

ІНСТРУМЕНТАЛЬНІ ЗАСОБИ АВТОМАТИЗОВАНОГО ПРОЕКТУВАННЯ

УДК 621

Р.А. Мельник, Р. Тушницький
Національний університет “Львівська політехніка”,
кафедра програмного забезпечення

ЗАСТОВУВАННЯ ПАКЕТА КЛАСТЕРИЗАЦІЇ ОБРАЗІВ ДЛЯ ОЦІНЮВАННЯ ПОКАЗНИКІВ СТАНУ КРОВІ

© Мельник Р.А., Тушницький Р., 2007

Наведено результати застосування пакета кластеризації візуальних образів для кількісного аналізу зразків крові, а саме визначення розмірів та кількості різнотипних клітинок зразка крові. Пакет містить алгоритми трьох рівнів кластеризації: покриття образу прямокутниками, побудови з них зв'язних областей та кластеризації зв'язних областей – образів клітинок.

The clustering package to decompose visual patterns was used to analyze the blood sample to get a number and dimensions of its different cells. The package contains three level clustering algorithms : coverage of patterns by rectangles, formulation of linked areas and clustering of the linked areas – blood cells.

Вступ. Для декомпозиції, дослідження та розпізнавання зображень використовують підходи, що ґрунтуються на методах кластеризації [1, 2, 3]. Характерною ознакою робіт у цьому напрямку є кластеризація однотонних пікселів і утворення фігур з характеристиками кольору чи відтінку. Метою роботи є розроблення та застосування програмного пакета до визначення характеристик зразка крові. Пакет ґрунтується на використанні алгоритмів кластеризації образів у три етапи: попереднього розбиття образів на прямокутники, згортання останніх у зв'язні області конкретного фізичного змісту [4, 5] та безпосереднього кластерного аналізу клітин, виділених в регіони з зображення. Зменшення алгоритмічної складності досягається динамічним керуванням розміру поля образу, у якому ведеться пошук кластерів.

Постановка задачі декомпозиції образів. Визначити тип клітин крові, їхні кількість та розміри можна під час їхнього виділення та кластерного аналізу. Об'єднання розроблених алгоритмів в пакет, аналіз зображення та його дослідження є завданнями цієї роботи. В основу інструменту покладено алгоритми декомпозиції візуальних образів на геометричні фігури покриттям фрагментів образу прямокутниками та об'єднання їх у зв'язні області. Алгоритми ґрунтуються на принципах побудови ієрархічного дерева. Кластеризація кольорових зображень здійснюється з попереднім фільтруванням і перетворенням до чорно-білого зображення.

Згортання полягає в поділі візуальних образів на класи – групи об'єктів, схожі між собою за певними ознаками. Початково образ (об'єкт) розбивається на мікрооб'єкти за певним правилом. Надалі об'єднанням мікрооб'єктів утворюються мікрокластери, характеристики F_1, F_2, \dots, F_n яких формуються як суми чи інші функції характеристик мікрооб'єктів, що повинні задовольняти обмеження певного виду. Значення функцій ознак мікрокластерів, що формуються, відповідають за яскравість, колір, співвідношення заповнених та пустих клітинок тощо. Вершини, що не підлягають подальшому згортанню, відповідають кластерам результату.

Зменшення складності алгоритму досягають на етапі пошуку кандидатів кластерів для об'єднання розділенням об'єктів, що розглядаються. Пошук кандидатів мікрокластерів на згортання здійснюється не у повному зображенні, а в його частині, покритій прямокутником з заданими

розмірами (рис. 1, а). Останній рухається за певними правилами, скануючи всю поверхню зображення. На рис. 1, б показано, що алгоритм сканування образу працює зліва направо, а на рис. 1, в показано, що робоча область обмежена справа і зліва.

Одна розмірність прямокутника відповідає розміру зображення, а інша формується динамічно залежно від розмірів кластерів. Наприклад, спочатку він збігається з розмірністю сітки розбиття, а надалі під час переходу на наступні рівні дерева згортання розмірність збільшується вдвічі. Отже, одна розмірність прямокутника сканування адаптується до розмірів кластерів, що зменшує кількість перегляду малоймовірних сусідів поточного кластера.

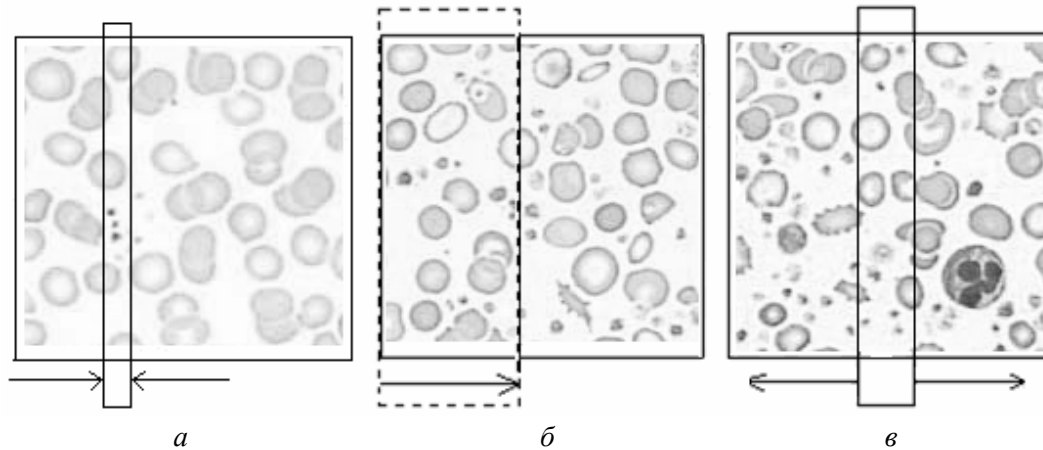


Рис. 1. Сканувальна область алгоритму згортання

Метою другого етапу опрацювання зображення є його декомпозиція на зв'язні області, що реалізується алгоритмом кластеризації у просторі отриманих прямокутників покриття зображення та за іншим критерієм об'єднання. Множина всіх прямокутників $Q(Q_1, Q_2, Q_3, \dots, Q_N)$ є об'єктом алгоритму згортання, в результаті роботи якого формуються p підмножин $O_1(Q_1, Q_2, Q_3, \dots, Q_z)$, $O_2(Q_{z+1}, Q_{z+2}, Q_{z+3}, \dots, Q_l)$, \dots , $O_p(Q_{t+1}, Q_{t+2}, Q_{t+3}, \dots, Q_N)$, кожна з яких визначає фрагмент зображення. За критерієм об'єднання прямокутників приймають кількість суміжних між прямокутниками клітинок, а за необхідності врахування можливих втрат критерієм вважають віддаль s між зв'язними областями – кластерами нижчих рівнів.

Керування значенням відстані s між кластерами дає змогу формувати зв'язні області, що відповідають реальним фрагментам зображення. Зокрема, розподіл відстаней на головну та діагональну корисний для опрацювання зображень із кільцевою структурою – відбитку пальця з “ярами” та “пагорбами”. На рис. 2 наведено результати керування головною та діагональною відстанями в алгоритмі формування зв'язних областей: а – відстань = 1, отримано областей – 333; б – відстань = 2, областей – 235. У реальному тестуванні наведені клітинки зображення є різнокольоровими для виділення окремих регіонів.

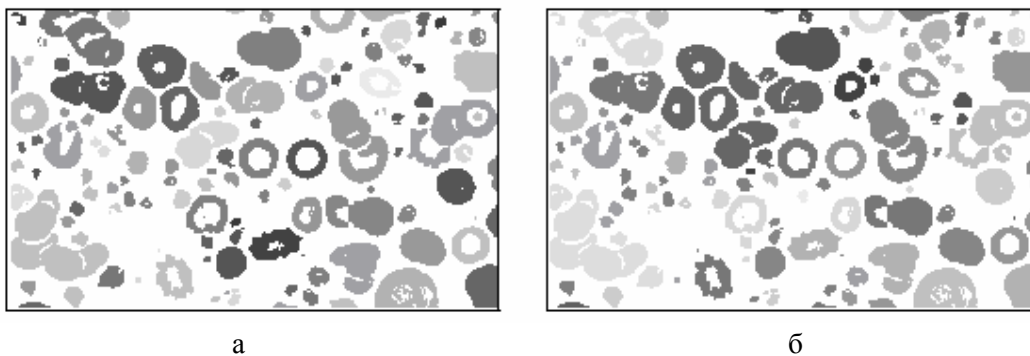


Рис. 2. Приклади керування формування зв'язних областей

Застосування кластеризації для підрахунку клітин крові. Складовою частиною аналізу крові є визначення кількості та розмірів складових клітинок крові: лейкоцитів, еритроцитів тощо. Розроблений алгоритм було застосовано для опрацювання зображення еритроцитів, а саме визначення їхньої кількості та розмірів; виділення тромбоцитів та інших елементів крові з метою визначення їх розмірів та кількості. На рис. 3 наведено декілька зразків крові із нормальним та завищеним числом кров'яних клітин, які приймаються як вхідні дані для розробленого пакета програм. Реальні зображення подаються в кольорі, який на цих рисунках не відтворено.

Програма кластеризації образів працює в чорно-білому форматі. Тому для опрацювання кольорових образів використовують два фільтри, які спочатку перетворюють кольорове зображення у відтінки сірого як середнє арифметичне червоної, зеленої та синьої компоненти кольору кожного пікселя. Надалі отримане сіре зображення перетворюється у чорно-біле.

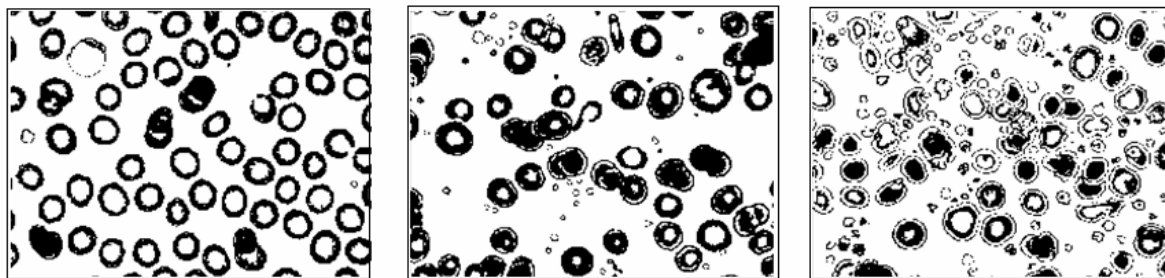


Рис. 3. Зображення зразків крові

Для фільтра встановлюється нижня (*lo*) та верхня (*hi*) межі, що становлять *RGB*-компоненти кольору. Так, якщо нижня або верхня межа є 200, то це колір *RGB*(200, 200, 200). Діапазон дії фільтра – це гама відтінків сірого кольору, які будуть інтерпретуватися як чорний (рис. 4).



Рис. 4. Діапазон дії фільтра

Якщо в образі, який фільтрують, значення пікселя знаходиться у діапазоні між нижньою і верхніми межами, то піксель інтерпретується як чорний. Отже, керування відтінком сірого кольору як параметром фільтра дає змогу керувати контрастністю зображення. У разі збільшення верхньої межі менш насичений сірий піксель інтерпретується як чорний, чим підвищується чіткість образу. Внаслідок цього програма кластеризації зображень виділяє окремі зв'язні області, що являють собою конкретні клітинки крові. На рис. 5 наведено результати застосування фільтра із різною верхньою межею: *а* – вихідний образ у тонах сірого; *б* – верхня межа 220; *в* – межа 230; *г* – 240. При значенні верхньої межі фільтра 220 границі окремих еритроцитів є менш чіткими на відміну від параметра 230. Подальше підвищення верхньої межі фільтра призводить до злиття об'єктів

Отже, фільтри дають змогу, по-перше, опрацювати кольорові зображення, по-друге, підсилити їхні контрастність та чіткість, впливаючи цим на кількість зв'язних областей.

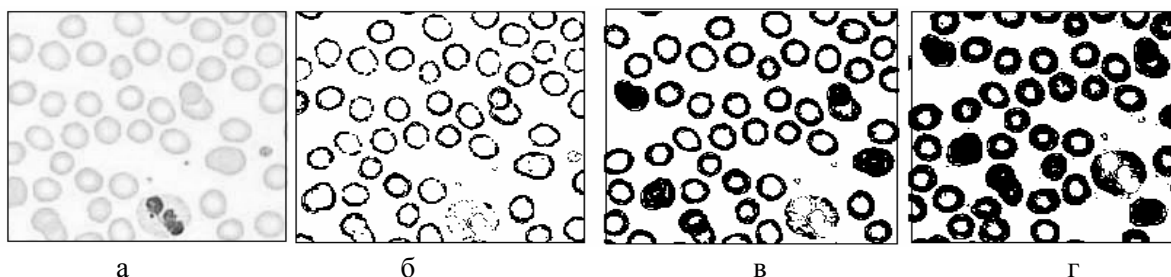


Рис. 5. Фільтрування з різними верхніми межами

Кластеризація регіонів для аналізу зображення. Регіони – зв’язні області – подамо множиною $Q(Q_1, Q_2, Q_3, \dots, Q_N)$, де N – достатньо велике значення. Кожному регіону як об’єкту певного фізичного змісту притаманні характеристичні ознаки a, b, c, \dots (координати центра, площа, яскравість тощо), які формують його ключ та виражаються числовими значеннями. Образу як сукупності регіонів відповідає множина його характеристик $Q_i = Q_i(a_i, b_i, c_i, \dots)$.

Для кластеризації регіонів використаємо процедуру, подібну до використаної при формуванні прямокутних кластерів чи регіонів довільної форми, але реалізованої для ключів образів [6]. Критерій об’єднання для побудови дерева згортання утворимо як зважену суму квадратів різниць між характеристиками

$$F_{ij} = w_1 [a_i - a_j]^2 + w_2 [b_i - b_j]^2 + w_3 [c_i - c_j]^2 + \dots,$$

де сумування відбувається за всіма ознаками, які формують ключ образу – клітини.

У процесі побудови дерева утворюються характеристики новоутворених вершин за зваженим усередненням характеристик складових вершин:

$$Q_k = Q_k((k \cdot a_i + r \cdot a_j) / (k+r), (k \cdot b_i + r \cdot b_j) / (k+r), (k \cdot c_i + r \cdot c_j) / (k+r), \dots),$$

де k, r – кількості ключів у i -му j -му вузлах дерева згортання.

Для оцінки якості результату кластеризації на основі введених ознак утворимо критерій середньої питомої густини ключів результатів цього опрацювання

$$S = \{ \sum L_k / D_k \} / M,$$

де доданками є питомі густини окремих кластерів, L_k , D_k – кількість ключів та діаметр k -го кластера, M – кількість кластерів.

Із збільшенням кількості кластерів величини: S – збільшується, V – зменшується. Точку різкої зміни градієнта цих величин приймаємо за відлік оптимальної кількості кластерів.

Програмний пакет. Розроблено експериментальний програмний пакет з інтерфейсом користувача, що контролює всі етапи роботи: введення образу, контроль параметрів керування, виведення варіантів кластеризації (повної, уточненої), відновленого зображення та повний звіт про виконання алгоритму, числові параметри результатів тощо. Пакет реалізує два етапи алгоритму кластеризації: покриття образу прямокутниками та формування зв’язних областей з прямокутників. Пакет має засоби для керування як для покриття прямокутниками, так і для формування зв’язних областей. Зокрема, розмір сітки покриття, яскравість кластера, головна та діагональна відстані. Протокол роботи програми заноситься у вихідний інтерфейс.

Обговорення результатів. У табл. 1 наведено результати дослідження впливу параметрів фільтрування на вхідне зображення, кількісні та якісні показники сегментів зображення. З таблиці видно, що параметри фільтра істотно впливають на вхідні дані (кількість мікрокластерів), а також на вихідні результати. Для більшої верхньої межі фільтра кількість мікрокластерів та кластерів зростає.

Таблиця 1

Вплив параметрів фільтра на вхідні і вихідні дані алгоритму кластеризації

Образ	Розміри	Параметри фільтра		Кількість мікрокластерів	Кількість кластерів	К-сть всіх зв’язних областей*	Кількість еритроцитів**	Реальна кількість еритроцитів
		Нижня межа	Верхня межа					
1274N	260 x 420	200	220	17 661	4 537	311	163	150
		200	230	31 053	5 486	200	146	
		200	240	44 355	6 037	289	152	
1691O	260 x 420	200	220	22 195	6 611	838	78	138
		200	230	35 903	6 935	458	125	
		200	240	46 528	7 471	442	140	

* Еритроцити, нейтрофіли, тромбоцити, стовпчики еритроцитів, лімфоцити.

** Кількість еритроцитів, підрахована програмно, залежить від способу підрахунку.

Кількість сформованих зв'язних областей прямо незалежить від кількості мікрокластерів, оскільки у разі підвищення контрастності об'єкти починають дотикатися, а це приводить до їх об'єднання. Загалом із підвищенням контрастності кількість сформованих зв'язних областей збільшується.

Зауважимо, що всі зв'язні області містять еритроцити та всі інші клітини крові. На рис. 2 можна побачити нейтрофіли, стовпчики із еритроцитів, тромбоцити та лімфоцити.

У табл. 2 наведено результати кластеризації регіонів (образів клітинок). Клітинок з площею, меншою ніж 26, є 54 (тромбоцити). Якщо площа перевищує значення 400 (їх кількість 15), то це є стовпчики еритроцитів. Площу в межах значень 76–400 мають еритроцити, нейтрофіли та лімфоцити.

Таблиця 2

Розподіл регіонів за площею

	Межі значень площі зв'язних областей						
	<= 26	27–75	76–118	119–162	163–200	201–400	> 400
Кількість	54	13	10	4	40	64	15
Площа	266	638	946	557	7495	14431	6720
Загальна кількість	200						
Загальна площа	31 053						
Кількість, %	27	6,5	5	2	20	32	7,5
Площа, %	0,85	2,06	3,05	1,79	24,14	46,47	21,64

Дерево кластеризації клітинок зображення зразка крові наведено на рис.6. За необхідності в ньому можна виділити групи клітинок з меншими діаметрами. Оптимальною кількістю груп клітинок для цього експерименту за алгоритмом є 11. На осі ординат рис. 6 наведені значення функції для об'єднання кластерів.

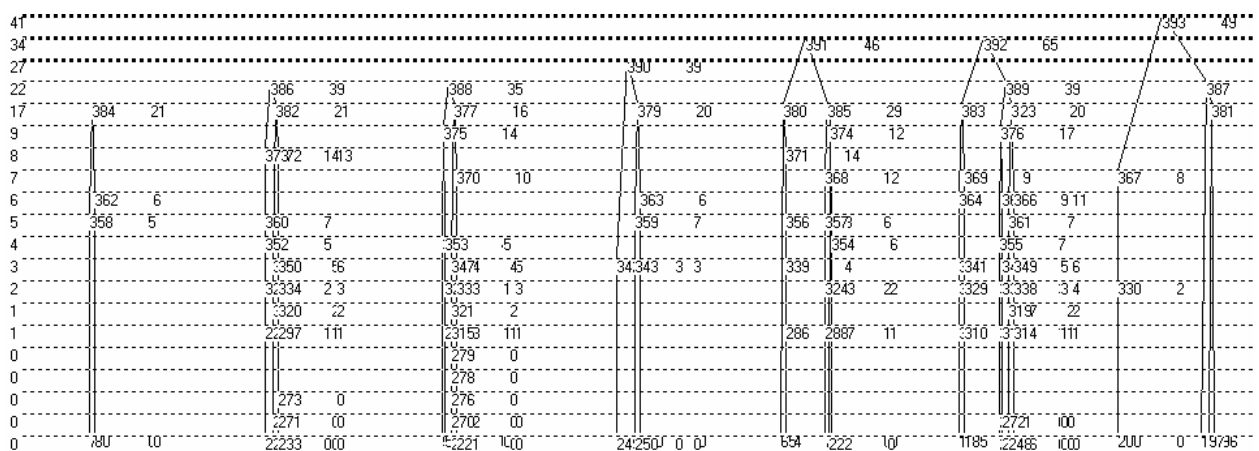


Рис. 6. Дерево кластеризації клітин зразка крові

Висновки. Проведені дослідження показали можливість застосування програмного пакета декомпозиції образів до опрацювання зображень, які містять складові частини – об'єкти вивчення фахівцями конкретних галузей, наприклад, медицини, металургії тощо. Зокрема, в роботі проілюстровано визначення якісних та кількісних показників зразка крові за часові проміжки, що не перевищують 20 секунд машинного часу.

1. Katarzyna Wilamowska, Milos Manic. *Unsupervised pattern clustering for data mining // Proceedings of the 27th Annual Conference of the IEEE Industrial Electronics Society. IECON'01. – P. 1862–1867.* 2. Katz S., Tal A. *Hierarchical mesh decomposition using fuzzy clustering and cuts // ACM Transactions on Graphics. – 2003. – Vol. 22, issue 3. – P. 954– 961.* 3. Dosil R., Pardo X.M., Fdez-Vidal X.R. *Decomposition of three-dimensional medical images into visual patterns // IEEE Transactions on*

biomedical engineering. – 2005. – Vol. 52, n. 12. – P. 2115–2121. 4. Мельник Р.А., Алексеев О.А. Кластеризація мікрообразів для кодування зображень // Пр. міжнар. конф. “Укробраз’2004”. – К., 2004. – С. 81–85. 5. Melnyk R., Tushnytskyu R. Decomposition of Visual Patterns // Досвід розробки та застосування приладо-технологічних САПР в мікроелектроніці: Матеріали ІХ Міжнар. наук.-техн. конф. CADSM 2007. – Львів. – С. 278–279. 6. Мельник Р.А., Алексеев О.А. Кластеризація ключів образів на основі декомпозиції їх множини // Відбір і обробка інформації. – 2006. – Вип. 24(100). – С. 110–114.

УДК 004.032.26:004.048

Р.О. Ткаченко, А.В. Дорошенко
Національний університет “Львівська політехніка”,
кафедра автоматизованих систем управління

ВДОСКОНАЛЕННЯ НЕЙРОМЕРЕЖНИХ МЕТОДІВ КЛАСИФІКАЦІЇ В ЗАВДАННЯХ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ ЗА ДОПОМОГОЮ МЕТОДУ ІМІТАЦІЇ ВІДПАЛУ МЕТАЛУ

© Ткаченко Р.О., Дорошенко А.В., 2007

Проаналізовано особливості постановки та підходи до розв’язання задач класифікації для випадків великорозмірних завдань інтелектуального аналізу даних. Подано основи нейромережного методу класифікації, вдосконаленого завдяки оптимізації методу імітації відпалу металу, та результати його апробації.

The article analyses the features of the target setting and the approach to solving a problem of classification task for Data Mining tasks where data are high-dimensional. Essential principles of the method of classification on the base of neural networks are proposed. This method of classification is improved by simulated annealing algorithm. The results of classification are proposed.

Вступ. Сучасні інформаційні технології надають можливість накопичувати величезні обсяги даних в усіх сферах людської діяльності, від повсякденних та ділових (дані про транзакції в супермаркетах, записи про використання кредитних карток, інформація про телефонні дзвінки та урядову статистику) до наукових (такі, як зображення астрономічних тіл, база даних молекул чи медичних записів). Враховуючи гігантські обсяги накопичених даних, було розроблено концепцію сховищ даних, що втілила досконалішу технологію запису і збереження інформації. Однак головною метою накопичення даних є не стільки їхнє збереження, скільки аналіз. З одного боку, збільшення обсягів даних надає більші можливості для отримання нової корисної інформації та знань, з іншого – зменшується можливість аналізу даних. Для точного й оперативного аналізу даних виникли нові, відмінні від традиційно статистичних, методи, об’єднані під назвою видобуток даних, або інтелектуальний аналіз даних.

Постановка задачі. В основу багатьох методів класифікації покладено гіпотезу компактності [1], яка передбачає, що об’єкти, які належать до одного класу, формують певні кластери в просторі ознак, а отже, можуть бути розділені гіперповерхнями простого вигляду. Однак у багатьох випадках для систем, представлених в умовах невизначеності, існує взаємне перекриття класів. Це спричиняється неповнотою інформаційного базису, суперечливістю даних та іншими факторами. Тому класичні методи класифікації, які ґрунтуються на основі гіпотези компактності, є неефективними, а отже, необхідно знаходити нові.

Пропонуємо розглянути поєднання методу кусково-лінійної класифікації на основі моделі геометричних перетворень (нейронні мережі типу “функціонал на множині табличних функцій”), який