



ISSN 2411-6602 (Online)

ISSN 1607-2855 (Print)

Том 14 • № 1 • 2018 С. 30 – 34

<https://doi.org/10.18372/2411-6602.14.04>

УДК 528.8+504.064+535.361.2

Методи підвищення точності класифікації агрофітоценозів за супутниковими зображеннями

О.В. Семенів^{1,2,3*}, Л.В. Підгородецька¹

¹Інститут космічних досліджень НАН України і ДКА України, 03187, м. Київ, пр. Академіка Глушкова, 40

²Svitla Systems, Inc., 02002, м. Київ, вул. Євгена Сверстюка, 2А

³Thermo Fisher Scientific, Inc., 3380 Central Expressway, 95051 Санта Клара, США

Стаття присвячена проблемі класифікації агрофітоценозів за даними супутникового спостереження та методами підвищення точності оцінювання. Побудовано модель визначення типу сільськогосподарських культур за мультиспектральними даними на основі методу опорних векторів. Апробація модифікованого алгоритму класифікації сільськогосподарських культур здійснювалася на основі результатів супутникової зйомки пілоотної території господарства "Степовий" Кам'янсько-Дніпровського району Запорізької області. Для цифрової обробки використано наземні опорні дані щодо сівозмін господарства, нормативну базу фаз розвитку сільськогосподарських культур для важко-суглинистих та середньо-суглинистих ґрунтів, характерних для степової зони півдня України та дані ДЗЗ досліджуваної території апаратурою супутників Landsat TM та ETM+ за 2001 та 2003 рр. Досліджено, що значна кількість хибно класифікованих пікселів (> 5%) припадають на зони стику полів (межі полів) та хаотично розкиданих поодиноких сегментів невірно класифікованих культур в межах поля (≈ 12%). В основному дана похибка зумовлена просторовими коливаннями у вологості ґрунту, станом здоров'я рослин, типом проективного покриття, спектральною схожістю культур, фенологічною стадією розвитку рослин, часом супутникового спостереження. Для мінімізації впливу вищезгаданих факторів на кінцевий результат класифікації було використано методи медіанної та рангової фільтрації. Метод медіанної фільтрації дозволив підвищити на 3% точність класифікації, а метод рангової фільтрації — на 2% в порівнянні із простою процедурою методу опорних векторів. Похибка класифікації та недоліки фільтрації значною мірою пов'язані із зонами переходу площ посівів різних культур, так званими межами полів. Оскільки просторове розрізнення становить 30 × 30 метрів, то для навчання моделі класифікації доцільно було б відкидати пікселі переходу (межові) із навчальної вибірки.

Ключові слова: ДЗЗ; оцінювання; стан рослинності; класифікація; фільтрація.

1. ВСТУП

В наш час дані ДЗЗ є важливим джерелом отримання оперативної та об'єктивної інформації про стан сільськогосподарського виробництва на великих і віддалених від центрів управління територіях, що особливо актуально для нашої країни [2, 4, 8, 10, 12, 14]. Сільськогосподарський моніторинг виник на базі використання матеріалів наземних спостережень та оцінок, а також використання моделей прогнозування врожаю за метеорологічними даними [2, 4, 10, 14]. В умовах реального часу, коли мережу наземних пунктів агрономічного спостереження практично згорнуто, а моніторинг засобами авіації, зважаючи на високу вартість, майже не проводиться, використання супутникових даних ДЗЗ є дуже перспективним напрямком [2, 4, 10]. За допомогою даних ДЗЗ здійснюється комплексний підхід до вирішення завдань сільського господарства [2, 4, 10], зокрема: загальний моніторинг сільськогосподарських територій, у тому числі цілих регіонів; спостереження за сніговим покривом та оцінка вологонакопичення; оцінка температури та вологості ґрунту; визначення площі полів під різними культурами; контроль за станом рослинності різних сільськогосподарських культур; виявлення ділянок деградації ґрунту; прогнозування врожайності; моніторинг темпів збирання врожаю тощо.

Одним з найбільш вживаних методів дослідження даних дистанційного зондування є класифікація [7, 9, 12, 13]. Зазвичай результатом обробки даних дистанційного спостереження є класифікаційні карти [7, 9], які дають уявлення про миттєве просторове розподілення різнорідних об'єктів, зокрема, рослинного покриву земної поверхні. Подібні карти є необхідними в багатьох прикладних задачах, наприклад, при вивченні змін рослинності. Однак такі напівавтоматизовані підходи в значній мірі залежать від спектральної чутливості приладів, рівня шумів, що обумовлені просторовими коливаннями у вологості ґрунту, здоров'ям рослин, типом проективного покриття, спектральною схожістю культур, фенологічною стадією розвитку рослин, часом супутникового спостереження. Вищезгадані фактори вимагають докладання значних зусиль на етапі навчання моделей класифікації, що звичайно не гарантує отримання задовільних результатів. Ще одним недоліком є необхідність групування класифікованих пікселів у сільськогосподарські об'єкти (поля), що деколи є доволі не тривіальною задачею.

* Семенів Олег Володимирович; ✉ oleg.semeniv@gmail.com

Значне місце у виділенні інформації та її структуризації за супутниковими знімками займає сегментація зображення на основі спектральних відмінностей [6, 15, 16]. На відміну від методів класифікації алгоритми сегментації дозволяють одразу здійснювати ідентифікацію об'єктів (в тому числі і полів) на супутниковому зображенні, адже в даних підходах агрегується різнорідна геопросторова інформація. Проте даний аспект слугує значним обмеженням, тому що не завжди можливо отримати повну інформацію щодо досліджуваної ділянки земного покриву. Також значним обмеженнями є необхідність проведення додаткової верифікації, накопичення значних баз даних та комп'ютерних ресурсів для проведення сегментації.

Усе вище описане стимулює до пошуку інтегральних підходів для підвищення точності виявлення та класифікації агрофітоценозів на основі супутникових зображень при мінімальній кількості супутньої інформації. Враховуючи можливість отримання даних ДДЗ у вільному доступі більш доцільним виглядає покращення методів класифікації із їх модифікацією для більш якісного групування прокласифікованих пікселів за типом культури (групування у поля). В даній роботі розглянуті підходи, що дозволяють в значній мірі підвищити точність класифікації пікселів в межах сільськогосподарських об'єктів.

2. КЛАСИФІКАЦІЯ АГРОФІТОЦЕНОЗІВ НА ОСНОВІ ДДЗ

Спираючись на попередньо проведені дослідження [3, 12, 13], процес класифікації супутникового зображення, а саме пікселів, розглядається в рамках класифікації вхідних векторів \mathbf{x}_n , $n = 1, \dots, N$ на багато класів, де кожен вектор містить спектральні характеристики пікселя і розмірність вектору залежить від кількості спектральних каналів приладу. Задачу класифікації, як показано в [12, 13], можна представити у вигляді поетапної класифікації на два класи $z_n \in \{-1, 1\}$ із подальшим розрахунком максимальної ймовірності (інтегральної характеристики) потрапляння одного із пікселів до одного із заданих класів, що здійснюється шляхом розв'язання задачі

$$\min_{\mathbf{w}, b, \xi} \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{n=1}^N \xi_n, \quad z_n (\mathbf{w}^T \varphi(\mathbf{x}_n) + b) \geq 1 - \xi_n, \quad \xi_n \geq 0, \quad n = 1, \dots, N, \quad (1)$$

де $\varphi(\mathbf{x}_n)$ — функція, що проектує набір векторів \mathbf{x}_n в простір вищого порядку, C — регуляризаційний параметр. Задачу (1) зазвичай зводять до вирішення дуальної задачі

$$\min \frac{1}{2} \boldsymbol{\alpha}^T \mathbf{Q} \boldsymbol{\alpha} - \mathbf{e}^T \boldsymbol{\alpha}, \quad 0 \leq \alpha_n \leq C, \quad n = 1, \dots, N, \quad \mathbf{z}^T \boldsymbol{\alpha} = 0,$$

де \mathbf{e} — одиничний вектор; C — параметр, що характеризує верхню допустиму границю значень; \mathbf{Q} — позитивна матриця $N \times N$, $Q_{ni} = z_n z_i K(\mathbf{x}_n, \mathbf{x}_i)$, $n \neq i$; $K(\mathbf{x}_n, \mathbf{x}_i) = \exp(-\gamma \|\mathbf{x}_n - \mathbf{x}_i\|^2)$, $\gamma > 0$ — RBF kern-функція.

В ході досліджень побудовано модель на основі опорних векторів класифікації сільськогосподарських угідь за типом культури та шляхом комп'ютерного моделювання отримано оптимальні статистичні параметри, що дозволяють забезпечити високу величину показника узагальнення та регулярності. Проте в ході дослідження результатів виявлено, що значна кількість хибно класифікованих пікселів (> 5%) припадають на зони стиків полів (межі полів) та хаотично розкиданих поодиноких сегментів невірно класифікованих культур в межах поля ($\approx 12\%$). Таким чином, постала задача мінімізувати похибку, що викликана хаотично розкиданими поодинокими сегментами в межах поля одного типу культур.

3. ПІДВИЩЕННЯ ТОЧНОСТІ КЛАСИФІКАЦІЇ СТАНУ АГРОФІТОЦЕНОЗІВ

3.1. Медіанна фільтрація

Одним з методів цифрової обробки зображень, що застосовуються для усунення дефектів зображення, зумовлених перешкодами та шумами, є медіанна фільтрація [5, 11]. Медіанні фільтри відрізняються робастністю та використовуються для згладжування інформації у випадках, коли шумові характеристики невідомі. В загальному випадку медіану можна визначити як величину z_{med} , для якої при будь-яких значеннях y є справедливим вираз

$$\sum_{n=1}^N |z_{\text{med}} - z_n| \leq \sum_{n=1}^N |y - z_n|,$$

де z_n — деяка точка на площині, що відображає приналежність пікселя n до одного із класів. Медіанна фільтрація реалізує вибір z_{med} для непарного n однозначно (у той же час для парного значення n існує нескінченне число можливих значень z_{med}). Так, для двовимірного вікна (випадку класифікованої карти земної поверхні), що містить 3×3 елементів зображення (пікселів), медіанний фільтр вибирає вектор із середнім значенням, яке потім присвоюється центральному пікселю вікна. Незалежно від форми вікна фільтр реалізує той самий алгоритм і характеризується кількістю (n) та розрядністю (m) оброблюваних пікселів. Таким чином, медіаною z_{med} дискретної послідовності двійкових векторів z_n ($n = 1, N$) для непарного n є той її елемент, для якого існує $(n-1)/2$ елементів, менших або рівних йому за значенням, та $(n-1)/2$ елементів, більших або рівних йому за значенням. Враховуючи, що класифікаційна оцінка завжди є цілочисельною, в межах одного поля будуть відкинуті поодинокі пікселі або незначні сегменти, що виділяються на фоні оточуючих їх.

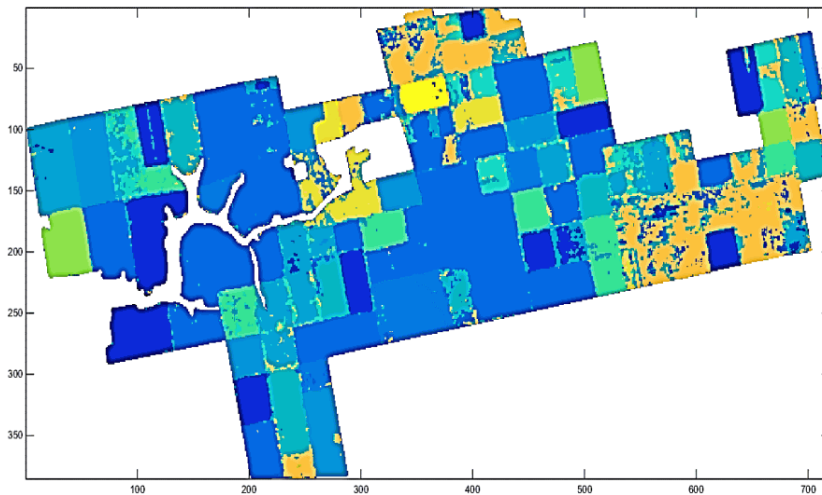


Рис. 1. Відображення результатів медіанної фільтрації результатів класифікації на основі методу опорних векторів

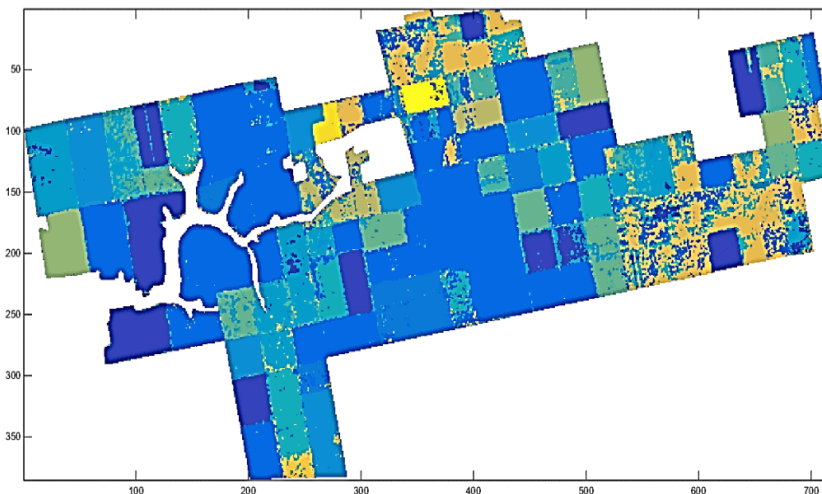


Рис. 2. Відображення результатів рангової фільтрації отриманої класифікації на основі методу опорних векторів

Таблиця 1. Похибка класифікації із застосуванням різних типів фільтрів у відсотковому відношенні

Тип фільтрації/Вибірки	Без фільтрації	Медіанна фільтрація	Рангова фільтрація
Включені непосівні поля	19,87%	16,47%	17,95%
Виключені непосівні поля	18,48%	15,32%	16,68%

Алгоритм фільтрації результатів класифікації на основі медіанного фільтру реалізований в якості ітеративної процедури, яка полягає в оцінюванні інтегрального показника заміщення

$$\varepsilon_j = \sum_{n=1}^N |z_n^j - z_n|, \quad j = \overline{1, J}, \quad J = \text{const}, \quad \varepsilon_j - \varepsilon_{j-1} > \delta N,$$

де J — кількість ітерацій; ε_j — інтегральний показник заміщення, для $j=0$ рівний $\varepsilon_0=0$; δ — граничний відсоток для зупинки роботи алгоритму $\delta=0,1$, $\delta \leq 0,2$. Результати фільтрації результатів класифікації продемонстровано на рис. 1, а статистичні показники точності — в табл. 1.

3.2. Рангова фільтрація

Рангова фільтрація є безпосереднім узагальненням бінарної рангової фільтрації та спирається на поняття порядкової статистики [1]. Навколо кожного елемента зображення (пікселя) вибирається околі, елементи зображення, що потрапляють в цей околі впорядковуються за зростанням. Ранговий фільтр порядку r ($1 < r < N^*$, де N^* — число відліків в околі) вибирає з отриманого ряду елемент з номером r і присвоює його значення вихідному елементу зображення. Часто в якості відібраного елемента виступає мінімальне чи максимальне значення, присутне у вікні (околі) N^* . З метою збереження форми корисного сигналу потрібна послідовна схема мінімаксної фільтрації, що складається з двох проходів по зображенню і обробки спочатку мінімальної (максимальної), а потім максимальної (мінімальної) за рангом локальної статистики. Така схема збільшує ефективність фільтрації. Таким чином, після проходження такої схеми отримуємо незмінну оцінку інтегрального показника заміщення. Результати застосування рангової фільтрації до результатів класифікації представлені на рис. 2, а статистичні показники — в

табл. 1.

Підвищення якості рангової фільтрації, а відповідно і точності класифікації, можна досягти шляхом модифікації процедури подвійної прогонки із урахуванням контурів (меж) окремих полів (геометричних полігонів на зображенні). В межах кожного полігону проводиться підрахунок кількості присутніх представників класів, після чого відбувається визначення класу, представники якого є більшістю з представлених на полігоні, потім відбувається процес перейменування усіх представників полігону класу більшості. Даному напрямку будуть присвячені наступні дослідження.

3.3. Порівняння результатів фільтрації класифікованих агропромислових культур

Вище представлені методи фільтрації, використані для підвищення точності результатів класифікації на основі методу опорних векторів [13]. Як видно із табл. 1, порівняно з простою процедурою методу опорних векторів метод медіанної фільтрації дозволяє підвищити точність класифікації до 3%, а метод рангової фільтрації — до 2%. Похибка класифікації та недоліки фільтрації пояснюються наявністю зон переходу площ посівів різних культур, так званими межами полів. Оскільки просторове розрізнення становить 30×30 метрів, то для навчання моделі класифікації із навчальної вибірки доцільно було б відкидати пікселі переходу (на межі).

Побудова алгоритму виділення контурів полів дозволить в значній мірі підвищити точність класифікації як на етапі безпосередньої класифікації, так і на етапі фільтрації. В даному аспекті більш перспективною виглядає рангова фільтрація, а саме її модифікація шляхом застосування процедури визначення приналежності пікселя до більшості в межах одного поля. Дана процедура полягає у підрахунку усіх результатів класифікації, що знаходяться в межах одного поля, та вибору класу більшості представників. Таким чином рангова фільтрація є більш перспективна в комплексному алгоритмі класифікації агрофітоценозів, хоча на даному етапі вона демонструє гірші результати.

4. ВИСНОВКИ

Продовжено розробку математичної моделі дистанційного оцінювання стану агрофітоценозів з використанням мультиспектральних даних. Набув подальшого розвитку метод опорних векторів та розроблено ряд модифікацій процедури обробки супутникових зображень та пост-фільтрації результатів класифікації. Розроблено комплексний підхід класифікації агрофітоценозів, який полягає у застосуванні моделі опорних векторів та методів просторової фільтрації. Проведено аналіз статистичних параметрів точності класифікації за наземними завірковими даними та розроблено процедуру валідації моделі. Здійснено порівняльний аналіз результатів класифікації із застосуванням різних схем фільтрації. Встановлено, що застосування методів рангової та медіанної фільтрації дозволяють підвищити точність класифікації на 2 та 3 відсотки відповідно. Запропоновано схему вдосконалення процедури навчання та післяобробки результатів класифікації для отримання ще більш точної класифікації.

1. *Визильтер Ю.В., Желтов С.Ю., Князь В.А., Ходарев А.Н., Моржин А.В.* Обработка и анализ цифровых изображений с примерами на LabVIEW IMAQ Vision. — М.: ДМК Пресс, 2009. — 464 с.
2. *Лялько В.І., Попов М.О.* Багатоспектральні методи дистанційного зондування Землі в задачах природокористування. — Київ: Наукова думка, 2006. — 360 с.
3. *Семенів О.В.* Оцінювання біохімічних компонент в рослинності на основі методів статистичного навчання та даних дистанційного зондування // Вісник Астрономічної школи. — 2014. — Том 10, № 1. — С.70–74. <https://doi.org/10.18372/2411-6602.10.1070>
4. *Шадчина Т.М.* Наукові основи дистанційного моніторингу стану посівів зернових. — Київ: Фітосоціоцентр, 2001. — 220 с.
5. *Astola J., Naavisto P., Neuvo Y.* Vector median filters // Proc. of the IEEE. — 1990. — Vol.78(4). — P.678–689. <https://doi.org/10.1109/5.54807>
6. *Bo Chen, Fang Qiu, Bingfang Wu, Hongyue Du.* Image segmentation based on constrained spectral variance difference and edge penalty // Remote Sens. — 2015. — Vol.7. — P.5980–6004. <https://doi.org/10.3390/rs70505980>
7. *Bruzzone L., Pioto D.F.* A partially unsupervised cascade classifier for the analysis of multitemporal remote-sensing images // Pattern Recognition Letters. — 2002. — Vol.23. — P.1063–1071. [https://doi.org/10.1016/s0167-8655\(02\)00053-3](https://doi.org/10.1016/s0167-8655(02)00053-3)
8. *Egbert S.L., Park S., Price K.P.* Using conservation reserve program maps derived from satellite imagery to characterize landscape structure // Compute Electron Agric. — 2002. — Vol.37. — P.141–156. [https://doi.org/10.1016/s0168-1699\(02\)00114-x](https://doi.org/10.1016/s0168-1699(02)00114-x)
9. *Foody G.M., Mathur A.* Toward intelligent training of supervised image classifications: directing training data acquisition for SVM classification // Remote Sensing of Environment. — 2004. — Vol.93. — P.107–117. <https://doi.org/10.1016/j.rse.2004.06.017>
10. *Kogan F.N.* Operational space technology for global vegetation assessment // Bull. Am. Meteorol. Soc. — 2001. — Vol.82(9). — P.1949–1964. [https://doi.org/10.1175/1520-0477\(2001\)082<1949:ostfgv>2.3.co;2](https://doi.org/10.1175/1520-0477(2001)082<1949:ostfgv>2.3.co;2)
11. *Lim J.S.* Two-dimensional signal and image processing. — Englewood Cliffs, NJ, Prentice Hall, 1990. — P.469–476.

12. Podgorodetskaya L.V., Prutsko Y.V., Semenov O.V. Farmland state estimation based on agrophytocoenoses classification and remote sensing satellite data // Journal of Automation and Information Sciences. — 2016. — Vol. 48(5). — P.55–63. <https://doi.org/10.1615/jautomatinfscien.v48.i5.50>
13. Semenov O.V., Pidgorodetska L.V. Biophysical Parameters of Water Ecosystem Estimation Using Satellite Images and Optimization Techniques // Journal of Automation and Information Sciences. — 2014. — Vol. 46(9). — P.68–77. <https://doi.org/10.1615/jautomatinfscien.v46.i9.60>
14. Schmedtmann J., Campagnolo M.L. Reliable crop identification with satellite imagery in the context of common agriculture policy subsidy control // Remote Sens. — 2015. — Vol. 7. — P.9325–9346. <https://doi.org/10.3390/rs70709325>
15. Yan L., Roy D.P. Automated crop field extraction from multi-temporal Web Enabled Landsat Data // Remote Sensing of Environment. — 2014. — Vol. 144. — P.42–64. <https://doi.org/10.1016/j.rse.2014.01.006>
16. Xi Chen, Tao Fang, Hong Huo, Deren Li Measuring the effectiveness of various features for thematic information extraction from very high resolution remote sensing imagery // IEEE Transactions on geoscience and remote sensing. — 2015. — Vol. 53(9). — P.4837–4851. <https://doi.org/10.1109/tgrs.2015.2411331>

Методы повышения точности классификации агрофитоценозов по спутниковым изображениям

Семенов О.В.^{1,2,3}, Подгородецкая Л.В.¹

¹Институт космических исследований НАН Украины и ГКА Украины, 03187, г. Киев, пр. Академика Глушкова, 40

²Svitla Systems, Inc., 02002, г. Киев, вул. Евгения Сверстюка, 2А

³Thermo Fisher Scientific, Inc., 3380 Central Expressway, 95051 Санта Клара, США

Статья посвящена проблеме классификации агрофитоценозов по данным спутникового наблюдения и методам повышения точности оценки. Построена модель определения типа сельскохозяйственных культур по мультиспектральным данным на основе метода опорных векторов. Апробация модифицированного алгоритма классификации сельскохозяйственных культур осуществлялась на основе результатов спутниковой съемки пилотной территории хозяйства «Степной» Каменско-Днепровского района Запорожской области. Для цифровой обработки использованы наземные опорные данные по севооборотам хозяйства, нормативная база фаз развития сельхозкультур для тяжело-суглинистых и средне-суглинистых почв, характерных для степной зоны юга Украины и данные ДЗЗ исследуемой территории аппаратурой спутников Landsat TM и ETM+ за 2001 и 2003 гг. Доказано, что значительное количество неправильно классифицированных пикселей (> 5%) приходится на зоны стыков полей (границы полей) и хаотично разбросанных отдельных сегментов неверно классифицированных культур в пределах поля ($\approx 12\%$). В основном данная погрешность обусловлена пространственными колебаниями влажности почвы, состоянием здоровья растений, типом проективного покрытия, спектральной схожестью культур, фенологическими стадиями развития растений, временем спутникового наблюдения. Для минимизации влияния вышеупомянутых факторов на конечный результат классификации были использованы методы медианной и ранговой фильтрации. Метод медианной фильтрации позволил повысить на 3% точность классификации, а метод ранговой фильтрации — на 2% по сравнению с простой процедурой метода опорных векторов. Погрешность классификации и недостатки фильтрации в значительной степени связаны с зонами перехода площадей посевов различных культур, так называемыми границами полей. Поскольку пространственное разрешение составляет 30×30 метров, то для обучения модели классификации целесообразно было бы отвергать пиксели перехода (пограничные) из обучающей выборки.

Ключевые слова: ДЗЗ; оценка, состояние растительности; классификация; фильтрация.

Methods of increasing the accuracy of classification of agrophytocoenoses by satellite images

Semenov O.V.^{1,2,3}, Pidgorodetska L.V.¹

¹Space Research Institute NASU & SSAU, 03187, Hlushkova Avenue 40, Kyiv, Ukraine

²Svitla Systems, Inc., 02002, Yevhena Sverstyuka St. 2A, Kyiv, Ukraine

³Thermo Fisher Scientific, Inc., 3380 Central Expressway, Santa Clara, CA 95051, USA

The article is devoted to the problem of the agrophytocoenoses classification with satellite observation data and noise filtration methods. The model of agricultural crop types determination with multispectral data and the support vector machine method was conducted. The modified algorithm for agricultural crops classification approbation was carried out on the satellite imageries of the “Steppe” farm of the Kamyansko-Dniprovsky district of the Zaporizhzhya region. Ground-based reference data on crop types, normative base of agricultural development phases for heavy-loamy and medium-loamy soils characteristics of the southern Ukraine steppe zone and remote sensing data of the studied territory gathered by Landsat TM and ETM+ satellites for 2001 and 2003 were used for digital processing. It was investigated that a significant number of false-detected pixels (> 5%) belong to the interface zones (field boundaries) and chaotic scattered segments of incorrectly classified crops within the field ($\approx 12\%$). This error is mainly due to spatial variations in soil moisture, plant health, type of projective surface, cultures spectral similarity, phenological stage of plants development, time of satellite observation. Median and rank filtration methods were used to minimize the effect of the above-mentioned problems on the final classification result. The method of median filtration allowed to increase the classification accuracy up to 3%, and the rank filtration method up to 2% compared with the simple procedure of the support vector machine method. The classification error and the disadvantages of filtration are largely due to the interface zones of the different crops, the fields boundaries. Since the spatial resolution is 30×30 meters, it would be advisable to reject the pixels of the transition (on the boundary) from the training sample to identify more accurate classification model.

Keywords: remote sensing; estimation; vegetation state; classification; filtration.

Надійшла до редакції / Received 31.07.2018

Виправлена авторами / Revised 14.08.2018

Прийнята до друку / Accepted 15.08.2018