачных образований по данным геостационарных спутников Земли // Автометрия. 2009. № 4, С. 40-47.