

УДК 004.925.5

*Ботвін М.М. (аспірант, Державний університет інфраструктури та технологій)*

*Бутрик Н.О. (аспірант, Державний університет інфраструктури та технологій)*

*Герцій О.А., к.т.н., доц. (завідувач кафедри «Автоматизація та комп'ютерно-інтегровані технології транспорту», Державний університет інфраструктури та технологій)*

### АЛГОРИТМ КЛАСТЕРИЗАЦІЇ K-MEANS В ЗАДАЧАХ РОЗПІЗНАВАННЯ ЗОБРАЖЕНЬ

*В статті розглянуто ідентифікацію зображень на основі алгоритму кластеризації k-means. Проаналізовано принцип роботи алгоритму, проведено моделювання по сегментації теплових зображень для розпізнавання теплових характеристик досліджуваних об'єктів. Запропоновано схему реалізації системи машинного зору в процесі виявлення перегрітих букс на залізничному транспорті.*

***Ключові слова:** зображення, алгоритм, кластеризація, моделювання, сегментація, температурні показники, букса.*

**Вступ.** У даний час напрямок ідентифікації графічних зображень набув величезної популярності серед дослідників та науковців. Це пов'язано з тим, що розв'язання задач у даній області знаходить все більш прикладне застосування в повсякденній діяльності людини. В якості прикладу можна навести розпізнавання текстів, відбитків пальців, автомобільних номерів, облич та ін. Достатньо важливим фактором в популярності ідентифікації зображень є постійний розвиток теоретичного та математичного апаратів. Серед основних підходів для вирішення завдань розпізнавання варто виділити: класифікації за допомогою вирішальних функцій і функцій відстані; машини опорних векторів; нейронні мережі; різні статистичні методи та алгоритми кластеризації [1].

**Аналіз останніх досліджень і постановка проблеми.** Кластеризація чи кластерний аналіз в задачах ідентифікації являє собою процес розбиття заданої вибірки об'єктів на непересічні підмножини (кластери) так, щоб кожен кластер складався з схожих об'єктів. Подібність об'єктів визначається на основі метрики, обраної відповідно до критерію кластеризації.

Вихідні дані для процедури кластеризації – це набір об'єктів, кожен з яких задається вектором своїх характеристик. У цих алгоритмах векторами зазвичай є пікселі або околиці пікселів. Множинність векторів може містити такі компоненти [2]:

- значення інтенсивності;
- коди кольору або кольорові характеристики;

**DOI: 10.32703/2617-9040-2019-34-2-3**

- обчислені характерні ознаки;
- текстурні характеристики.

Сам алгоритм кластеризації – це функція  $a: X \rightarrow Y$ , яка будь-якому об'єкту  $x \in X$  ставить у відповідність номер кластера  $y \in Y$ . Множинність  $Y$  у деяких випадках відома заздалегідь, однак частіше ставиться завдання визначити оптимальну кількість кластерів, з точки зору того чи іншого критерію якості кластеризації.

Але рішення завдання кластеризації принципово неоднозначне, і тому є кілька причин:

- не існує однозначно найкращого критерію якості кластеризації;
- кількість кластерів, як правило, невідомо заздалегідь і встановлюється у відповідності з деяким суб'єктивним критерієм;
- результат кластеризації істотно залежить від метрики, вибір якої, як правило, також суб'єктивний і визначається експертом [3].

Така неоднозначність потребує гнучкого налаштування засобів обробки даних, що необхідно враховувати при розробці відповідних методів ідентифікації графічних зображень.

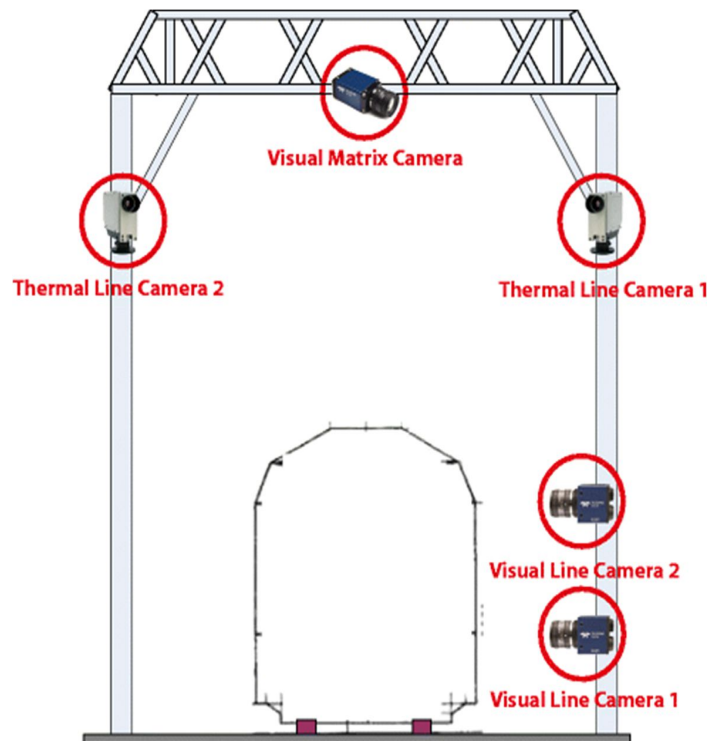
Над даним питанням в області цифрової обробки зображень працюють дослідники та науковці багатьох країн світу, зокрема, питання щодо сегментації кольорових зображень на основі кластеризації з метою ідентифікації зображень набуло великої популярності [4-6].

Так, наприклад, на залізниці для інспекції вагонів та локомотивів на предмет виявлення технічних несправностей, а саме, перегрітих букс застосовується широко відома система приладів виявлення нагрітих букс (ПОНАБ) [7]. Принцип дії даного обладнання полягає у вимірюванні температури букс в процесі проходження поїзда повз інфрачервоних датчиків. Але дана система має ряд недоліків: ПОНАБ може помилково спрацювати при перевезенні гарячих вантажів (агломерат-окатиші), при течі гарячої води з системи опалення пасажирських вагонів або при проходженні ПОНАБ в режимі гальмування фрикційним гальмом (через нагрівання гальмівних колодок) [7].

В одній із іноземних статей авторами представляється система машинного зору для ефективного моніторингу, аналізу та представлення візуальних даних, отриманих з декількох камер (рис. 1). Це рішення спрямоване на підвищення безпеки повсякденного залізничного перевезення двома способами: 1) шляхом оцінки безлічі вимог безпеки з використанням алгоритмів аналізу зображень, які можуть обробляти великі зображення поїздів, 2) допомагаючи операторам безпеки поїздів виявляти будь-які можливі несправності в поїзді. Система використовує високошвидкісні видимі і теплові камери, які спостерігають за поїздом, що проходить під залізничною рамою [8].

Тому, на заміну таких застарілих систем як ПОНАБ, можна було б застосувати системи технічного зору на базі термографічних камер [9]. Термографічні камери виявляють випромінювання в інфрачервоному діапазоні електромагнітного спектра (приблизно 0,9-14 мкм) і на основі цього випромінювання створюють зображення, що дозволяють визначити перегріті або переохолоджені об'єкти. В таких зображеннях колір являється показником рівня температури, наприклад, синій колір означає, що температура об'єкта складає 27-28 °С, фіолетовий – 29-31 °С, оранжевий – 32-34 °С, жовтий – 35-36 °С, білий – вище 36 °С і т. д.

- Переваги таких камер полягають в такому:
- можуть показувати візуальні зображення, що допомагає в порівнянні температур на великій площі;
  - дозволяють знаходити аварійні елементи до їх виходу з ладу;
  - вимірювання в областях, де інші методи неможливі (об'єкти з малою теплоємністю) або пов'язані з ризиком для здоров'я;
  - неруйнівний контроль;
  - полегшує пошук дефектів (тріщин) в колонах або інших металевих частинах;
  - можливість бачити тепло зі швидкістю 1 кадр за секунду, навіть з невисоким просторовим розширенням [10].



*Рис. 1. Багатокамерна система обробки і візуалізації зображень для оцінки безпеки руху поїздів*

Для реалізації системи машинного зору з метою виявлення перегрітих букс на залізниці на рис. 2 представлена структурна схема. Принцип дії даної системи наступний: коли до датчика S1 наближається колісна пара – він видає сигнал на систему керування CS (яка фіксує порядковий номер колісної пари) та на термографічні камери C1, C2 для старту зйомки. Камери опрацьовують термічні зображення і, в залежності від результату, видають сигнал високого рівня на CS, якщо зафіксовано перевищення температури. CS поєднує номер колісної пари із сторонністю букси, фіксує результати інспекції до бази даних та, наприклад, видає сигнал на контролер, який, в свою чергу, маркує букси фарбою, для сигнал для машиністів про необхідність зменшити швидкість і т. д.

Тобто, проблема полягає в розробці якісного алгоритму розпізнавання зображень, в даному випадку, термографічних, який водночас був би швидкодіючим та ефективним на фоні інших теплових завод, які присутні на залізниці.

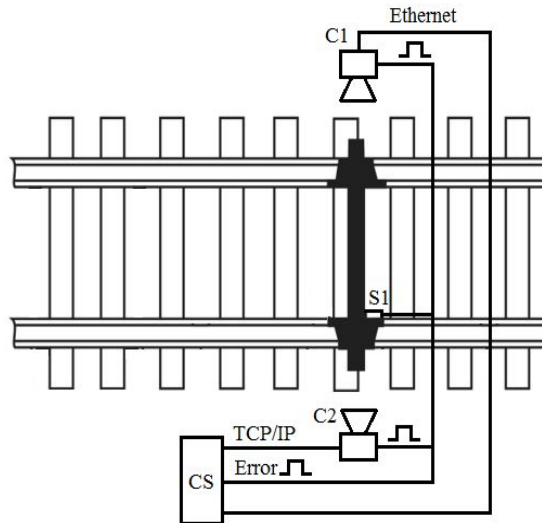


Рис. 2. Структурна схема реалізації системи машинного зору в процесі виявлення перегрітих буксів на залізничному транспорті

**Мета і завдання дослідження.** Проаналізувати принцип роботи алгоритму кластеризації k-means в задачах розпізнавання графічних зображень, провести експериментальне моделювання даного алгоритму, сегментуючи теплові зображення на основі кольору з метою розпізнавання теплових характеристик досліджуваних об'єктів в середовищі MATLAB.

**Матеріали та методи дослідження.** Алгоритм кластеризації k-means розбиває множинність елементів n-векторного простору на кількість кластерів k, в яких кожен об'єкт належить кластеру з найближчим середнім значенням. Цей метод видає рівно k різних кластерів з найбільшою можливою різницею. Найкраща кількість кластерів k, що призводить до найбільшого відокремлення, не відома і тому повинна бути обчислена з даних. Ціль кластеризації алгоритмом k-means полягає в тому, щоб мінімізувати загальну внутрішньо-кластерну дисперсію або функцію квадратичної помилки [11, 12]:

$$J = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - c_j\|^2, \quad (1)$$

де  $J$  – цільова функція;  
 $k$  – кількість кластерів;  
 $n$  – кількість спостережень;  
 $x_i$  – спостереження  $i$ ;  
 $c_j$  – центроїд для кластера  $j$ ;  
 $\|x_i^{(j)} - c_j\|^2$  – дистанційна функція.

**Опис алгоритму.** Спочатку відбувається розподілення об'єктів по кластерам. Обирається кількість кластерів  $k$ , і на першому кроці ці точки вважаються «центрами» кластерів. Кожному кластеру відповідає один центр. Вибір початкових центроїдів може здійснюватися наступним чином:

- вибір  $k$ -спостережень для максимізації початкової відстані;
- випадковий вибір  $k$ -спостережень;
- вибір перших  $k$ -спостережень.

В результаті кожен об'єкт призначений певному кластеру.

Далі відбувається ітеративний процес. Вираховуються центри кластерів, якими потім і далі вважаються покоординатні середні кластерів. Об'єкти знову перерозподіляються.

Процес обчислення центрів і перерозподілу об'єктів триває до тих пір, поки не виконана одна з умов:

- кластерні центри стабілізувалися, тобто всі спостереження належать кластеру, до якого належали до поточної ітерації;
- кількість ітерацій дорівнює максимальній кількості ітерацій.

На рис. 3 та 4 зображено принцип роботи алгоритму  $k$ -means (де  $k = 2$ ) та його блок-схему [13].

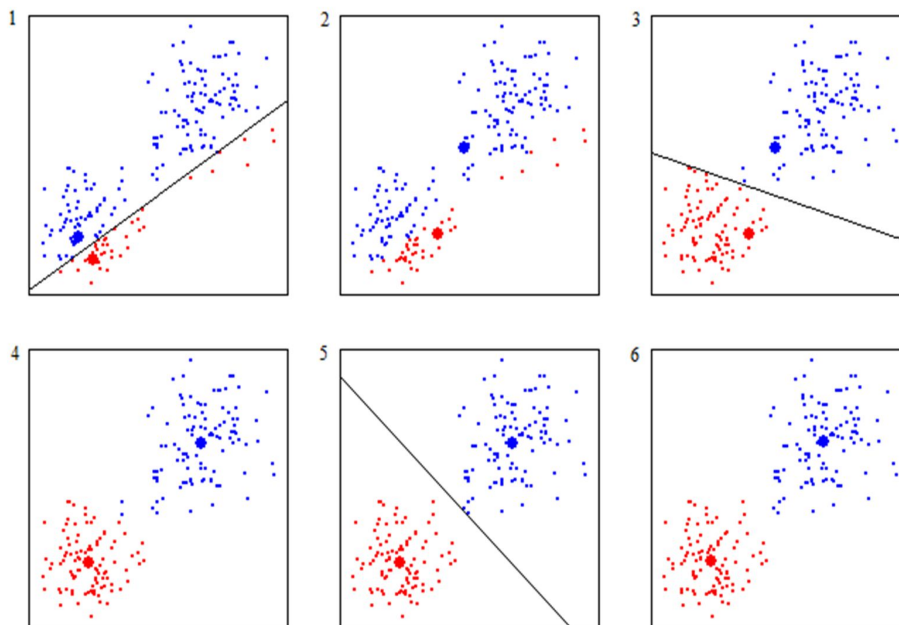


Рис. 3. Ілюстрація роботи алгоритму  $k$ -means

Вибір кількості кластерів являється складним питанням. Якщо немає припущень відносно цієї кількості, то рекомендується створювати 2 кластера, потім 3, 4, 5 і т.д. Після отримання результатів кластерного аналізу варто перевірити правильність кластеризації, тобто оцінити, наскільки кластери відрізняються один від одного. Для цього розраховуються середні значення для кожного кластера. При хорошій кластеризації повинні бути отримані середні, які сильно відрізняються для всіх вимірювань або хоча б для більшої їх частини [14].

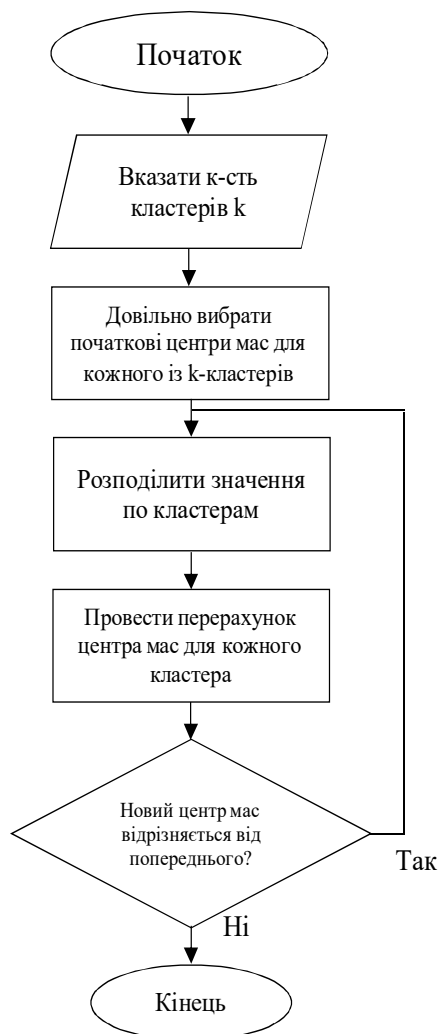
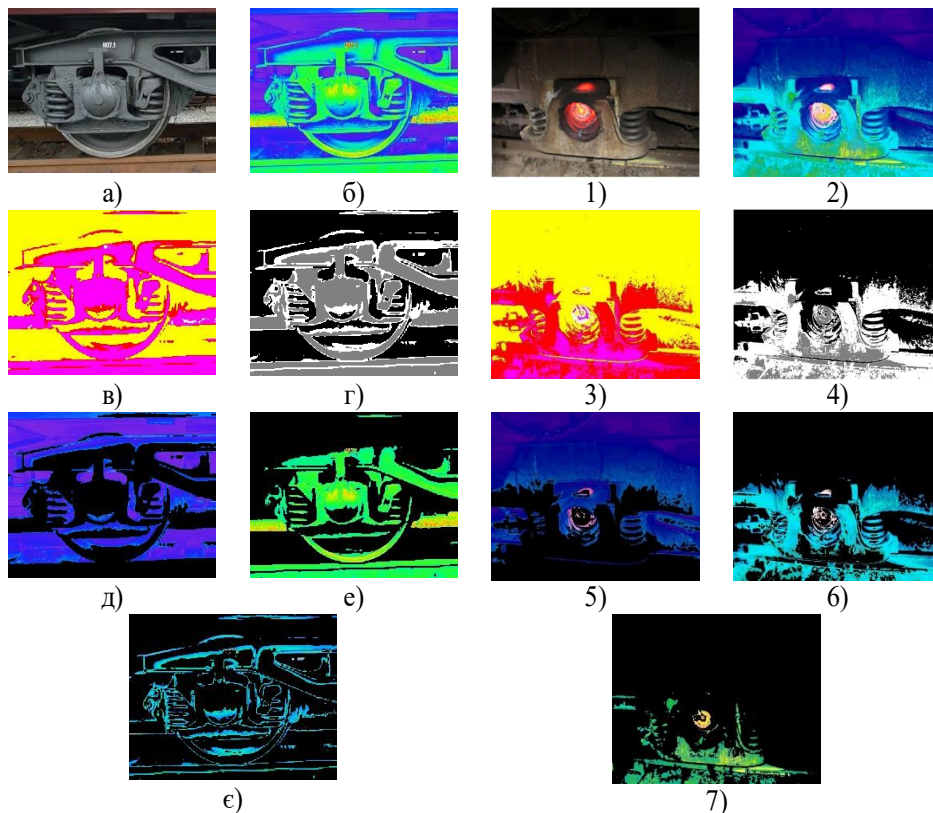


Рис. 4. Блок-схема алгоритму k-means

**Алгоритм кластеризації k-means в задачах ідентифікації образів.** Вилучення інформації із зображень та розуміння її так, що отримана інформація може бути використана для виконання декількох завдань, є важливою характеристикою машинного навчання. Сегментація зображення – це один із перших кроків у напрямку розуміння зображень, а потім знаходження в них різних об'єктів. Сегментація зображень тягне за собою поділ або розділення зображення на області аналогічних атрибутів. Це один з найважливіших компонентів аналізу зображень та розпізнавання образів, який і до цих пір, як і раніше, вважається найбільш складним завданням для обробки і аналізу зображень. Він має застосування в кількох областях, таких як аналіз віддалених зображень, медицина, моніторинг руху транспорту та розпізнавання відбитків пальців, тощо. Метод сегментації, що включає класифікаційні підходи, стикається з великими труднощами при обчисленні кількості кластерів, які присутні в просторі об'єкта чи витягуючи відповідну функцію. Цей тип сегментації зображень широко використовується

завдяки простоті розуміння і більшій точності результату. Існують різні методи, і одним з найпопулярніших методів є алгоритм кластеризації k-means. Даний алгоритм кластеризації – неконтрольований, і він використовується для сегментації області інтересу з фону, наведемо приклад використання цього алгоритму в розпізнаванні образів.

**Виявлення перегрітої букси.** Як ми зазначили раніше, на залізниці присутня проблема моніторингу технічного стану букс на предмет аварійності. Тому нами пропонується застосування системи машинного зору, яка б відслідковувала теплові характеристики рухомого складу на базі термографічних камер з використанням алгоритму неконтрольованої кластеризації k-means. Для моделювання взято зображення двох букс: нормальна букса (рис. 5 (а)) та перегріта букса (рис. 5 (1)).



*Рис. 5. Сегментація на основі кольору з використанням алгоритму кластеризації k-means: а) і 1) - умовні зображення букс; б) і 2) - ілюстрація нормальної та перегрітої букс в інфрачервоному діапазоні; в) і 3) - теплові зображення, конвертовані з кольорового простору RGB в  $L^*a^*b^*$ ; г) і 4) - зображення, позначені індексом кластера; д) і 5) - об'єкти в кластері 1; е) і 6) - об'єкти в кластері 2; є) і 7) – об'єкти в кластері 3*

Примітка стосовно зображень: оскільки відсутня можливість зробити реальні знімки букс в інфрачервоному діапазоні через високу вартість термографічних камер, прийнято рішення конвертувати звичайні RGB зображення в такий вид, який візуально створює ефект, що зображення виконано в тепловому діапазоні. Тому

зображення букси, яку ми умовно позначили як нормальну, насправді, теж може бути перегрітою, – і, нажаль, цей фактор спотворює результати моделювання.

Таким чином, для моделювання використовуються кольорові зображення нормальної та перегрітої букс для дефектної сегментації теплових показників. Дефектна сегментація виконується в два етапи. Спочатку пікселі кластеризуються на основі їх кольорних і просторових характеристик, де виконується процес кластеризації. Потім кластерні блоки об'єднуються в певну кількість регіонів. Використовуючи цю двоетапну процедуру, можна збільшити обчислювальну ефективність уникаючи виділення ознак для кожного пікселя в зображеннях букс. Хоча колір зазвичай не використовується для сегментації дефектів, він створює високу дискримінаційну здатність для різних областей зображення. Таким чином, цей підхід забезпечує надійне рішення для сегментації температурних показників. Експериментальні результати показують ефективність запропонованого підходу для поліпшення якості сегментації теплових зображень в аспектах точності і обчислювального часу. Результати моделювання показують що пропонуванний підхід є багатообіцяючим. Сегментація з використанням алгоритму k-means дуже корисна для аналізу зображення. Важливою метою сегментації зображення є поділ об'єкта і фону, незалежно від того, чи має зображення кордон розмиття. Дефектну сегментацію теплових характеристик можна розглядати як приклад сегментації зображення, в якому кількість сегментації не відомо. Основна мета запропонованого підходу – автоматичне сегментування кольорів з використанням методу кластеризації k-means і кольорного простору  $L^*a^*b^*$ .

Введена структура сегментації працює в п'ять кроків наступним чином:

- 1) Зчитування зображень в інфрачервоному діапазоні.
- 2) Конвертація зображення з кольорового простору RGB в  $L^*a^*b^*$ . Використовується кольорний простір  $L^*a^*b^*$ , тому що він складається з шару світності в каналі «L» і двох шарів кольоровості в каналах «a\*» і «b\*». Застосування кольорного простору  $L^*a^*b^*$  є безкомпромісним, тому що вся кольорна інформація присутня в «a\*» і «b\*».
- 3) Класифікація кольору за допомогою алгоритму кластеризації k-means в просторі « $a^*b^*$ ». Щоб виміряти різницю між двома кольорами, використовується метрика евклідової відстані.

4) Позначення кожного пікселя в зображенні за допомогою результату алгоритму k-means. Для кожного пікселя на нашому вході алгоритм обчислює індекс, що відповідає кластеру. Кожен піксель зображення буде позначений індексом кластера.

5) Створення зображень, які сегментують вхідному зображенню за кольором. Ми повинні відокремити пікселі в зображенні за кольором за допомогою міток пікселів, що призведе до різних зображень на основі кількості кластерів. Програмуємо визначення індексу кожного кластера, що містить завищені показники температури, тому що алгоритм не повертає одне і теж значення індексу кластеру кожен раз. Але ми можемо зробити це, використовуючи центральне значення кластерів, яке містить середнє значення «a\*» і «b\*» для кожного кластера.

**Висновки.** Алгоритм кластеризації k-means відноситься до неконтрольованих алгоритмів кластеризації даних (неконтрольоване навчання). До переваг даного алгоритму належать: зрозумілість, простота та швидкість використання.



Недоліками алгоритму є повільна робота з великими базами даних та чутливість до викидів, які можуть спотворювати середнє значення.

Запропоновано схему реалізації системи машинного зору в процесі виявлення перегрітих букс на залізничному транспорті з використанням алгоритму кластеризації k-means.

Експериментальні результати моделювання показали, що сегментація на основі кольору з використанням алгоритму кластеризації k-means дозволяє точно сегментувати температурні показники рухомого складу. Так, зображення на рис. 5 е) відображає сегментацію об'єктів в другому кластері й, в порівнянні з умовним тепловим зображенням (рис. 5 б)), представляє розпізнавання домінантної частки зеленого кольору, що свідчить про задовільний температурний стан букси, чого не скажеш про зображення перегрітої букси (рис. 5 б)), де білий колір сигналізує про надмірне перевищення температури, що свідчить про аварійний стан букси.

Слід зазначити, що застосування даного алгоритму в процесі створення програмного забезпечення для термографічних камер з метою розпізнавання температурних показників є ефективним, оскільки цей алгоритм дозволяє стискати зображення [15-17], що призведе до збільшення швидкості опрацювання даних та дасть можливість проводити моніторинг теплових характеристик рухомого складу на вищих швидкостях.

#### ЛІТЕРАТУРА

1. *Татьянкін В. М.* Спосіб ідентифікації образів. // Шостий технологічний уклад: механізми та перспективи розвитку. 2015. С. 79-81. URL: <https://cyberleninka.ru/article/v/sposob-identifikatsii-obrazov>. (дата звернення 25.10.2019).
2. *Селянкін В. В., Скороход С. В.* Анализ и обработка изображений в задачах компьютерного зрения: учебное пособие. Таганрог: Изд-во ЮФУ, 2015. 82 с.
3. *Яшина Е. С.* Использование метода кластеризации в информационно-аналитической системе / Е. С. Яшина, М. А. Щербак // Радиоэлектронні і комп'ютерні системи. 2016. № 2 (76). С. 62-69. URL: [http://nbuv.gov.ua/UJRN/recs\\_2016\\_2\\_11](http://nbuv.gov.ua/UJRN/recs_2016_2_11). (дата звернення 25.10.2019).
4. *Neelambike S.* Color image segmentation by clustering // International journal of advanced research in computer science & technology. 2014. Vol. 2. № 1. P. 95-97.
5. *Aqil Burney S.M.* K-means cluster analysis for image segmentation // International Journal of Computer Applications. 2014. Vol. 96. № 4. P. 8.
6. *Nameirakram Dh.* Image segmentation using k-means clustering algorithm and subtractive clustering algorithm. Eleventh international multi-conference on information processing-2015 (IMCIP-2015). India. P. 764 – 771.
7. Прибор обнаружения неисправных (нагретых) аварийных букс (ПОНАБ). // Учебно-образовательный портал «Все лекции». URL: <http://vse-lekcii.ru/zheleznodorozhnyj-transport/ats/ponab>. (дата звернення 25.10.2019).
8. *Karatan S., Pezzatini D., Bimbo A.* A multi-camera image processing and visualization system for train safety assessment. // Multimedia Tools and Applications. 2018. Vol. 77. № 2. P. 1583-1604.
9. Термографія. // Файловий архів студентів: Фізико-технічні основи рентгенології. URL: <https://studfile.net/preview/1785100/page:5/>. (дата звернення 25.10.2019).
10. Принципы получения термографического изображения. // Файловый архив студентов: Лекция 8 «Принципы получения термографического изображения». URL: <https://studfile.net/preview/2716754/>. (дата звернення 25.10.2019).
11. K-Means Clustering. // An Introduction to Data Science: Modeling. URL: [http://www.saedsayad.com/clustering\\_kmeans.htm](http://www.saedsayad.com/clustering_kmeans.htm). (дата звернення 25.10.2019).
12. K-Means Clustering. // A Tutorial on Clustering Algorithms: K-means. URL: [https://home.deib.polimi.it/matteucc/Clustering/tutorial\\_html/kmeans.html](https://home.deib.polimi.it/matteucc/Clustering/tutorial_html/kmeans.html). (дата звернення 25.10.2019).
13. *Madushan D.* Introductions to K-means Clustering. URL: <https://medium.com/@dilekamadushan/introduction-to-k-means-clustering-7c0ebc997e00>. (дата звернення 25.10.2019).

14. Методы кластерного анализа. Итеративные методы. // Интуит: Национальный открытый университет (Курс лекций). URL: <https://www.intuit.ru/studies/courses/6/6/lecture/184>. (дата звернення 25.10.2019).
15. Wu Wenbin, Yue Wu and Jintao Li. The Hyper-spectral Image Compression Based on K-Means Clustering and Parallel Prediction Algorithm. // International Conference on Smart Materials, Intelligent Manufacturing and Automation (SMIMA 2018), section: Digital Signal and Image Processing. 2018. Vol. 173: MATEC Web of Conferences. P. 4.
16. Karri Tata Rao, Ch. Ramesh. Baseline JPEG Image Compression with K-Means Clustering Based Algorithm. // International Journal of Computer Science and Mobile Computing. Vol. 4. № 4. 2015. P. 379-386.
17. Tichun Wang, Hongyang Zhang, Lei Tian. Research of Computer Desktop Image Compression Clustering Algorithm. // CHEMICAL ENGINEERING TRANSACTIONS. 2016. Vol. 51. P. 1-6.

## REFERENCES

1. Tatyankin V.M. (2017). *Sposob identifiacui obrazov. [Method of image identification]*. The Sixth Technological Structure: Mechanisms and Prospects of Development, 79-81. Retrieved from <https://cyberleninka.ru/article/v/sposob-identifikatsii-obrazov>. (Accessed 25 October 2019).
2. Selyankin V.V., Skorokhod S.V. (2015). *Analyze i obrabotka izobrajenie v zadachah computernoho zreniya [Analysis and processing of images in computer vision problems]*. Taganrog, SFedU Publ., 82.
3. Yashina E.S. (2016) *Ispolzovanie metoda clusterizacii v informacionno-analiticheskoi sisteme [The use of the clustering method in the information-analytical system]*. *Radio electronic and computer systems*, 76, 2, 62-69. Retrieved from [http://nbuv.gov.ua/UJRN/recs\\_2016\\_2\\_11](http://nbuv.gov.ua/UJRN/recs_2016_2_11). (Accessed 25 October 2019).
4. Neelambike S. (2014). *Color image segmentation by clustering*. International journal of advanced research in computer science & technology, 2, 1, 95-97.
5. Aqil Burney S.M. (2014). *K-means cluster analysis for image segmentation*. International Journal of Computer Applications (0975 – 8887), 96, 4, 8.
6. Nameirakpam Dh. (2015). *Image segmentation using k-means clustering algorithm and subtractive clustering algorithm*. Eleventh international multi-conference on information processing-2015 (IMCIP-2015), 764 – 771.
7. *Pribor obnaruzeniia neispravnyh (nagretyh) avaryinyh buks. [The device for detecting faulty (heated) emergency axle boxes (DDFEAB)]*. Educational portal: All Lectures. Retrieved from <http://vselekcii.ru/zheleznodorozhnyj-transport/ats/ponab>. (Accessed 25 October 2019).
8. Svebor Karaman, Daniele Pezzatini, Alberto Del Bimbo (2018). *A multi-camera image processing and visualization system for train safety assessment*. Multimedia Tools and Applications, 77, 2, 1583–1604.
9. *Termografiya [Thermography]*. Students archive: Physical and technical basics of radiology. Retrieved from <https://studfile.net/preview/1785100/page:5/>. (Accessed 25 October 2019).
10. *Printsupy polucheniya termographicheskogo izobrazeniya [Thermographic imaging principles]*. Students archive: Lecture 8 «Thermographic imaging principles». Retrieved from <https://studfile.net/preview/2716754/>. (Accessed 25 October 2019).
11. K-Means Clustering. // An Introduction to Data Science: Modeling. Retrieved from [http://www.saedsayad.com/clustering\\_kmeans.htm](http://www.saedsayad.com/clustering_kmeans.htm). (Accessed 25 October 2019).
12. K-Means Clustering. // A Tutorial on Clustering Algorithms: K-means. Retrieved from [https://home.deib.polimi.it/matteucc/Clustering/tutorial\\_html/kmeans.html](https://home.deib.polimi.it/matteucc/Clustering/tutorial_html/kmeans.html). (Accessed 25 October 2019).
13. Madushan D. *Introductions to K-means Clustering*. Retrieved from <https://medium.com/@dilekamadushan/introduction-to-k-means-clustering-7c0ebc997e00>. (Accessed 25 October 2019).
14. *Metody clasternogo analiza. Interaktyvnyye metody. [Cluster analysis methods. Iterative methods]*. Intuit: National Open University (Lecture Course). Retrieved from <https://www.intuit.ru/studies/courses/6/6/lecture/184>. (Accessed 25 October 2019).
15. Wu Wenbin, Yue Wu and Jintao Li (2018). *The Hyper-spectral Image Compression Based on K-Means Clustering and Parallel Prediction Algorithm*. MATEC Web of Conferences: International Conference on Smart Materials, Intelligent Manufacturing and Automation (SMIMA 2018), section: Digital Signal and Image Processing, 173, 00004. doi: 10.1051/mateconf/201817303070.
16. Karri Tata Rao, Ch. Ramesh (2015). *Baseline JPEG Image Compression with K-Means Clustering Based Algorithm*. International Journal of Computer Science and Mobile Computing, 4, 379-386.
17. Tichun Wang, Hongyang Zhang, Lei Tian (2016). *Research of Computer Desktop Image Compression Clustering Algorithm*. CHEMICAL ENGINEERING TRANSACTIONS, 51, 1-6.

*Ботвин М.М.,  
(аспирант, Государственный университет инфраструктуры и технологий)  
Бутрик Н.О.,  
(аспирант, Государственный университет инфраструктуры и технологий)  
Герций А.А., к.т.н., доц.  
(заведующий кафедры «Автоматизация и компьютерно-интегрированные технологии транспорта», Государственный университет инфраструктуры и технологий)*

### **АЛГОРИТМ КЛАСТЕРИЗАЦИИ K-MEANS В ЗАДАЧАХ РАСПОЗНАВАНИЯ ИЗОБРАЖЕНИЙ**

*В статье рассмотрена идентификация изображений на основе алгоритма кластеризации k-means. Проанализирован принцип работы алгоритма, проведено моделирование по сегментации тепловых изображений для распознавания тепловых характеристик исследуемых объектов. Предложена схема реализации системы машинного зрения в процессе выявления перегретых букс на железнодорожном транспорте.*

*Ключевые слова:* изображение, алгоритм, кластеризация, моделирование, сегментация, температурные показатели, букса.

*Maksym Botvin,  
(PhD-student, State University of Infrastructure and Technology)  
Nataliia Butryk,  
(PhD-student, State University of Infrastructure and Technology) Oleksandr Gertsii,  
PhD (Technical Sciences), Associate Professor, (Head of the Department of Automation and Computer-Integrated Transport  
Technologies, State University for Infrastructure and Technology)*

### **K-MEANS CLUSTERING ALGORITHM IN IMAGE RECOGNITION TASKS**

*The article is an overview of the scientific direction of graphic image recognition based on the k-means clustering algorithm. In particular, we have analyzed the clustering of data in digital image identification problems; we have described the operation principle of this algorithm with an illustration, where the number of selected clusters is equal to two and we presented a block diagram of the algorithm. We have proposed a scheme for implementing a machine vision system using this algorithm in the process of detecting overheated sockets on rail. We have performed color-based segmentation simulations over infrared images to detect high temperature sections in MATLAB. More specifically: two conditional images are taken, showing the railroad socket, one of which is normal and the other is overheated. The simulation results show that color-based segmentation allows accurate segmentation of thermal imaging. The application of this algorithm in the process of creating software for thermographic cameras for the purpose of recognizing temperature is effective, because this algorithm enables image compression, speeds up data processing, and enables the thermal performance of railroad socket to be monitored at high speeds.*

*Keywords:* image, algorithm, clustering, simulation, segmentation, temperature, railroad socket.