ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

Е.А. Чернецова, А.Д. Шишкин

ОПРЕДЕЛЕНИЕ РАЗМЕРА БАНКА ДАННЫХ ИЗОБРАЖЕНИЙ, ДОСТАТОЧНОГО ДЛЯ РАБОТЫ АВТОМАТИЧЕСКОГО КЛАССИФИКАТОРА ОБРАЗОВ НЕФТЯНОГО ПЯТНА И ВЕТРОВОГО СЛИКА

E.A. Chernetsova, A.D. Shishkin

DEFINITION OF THE DATA SIZE OF THE BANK OF IMAGES, SUFFICIENT FOR PERFORMANCE OF AUTOMATIC CLASSIFIER OF IMAGES OF THE OIL SPILL AND WIND SLICK

Рассматривается проблема классификации аномалий на морской поверхности на примере радиолокационных изображений нефтяного пятна и ветрового слика. В качестве признаков объектов рассматриваются ряды параметров распределений, аппроксимирующих гистограммы изображений и ряды интенсивности пикселей изображений. Рассмотрен последовательный алгоритм классификации для определения необходимого количества образцов объектов в банке данных.

Ключевые слова: морские аномалии, радиолокационные изображения, классификация объектов, решающее правило, охрана окружающей среды, непараметрический алгоритм.

The problem of classification of anomalies in the sea surface by the example of radar images of oil spills and wind slicks is examined. As the attributes of the objects the ranks of the distribution parameters, approximating the histogram of images and rows of pixel intensity images are considered. The sequence of the classification algorithm is examined to determine the required number of samples of objects in a database.

Key words: marine anomalies, radar images, classification of objects, the decision rule, environment protection, nonparamethric algorithm.

Ведение

При дистанционном зондировании морских акваторий данные представляют собой монохромные радиолокационные изображения (РЛИ), которые используются для обнаружения и классификации выделенных аномалий. Эти данные отправляются на хранение в банк данных, либо претерпевают в системе некоторую обработку и затем передаются потребителю в виде, удобным для принятия решений. Решение задачи классификации базируется на современных алгоритмах, вычислительных и программных средствах.

На рис. 1 приведены примеры обрабатываемых РЛИ [6].

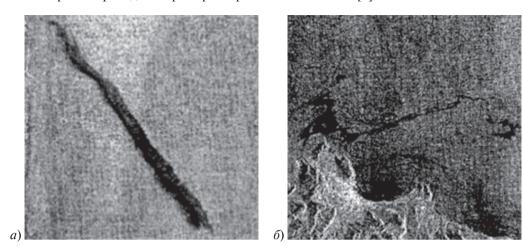


Рис. 1. РЛИ нефтяного пятна (a) и ветрового слика (δ)

Для автоматизации решения задачи распознавания нефтяного пятна на морской поверхности, и отличения его от ветрового слика необходимо иметь банк данных РЛИ, по которым производится оценка их параметров.

В работе [2] обсуждалось применение для целей распознавания образов нефтяного пятна и ветрового слика построенных гистограмм оттенков серого цвета на выделенном фрагменте РЛИ. Считаем, что на РЛИ проведена фрагментация выделенных участков и построены гистограммы распределений яркости. В дальнейшем предполагалась аппроксимация этих гистограмм законами распределения вероятностей с определением параметров этих распределений. Однако исследования показали, что, поскольку изображения являются двумерными, законы распределения вероятностей интенсивности пикселей на изображениях нефтяного пятна и ветрового слика могут, как отличаться друг от друга по параметрам, так и быть похожими друг на друга. На рис. 2 и 3 показаны расчетные и аппроксимирующие их по χ^2 плотности распределения вероятностей (ПРВ) пикселей пятна и слика по горизонтальной координате выделенного фрагмента РЛИ. На рис. 4 и 5 показаны аналогичные ПРВ по вертикальной координате. Поскольку реальные ПРВ не известны, то представляется наиболее целесообразным создание непараметрического автоматизированного классификатора рассматриваемых образов.

Теоретический анализ

Задача исследования состоит в определении минимального количества изображений каждого объекта (нефтяного пятна или ветрового слика), достаточного для работы автоматизированного классификатора объектов, если для работы классификатора используются: а) ряды параметров распределений, аппроксимирующих гистограммы изображений; б) ряды интенсивности пикселей изображений.

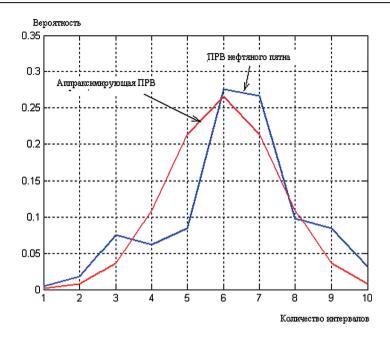


Рис. 2. Сравнение аппроксимирующей ПРВ и ПРВ нефтяного пятна по горизонтали

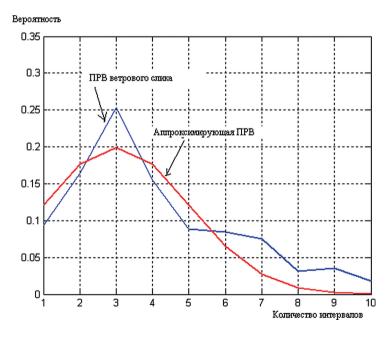


Рис. 3. Сравнение аппроксимирующей ПРВ и ПРВ ветрового слика по горизонтали

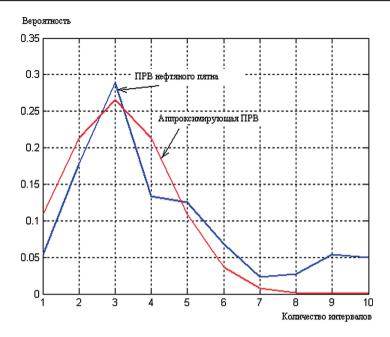


Рис. 4. Сравнение аппроксимирующей ПРВ и ПРВ нефтяного пятна по вертикали

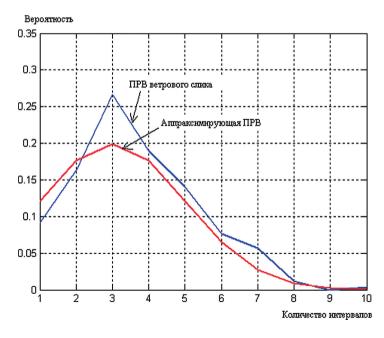


Рис. 5. Сравнение аппроксимирующей ПРВ и ПРВ ветрового слика по вертикали

В задачах непараметрического оценивания часто делается предположение альтернатив Лемана. В данном исследовании применяется оценка Ходжеса-Лемана, относительно которой было установлено, что она мало отличается от широко распространенной для решения подобных задач статистики Вилкоксона, но уменьшает количество необходимых вычислительных затрат [1].

Оценка сдвига распределений Ходжеса-Лемана. представляет собой медиану всех возможных пар разностей элементов одной и другой групп:

$$r_{ii} = med(x_{ui} - x_{vi}); \quad u = 1, ... n_i; \quad v = 1, ... n_i.$$
 (1)

где x_{ui} и x_{vi} — параметры пикселей по координате i и j.

Положительное свойство медианы (1) состоит в том, что насколько первая группа «больше» второй, настолько вторая «меньше» первой, то есть: $r_{ii} = -r_{ii}$.

Существенным недостатком медианы Ходжеса-Лемана является нетранзитивность. Если x больше y на a, а s больше y на b, то желательно, чтобы s было больше x на величину a+b. Медиана (1) этим свойством не обладает [3].

Однако, поскольку в нашем случае имеется только две альтернативы, то статистику Ходжеса-Лемана использовать правомерно.

В качестве решающей процедуры используем последовательный критерий отношения вероятностей Вальда (п. к. о. в.) [4, 5], основанный на последовательных рангах, расставив замеры сравниваемых выборок так, чтобы они чередовались: $x_1, y_1, x_2, y_2, ..., x_n, y_n$. Обозначим объединенные замеры на k-ом шаге вектором $V(k) = [v_1, v_2, ..., v_k]$, где $v_1 = x_1, v_2 = y_1$ и т.д. Пусть $S(k) = [S_1, S_2, ..., S_k]$ есть вектор последовательных рангов для V(k), а

$$\lambda_k = \frac{P_k(S(k)/H_1)}{P_k(S(k)/H_0)} \tag{2}$$

представляет собой последовательное отношение вероятностей на k-ом шаге процесса [6]. Если верна гипотеза H_0 , то для произвольного вектора S из S(k) имеем $P_k(S(k) = S/H_0) = 1/k!$ и, следовательно, можно вычислить $P_k(S(k) = S/H_1)$, учитывая, что каждый из полученных векторов S соответствует однозначным образом определенному порядку объединенных замеров x_i и y_i .

Таким образом, достаточно вычислить условную вероятность вариационного ряда

$$P(v_1 \le v_2 \le \dots \le v_k/H_1) = \int_{-\infty < t_1 \le t_2 \le \dots \le t_k < \infty} \dots \int \prod_{t=1}^k df_i(P(t_i)), \tag{3}$$

где $f_i(P(t_i)) = P(t_i)$, когда v_i есть x и $f_i(P(t_i)) = f_i(P(t_i))$, когда v_i есть y.

В случае альтернатив Лемана имеем две гипотезы.

Гипотеза о наличии на изображении образа нефтяного пятна

$$H_0: G = P(X),$$

против гипотезы о наличии на изображении ветрового слика

$$H_1: G = f(P(X)) = P^r(X); r > 0.$$

При этом последовательное отношение вероятностей (2) с учетом (3) на k-ом шаге будет

$$\lambda_{k} = \frac{P_{k}(S(k)/H_{1})}{P_{k}(S(k)/H_{0})} = \frac{k!r^{k/2}}{\prod_{i=1}^{k} \left(\sum_{j=1}^{i} A_{j}\right)}$$
 для четных значений k (4)

И

$$\lambda_{k} = \frac{P_{k}(S(k)/H_{1})}{P_{k}(S(k)/H_{0})} = \frac{k!r^{(k-1)/2}}{\prod_{i=1}^{k} \left(\sum_{j=1}^{i} A_{j}\right)}$$
 для нечетных значений k , (5)

где
$$A_j = \begin{cases} 1; & \text{если } v_j \text{ есть } x, \\ r; & \text{если } v_j \text{ есть } y. \end{cases}$$

Таким образом, непараметрическая процедура определения количества изображений в банке данных, необходимого для распознавания образов нефтяного пятна или ветрового слика с использованием последовательного критерия отношения вероятностей Вальда сводится к следующим шагам:

Шаг 1: получение последовательного ранга (k+1)-й выборки параметра классификации.

Шаг 2: Образовать вектор A(k + 1) из A(k) и S_{k+1} ...

Шаг 3. Вычислить последовательное отношение вероятностей по формуле (4) или (5) и сравнить с останавливающими границами. Решение принимается в пользу той или иной гипотезы, если выполнится одно из двух условий $\lambda_k \geqslant P_{\text{верх.}}$ или $\lambda_k \leqslant P_{\text{нижн.}}$, иначе измерения будут продолжены. Останавливающие границы (пороги) в п.к.о.в. Вальда приближенно равны

$$P_{ ext{верх.}} = \frac{1 - e_{01}}{e_{10}}, \quad P_{ ext{нижн.}} = \frac{e_{01}}{1 - e_{10}},$$

где e_{ij} — вероятность принятия гипотезы H_i , тогда как в действительности верна гипотеза H_j ; i,j=0,1. Для решения рассматриваемой задачи примем $e_{10}=e_{01}=0,1$. Тогда пороги п.к.о.в. Вальда будут равны $P_{\text{верх.}}=9,0$ и $P_{\text{нижн.}}=0,11$.

А связь между средним числом измерений E(k) и параметром r приближенно описывается зависимостью [6]

$$E_r(k) \cong \frac{\log[(1-e_{10})/e_{01}]}{\log\frac{1}{2}(r^{-1/2}+r^{1/2})}.$$

Компьютерное моделирование непараметрической процедуры п.к.о.в. Вальда

С целью проверки выдвинутых предположений был выполнен компьютерный эксперимент по моделированию описанного выше пошагового последовательного алгоритма. Определялось необходимое количество образцов РЛИ при двух видах исходных параметров: а) по параметрам аппроксимирующих распределений; б) по интенсивности пикселей изображений по вертикали и горизонтали. Результаты представлены на рис. 6, 7. На рисунках жирной линией выделен нижний порог, по которому определялось количество изображений нефтяного пятна и ветрового слика в банке данных, достаточных для корректной работы автоматического классификатора изображений. Расчеты проводились по параметрам аппроксимирующих распределений и интенсивности пикселей изображений по вертикали и горизонтали.

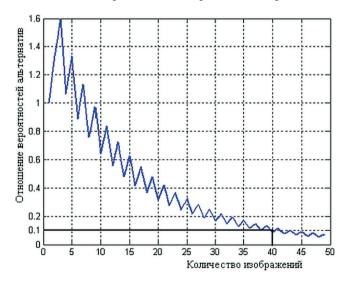


Рис. 6. Определение количества изображений нефтяного пятна и ветрового слика в банке данных, достаточное для корректной работы автоматического классификатора изображений по параметрам аппроксимирующих распределений

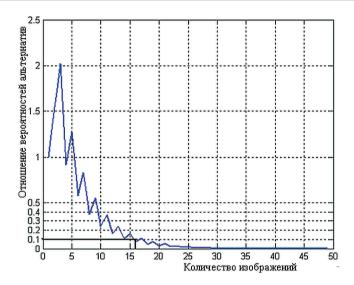


Рис. 7. Определение размера банка изображений, достаточного для работы автоматического классификатора нефтяного пятна и ветрового слика при обработке интенсивности пикселей изображений по вертикали

В таблице представлены результаты компьютерного моделирования п.к.о.в. Вальда для определения объема банка данных, необходимого для работы автоматического классификатора образов нефтяного пятна и ветрового слика на изображении морской поверхности.

Параметр классификации	Необходимое количество изображений, которое должно содержаться в банке данных
Параметры плотностей распределения вероятностей интенсивности пикселей изображения (различие по критерию χ^2 образцового и оцениваемого распределения $< 5 \%$)	40
Интенсивность пикселей изображения, обрабатываемого по горизонтали	16
Интенсивность пикселей изображения, обрабатываемого по вертикали	15

Выводы

Моделирование показало, что использование для работы автоматического классификатора образов нефтяного пятна и ветрового слика на изображении морской поверхности в качестве параметра интенсивности пикселей позволит снизить требования к объему банка данных по сравнению с использованием для той же целей аппроксимации гистограмм оттенков серого известными законами распределения вероятностей.

Литература

- 1. *Роенко А.А., Лукин В.В., Зеленский А.А.* Определение параметра сдвига выборки данных с симметричным негауссовым распределением на основе использования методов адаптивного робастного оценивания. // Радіоелектронні і комп'ютерні системи, 2005, № 2(10), с. 78—87.
- 2. Чернецова Е.А. Дистанционный мониторинг нефтяных загрязнений в водной среде. Lap Lambert Academic Publishing, 2011. 132 p.
- 3. *Шеломовский В.В.* Электронный учебник по дисциплине «Математическая статистика». Мурманский федеральный государственный педагогический университет. Электронный ресурс: [http://www.exponenta.ru/educat/systemat/shelomovsky/lab/lab07.asp].
- Шишкин А.Д., Чернецова Е.А. Классификация пространственно-распределенных объектов на основе использования последовательной выборки измерений. // Учёные записки РГГМУ, 2014, № 33, с. 155–163.
- 5. Φ у K. Последовательные методы в распознавания образов и обучении машин. Пер. с англ.. М.: «Наука», 1971. 256 с.
- Del Frate et al. Neural Networks for Oil Spill Detection Using ERS-SAR Data. // IEEE Transactions on Geoscience and Remote Sensing, 2000, vol. 38, no. 5, pp. 2282–2287.