

УДК 502:330.15

В. І. Лялько¹, М. О. Попов¹, О. Ю. Петроченко²,
О. Д. Рябоконеко², О. І. Сахацький¹, О. Г. Харечко²

¹Центр аерокосмічних досліджень Землі Інституту геологічних наук Національної академії наук України, Київ

²Український центр менеджменту землі та ресурсів, Київ

**Метод класифікації стану лісів
за матеріалами аерокосмічної зйомки
на основі принципу злиття даних**

Розглянуто методичний підхід до класифікації стану лісових ресурсів за багатоспектральними космічними зображеннями. Підхід у своїй теоретичній частині базується на синергетичній концепції злиття даних, його програмна реалізація передбачає використання елементів геоінформаційних технологій. Розроблено алгоритм класифікації, оснований на запропонованому критерії інформативності спектральних каналів і класифікаційному правилі визначення належності об'єкта до відповідного класу. На базі описаного алгоритму розроблено програмний модуль для класифікації багатоспектральних космічних зображень. Описано результати експериментальних досліджень запропонованого методу.

Як відомо, одним з ефективніших методів дослідження лісів є метод, оснований на використанні матеріалів багатоспектрального знімання поверхні Землі з космосу. Використання багатоспектральних знімків для вивчення рослинності базується на аналізі варіацій параметрів відбитого сонячного випромінювання в різних діапазонах спектру електромагнітних хвиль [2]. Для оброблення космічних зображень в останні часи залучають програмні системи ERDAS Imagine, PCI Geomatic, ErMapper та ін. [6]. Кожна така прикладна програмна система є високотехнологічним продуктом, але його повноцінне використання неможливе без потужної апаратної бази і високої кваліфікації фахівця-інтерпретатора даних.

Тому актуальною залишається проблема створення алгоритмів та програм, орієнтованих на вирішення тематичних задач, які можуть бути реалізовані на ПЕОМ типової конфігурації, і робота з якими потребує мінімальної комп'ютерної підготовленості користувача.

Нижче розглядається методичний підхід до вирішення однієї з таких задач, а саме до класифікування стану лісових ресурсів за багатоспек-

тральними космічними зображеннями. Підхід базується на синергетичній концепції злиття даних [3, 5], його програмна реалізація передбачає використання елементів геоінформаційних технологій [1].

Нехай S — область земної поверхні з лісами, стан яких підлягає вивченню методом дистанційного зондування Землі (ДЗЗ). Область S розкладається на сукупність елементарних однакових за формою та розміром ділянок місцевості $\{S_i / S_i \in S; i = 1, 2, \dots\}$; при цьому $S_i \cap S_k = \emptyset, (i \neq k); \cup_{i=1,2,\dots} S_i = S$.

Передбачається, що кожна ділянка місцевості S_i може належати до того чи іншого класу $K_m, m = 1, 2, \dots, M$, де M — загальна кількість класів. Метою вивчення кожної окремої ділянки є її класифікація, тобто визначення, до якого класу вона належить.

При використанні системи ДЗЗ, що має N спектральних каналів, для кожної ділянки S_i земної поверхні може бути отримано до N зональних зображень. Якщо прийняти, що елементарній ділянці місцевості S_i відповідає один піксел на зображенні, то задача вивчення ділянок зводиться до класифікування пікселів (їх також будемо позначати як s_i). При заданій геометричній формі пікселя

повний опис складається з набору таких даних, як ім'я багатоспектрального зображення, місцеположення у площині зображення, амплітуда (або інтенсивність чи оптична щільність). Таким чином, на багатоспектральному зображенні кожний піксел буде мати повний опис у вигляді кортежу

$$\langle \#, x, y, \pi_1, \pi_2, \dots, \pi_N \rangle, \quad (1)$$

де $\#$ — номер багатоспектрального зображення; x, y — координати піксела; $\pi_1, \pi_2, \dots, \pi_N$ — амплітуди ідентичних (за координатами) пікселів на відповідних спектральних зображеннях загальною кількістю N , що разом складають багатоспектральне зображення з номером $\#$.

Задача класифікування піксельних об'єктів ставиться таким чином. Кожний піксел s_i належить одному з K_m класів. Для кожного класу K_m ($m = 1, 2, \dots, M$) надана інформація I_m у вигляді навчальної вибірки $\langle v_m^+ \rangle$ множини позитивних прикладів $\langle v_m^+ \rangle \subseteq K_m$, що загалом утворюють загальну навчальну вибірку I :

$$I = I_1 \cup I_2 \cup \dots \cup I_M. \quad (2)$$

Необхідно: 1) сформулювати критерій інформативності спектральних каналів; 2) розробити правило, яке дозволить визначити для будь-якого об'єкта $s_i \in S$ ($i = 1, 2, \dots$) достовірність μ_m його належності до класу K_m , тобто отримувати пару $\langle K_m, \mu_m \rangle$.

При оцінці інформативності спектральних каналів будемо вважати, що відома навчальна вибірка для кожного з K_m класів. Кожна кількісна ознака π_n ($n = 1, 2, \dots, N$) у відповідному спектральному каналі має певний діапазон значень; позначимо його d_{π_n} . У межах загальної навчальної вибірки I діапазон значень n -ї ознаки розраховується як

$$d_{\pi_n} = \max_{\bar{I}}(\pi_n) - \min_{\bar{I}}(\pi_n), \quad (3)$$

де $\max_{\bar{I}}(\pi_n)$ та $\min_{\bar{I}}(\pi_n)$ — найбільше та найменше значення цієї ознаки.

Для дослідження кореляції значень ознаки π_n з класами K_1, K_2, \dots, K_M розподілимо діапазон значень ознаки на однакові за шириною інтервали. Ширина інтервалу визначається як

$$\Delta d_{\pi_n} = \frac{d_{\pi_n}}{|\bar{I}|},$$

де $|\bar{I}|$ — потужність загальної навчальної вибірки у теоретико-множинному розумінні.

Клас K_m будемо називати зіставленим з певним інтервалом ознаки π_n , яка розглядається, якщо в навчальній вибірці I_m позитивних прикладів по класу K_m матимемо приклад об'єкта, значення ознаки якого потрапляє до цього інтервалу.

Якщо кількісна ознака для об'єктів різних класів в кожній з N спектральних зон розподілена на відповідні інтервали, то, використовуючи загальну навчальну вибірку, для будь-якої спектральної зони можна визначитися, в який інтервал попали конкретні представники різних класів з цієї вибірки, а потім підрахувати кількість попадань класів у відповідні інтервали. Зрозуміло, розподіл попадань класів по інтервалах буде відрізнятися для різних спектральних каналів, а це означає, що різні спектральні канали неоднаково ефективні при класифікуванні об'єктів.

Впровадимо функцію F , яку будемо далі називати критеріальною функцією інформативності спектрального каналу:

$$F = 1 - \frac{1}{M(M-1)} \cdot \sum_{m=1}^M \frac{\sum_{j=1}^{nz} \left(l_{mj} \cdot \sum_{\substack{k=1 \\ k \neq m}}^M l_{kj} \right)}{\sum_{j=1}^{nz} l_{mj}}, \quad (5)$$

де M — кількість класів; nz — число інтервалів для ознаки спектрального каналу, що розглядається; l_{mj} — індикатор зіставлення,

$$l_{mj} = \begin{cases} 1, & \text{якщо клас } k_m \text{ зіставлений з інтервалом } j, \\ 0, & \text{в іншому випадку.} \end{cases}$$

Можна показати, що значення функції інформативності (5) лежать в межах від 0 до 1 і залежать від того, як зіставлені між собою класи та ознаки. Серед спектральних каналів той буде характеризуватися найбільшим значенням функції F , для ознаки якого кількість класів, одночасно зіставлених з кожним інтервалом, мінімальна. Будемо далі користуватися гіпотезою, що саме той спектральний канал i є найінформативнішим; тоді формула (5) надає можливість кількісно порівнювати інформативності спектральних каналів та відбирати серед них найефективніші для класифікування об'єктів.

Введемо оцінку μ , яку далі будемо називати достовірністю належності об'єкта класу. Оцінки μ обчислюються для кожного інтервалу кожної ознаки по усіх класах за правилом, описаним в [4].

Класифікаційне правило містить поінтервальне подання кожної зі спектральних ознак π_1, π_2, \dots ,

π_N , при цьому для кожної ознаки вказуються номер n спектрального каналу ($n = 1, 2, \dots, N$) та інформація про розподіл класів по усіх інтервалах ознаки.

Для кожного інтервалу вказуються:

- 1) його верхня та нижня границі;
- 2) кількість прикладів із загальної навчальної вибірки, які потрапили до даного інтервалу по кожному класу K_m ;
- 3) достовірності належності об'єкта кожному з

класів. Всі перелічені дані формуються шляхом оброблення навчальної вибірки і розміщуються в створюваній інформаційній базі.

Розглянемо, використовуючи цю базу, механізм застосування класифікаційного правила R до конкретного піксельного об'єкту s_i . Об'єкт на багатоспектральному зображенні за номером $\#$ може бути заданий деякими значеннями ознак по спектральних каналах. При класифікуванні здійснюється накладання значень ознак об'єкта на інтервальне

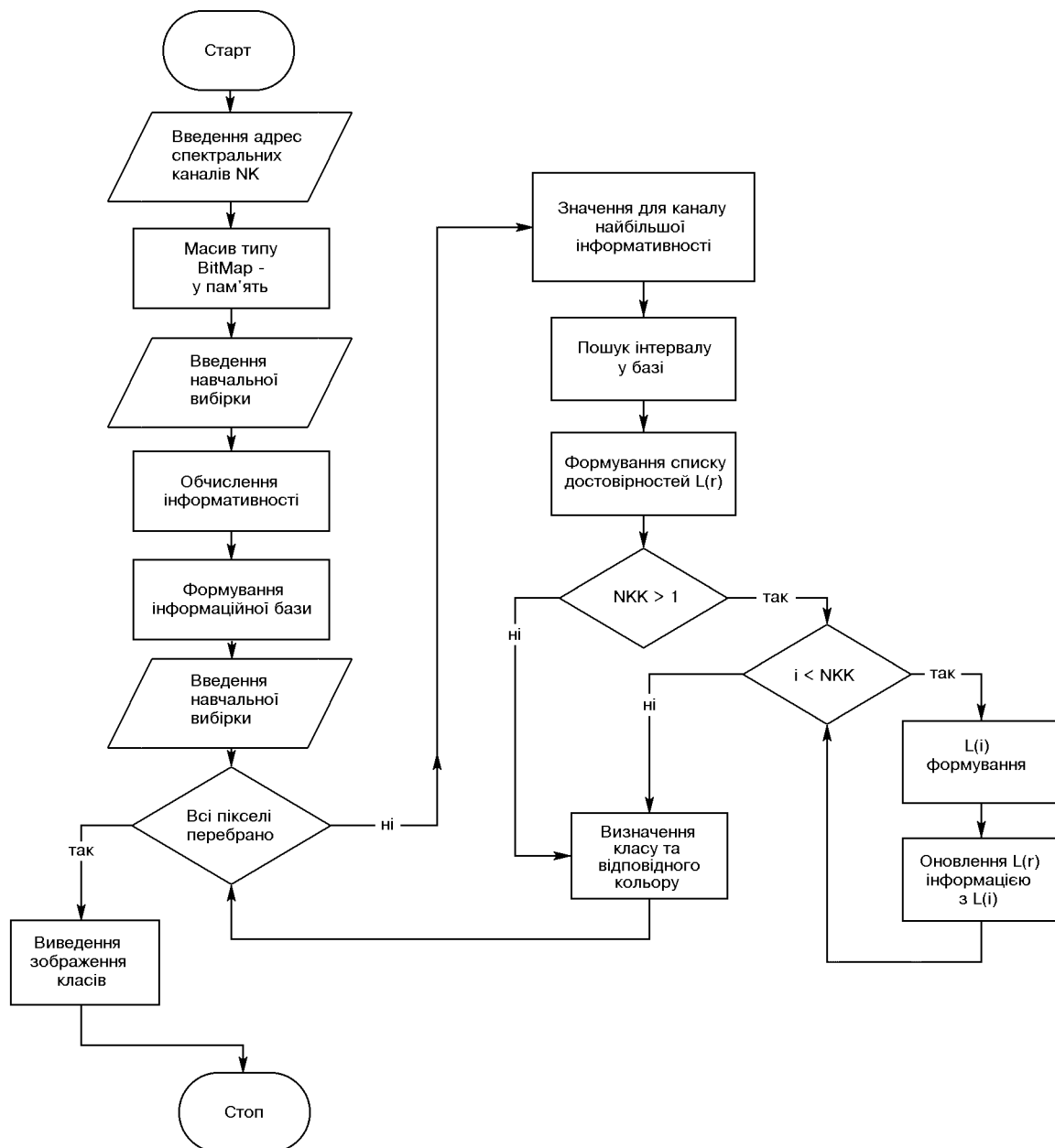


Рис. 1. Блок-схема алгоритму класифікування

подання ознак в інформаційній базі. Аналіз починається з розгляду найбільш суттєвої, з точки зору визначення належності об'єкта класу, ознаки, тобто з аналізу значень об'єкта у найбільш інформативному за формулою (5) спектральному каналі. У середині області значень цієї ознаки розшукується інтервал, до якого потрапляє значення ознаки об'єкта, що розглядається. З отриманого таким чином інтервалу береться інформація щодо достовірності зіставлених класів. Ця інформація заноситься до списку L_1 . Це — первинний список класів, до яких може належати об'єкт, та їхніх достовірностей. Інформація по наступному за своєю інформативністю спектральному каналу заноситься до списку L_2 . З тих двох списків формується новий список L_3 і т.д. [4]. Процес продовжується із залученням чергового спектрального каналу до тих пір, поки в остаточному списку не залишиться один клас або не буде проаналізована інформація про ознаки даного об'єкта по всіх спектральних каналах. Таким чином, у ході послідовного аналізу спектральних ознак об'єкта має місце злиття даних з відсічкою зайвих класів і коригуванням значень достовірності для класів, що залишаються у списку.

Для застосування класифікаційного правила до конкретних об'єктів з використанням запропонованого вище критерію інформативності спектральних каналів та інформаційної бази використаний алгоритм, блок-схема якого надана на рис. 1.

На базі даного алгоритму розроблено програмний модуль оброблення багатоспектральних зображень, для нього обрано середовище програмування Delphi Enterprise version 6.0.

Було проведено перевірку точності результатів класифікування за допомогою даного методу. Точність оцінюється таким чином. Спочатку для кожного класу підраховується R_i — відношення кількості правильно розпізнаних пікселів до кількості невірно розпізнаних пікселів у межах цієї тестової ділянки:

$$R_i = v_i / r_i. \quad (6)$$

Далі підраховується загальна оцінка точності роботи методу $R_{\text{заг}}$ як середнє значення R_i для кожного класу:

$$R_{\text{заг}} = \frac{\sum R_i}{N_{\text{cl}}} \cdot 100, \quad (7)$$

де N_{cl} — кількість класів, які були розпізнані.

З метою оцінки ефективності запропонованого алгоритму класифікування проведено два експериментальних дослідження з використанням космічного знімка, отриманого за допомогою космічного

апарата КА «Landsat-5 у 1988 р. для території зони відчуження Чорнобильської АЕС (рис. 2).

У експериментах використано такі тестові ділянки: 1) ліс, пошкоджений шовкопрядом; 2) здоровий хвойний ліс; 3) листяний ліс; 4) луки; 5) вода.

Експеримент 1. Оцінка інформативності спектральних каналів та точності класифікування.

Інформативність каналів в експерименті становила 0.7 для 3-го спектрального каналу; 0.74 для 4-го каналу і 0.49 — для 5-го каналу. В межах першого експерименту проводилось класифікування за навчальною вибіркою із застосуванням інформації двох спектральних каналів з найбільшими значеннями функції інформативності. Оцінка точності класифікування становила $R_{\text{заг}} = 87\%$.

Експеримент 2. Його метою було порівняння результатів роботи описаного вище алгоритму з результатами методу байєсівської класифікації пакету Erdas Imagine 8.4. Збирання сигнатур для навчання класів було проведено на тих самих ділянках. За допомогою мови макромодельовання пакету Erdas Imagine 8.4 було розроблено макрос для порівняння результатів класифікування, отриманих за допомогою двох алгоритмів. Для проведення порівняння зображення результатів класифікування було приведено до однакових значень псевдокольників відповідних класів. Алгоритм порівняння: проводиться попіксельне порівняння значень тематичних зображень — результатів класифікування; при збігу значень класів у відповідний піксель вихідного зображення пишемо значення 0, у іншому випадку — пишемо значення 1. В результаті було отримано підсумкове чорно-біле зображення, виведене на рис. 3. Тут значенню 0 відповідає білий колір, значенню 1 — чорний. Підрахунок статистики засобами пакету Erdas Imagine показав, що для всіх класів є 893471 пікселів, де значення класів збігаються, та 262670, що відрізняються. Таким чином, загальне розходження результатів класифікування п'яти класів становило 22% від загальної кількості пікселів.

Таким чином, запропоновано новий метод класифікування стану лісових масивів по багатоспектральних космічних зображеннях, який у своїй теоретичній частині базується на концепції злиття даних. Розроблено алгоритм класифікування, оснований на запропонованому критерію інформативності спектральних каналів і правилі визначення належності об'єкта до відповідного класу. На базі описаного алгоритму з використанням елементів геоінформаційних технологій розроблено програмний модуль для класифікування багатоспектральних космічних зображень.



Рис. 2. Знімок території зони відчуження Чорнобильської АЕС, отриманий космічним апаратом «Landsat-5»



Рис. 3. Підсумкове чорно-біле зображення

Оцінено значення інформативності каналів: 0.7 для 3-го каналу, 0.74 для 4-го каналу і 0.49 для 5-го каналу. В цілому точність класифікування становила 87 %.

1. Бугаевский Л. М., Цветков В. Я. Геоинформационные системы. — М.: Златоуст, 2000.—214 с.
2. Кронберг П. Дистанционное изучение Земли: Основы и методы дистанционных исследований в геологии. — М.: Мир, 1988.—343 с.
3. Попов М. О. Сучасні погляди на інтерпретацію даних аерокосмічного дистанційного зондування Землі // Космічна наука і технологія.—2002.—8, № 2/3.—С. 110—115.
4. Попов М. О., Рябоконеко О. Д., Петроченко О. Ю. Підхід до класифікації стану лісових ресурсів по багатоспектральних космічних зображеннях на основі принципу злиття даних та використання елементів ГІС-технологій // Ученые записки Таврического нац. ун-та им. В. И. Вернадского. Сер. География.—2003.—16 (55), № 2.—С. 142—150.
5. Эбелинг В., Энгель А., Файстель Р. Физика процессов

- эволюции. — М.: Эдиториал УРСС, 2001.—328 с.
6. Campbel J. B. Introduction to Remote Sensing, 2nd ed. — London: Taylor and Francis, 1996.—622 p.

METHOD FOR FOREST STATE CLASSIFICATION WITH THE USE OF AEROSPACE IMAGES BASED ON DATA FUSION METHOD

**V. I. Lyalko, M. O. Popov, O. Yu. Petrochenko,
O. D. Ryabokonenko, O. I. Sakhatskyi, O. G. Kharechko**

An approach to forest state classification on the basis of multi-spectral space images is considered. The approach is based on the data fusion method and synergism principle. The program application of the approach involves geoinformation technologies. A classification algorithm based on a proposed criterion of the information content of spectral bands is developed. Some results of experimental investigations of the method proposed are presented.