

ШАМУРАТОВ О. Ю.

Національний університет "Львівська Політехніка"

<https://orcid.org/0000-0003-1913-5362>e-mail: [oleksii.y.shamuratov@lpnu.ua](mailto:oleksii.y.shamuratov@lpnu.ua)

## МЕТОД КЛАСТЕРИЗАЦІЇ ОБ'ЄКТІВ НА ЗОБРАЖЕННІ НА ОСНОВІ ВИБОРУ ОЗНАК

У статті описується розробка методу, що дозволяє створювати кластери на основі вибору ознак. У сучасному світі індустрія розваг в Інтернеті швидко розвивається, створюючи попит на більш якісні продукти. Це в свою чергу призвело до використання штучного інтелекту не тільки в науці, а й у розвагах. На даний момент набирають популярності програми, що дозволяють створювати анімацію об'єктів на фотографіях. У цій статті представлений підхід до вирішення проблеми визначення об'єктів для анімації

Ключові слова: кластеризація, вибір ознак, анімація

Oleksiy SHAMURATOV

Lviv Polytechnic National University

## OBJECT CLUSTERIZATION METHOD IN PICTURES BASED ON FEATURE SELECTION

The article describes the development of a method that allows you to create clusters based on selecting feature features. In today's world, the entertainment industry on the Internet is developing rapidly, creating a demand for better products. This factor has led to the use of artificial intelligence not only in science but also in entertainment. Currently, applications that allow you to create animations of objects in photos are gaining popularity. This article presents an approach to solving the problem of defining objects for animation. To classify and further identify objects, their characteristics should be determined. This is one of the options for abstraction, in which the input set of properties of the object is reduced to the minimum required number of features by which you can identify the object.

The algorithm can be used to determine the main features of objects, such as area and perimeter, radii of inscribed and circumscribed circles, sides of the described rectangle, number and relative position of angles, gradient of the object histogram. Based on these features clustering and classification of the image are implemented. The artificial neural network was trained on image samples, each class contained from 2528 to 16185 images of 64x64 pixels. 1000 images of objects of each class were then selected for testing. The success of recognition based on a convolutional neural network was evaluated. According to the results, we can conclude that the smaller the invariance of the class, the greater the accuracy of recognition. The amount of data in the training sample has little effect on the accuracy of the algorithm. After calculating the intensity gradient, you should divide the image into a cell and build a histogram of the gradient object for each pocket of cell; the histogram module corresponds to the intensity gradient at the point.

Keywords: clustering, feature selection, animation

### Постановка проблеми у загальному вигляді та її зв'язок із важливими науковими чи практичними завданнями

У сучасному світі індустрія розваг в Інтернеті швидко розвивається, створюючи попит на більш якісні продукти. Це в свою чергу призвело до використання штучного інтелекту не тільки в науці, а й у розвагах. На даний момент набирають популярності програми, що дозволяють створювати анімацію об'єктів на фотографіях. У цій статті представлений підхід до вирішення проблеми визначення об'єктів для анімації.

### Аналіз досліджень та публікацій

Для класифікації та подальшої ідентифікації об'єктів необхідно спочатку визначити їх ознаки. Це один з варіантів абстракції, при якому вхідний набір властивостей об'єкта зводиться до мінімально необхідної кількості ознак, за якими можна буде ідентифікувати об'єкт. Вибір ознак починається з набору вхідних даних, який поділяється на первинний і вторинний. Передбачається, що первинні ознаки мають бути інформативними і не повинні бути зайвими [1]. Такий підхід сприяє прискоренню процесу навчання алгоритмам розпізнавання об'єктів та більш точної їх роботи. При аналізі складних об'єктів однією з головних проблем є кількість змінних. Цей підхід споживає більше обчислювальних ресурсів, а також може спричинити колізію щодо вибірки навчальних даних, що призведе до зниження точності розпізнавання. Ізоляція ознак є основним методом побудови комбінацій змінних, щоб уникнути цих проблем, але цього достатньо, щоб точно описати об'єкт. Правильно оптимізований розподіл функцій є ключем до побудови ефективної моделі [2].

Після визначення ознак об'єкта необхідно буде кластеризувати і класифікувати дані, передавати об'єкти в екземпляр. Зразок означає назву місцевості в просторі знаків, на якій відображено багато предметів або явищ матеріального світу. Ознака — це кількісний опис властивості предмета чи явища [3]. Простір ознак — це N-вимірний простір, визначений для даної задачі розпізнавання, де N — фіксована кількість вимірних ознак для будь-якого об'єкта. Вектор із простору ознак  $x$ , що відповідає об'єкту завдання розпізнавання, є N-вимірним вектором з компонентами  $(x_1, x_2, \dots, x_N)$ , які є значеннями ознак для цього об'єкта. Іншими словами, розпізнавання образів можна визначити як віднесення вихідних даних до

певного класу шляхом вибору істотних ознак або властивостей, що характеризують ці дані, із загальної маси несуттєвих деталей [4].

Проаналізувавши загальну інформацію про класифікацію, можна описати формальне завдання класифікації. Дано багато об'єктів. Їх потрібно класифікувати. Множина представлена підмножинами, які називаються класами. Вказано: відомості про класи, опис усього набору та опис інформації про об'єкт, належність якого до певного класу невідома. На основі наявної інформації про класи та опису об'єкта необхідно визначити, до якого класу належить цей об'єкт [5].

Проблема кластеризації схожа на проблему класифікації, але відмінність між методом полягає в тому, що класи не визначені. Кластеризація не робить ніяких статистичних висновків, але дає можливість проаналізувати структуру даних, тобто мету кластеризації. йде пошук подібних конструкцій.

Штучні нейронні мережі є одним з найбільш ефективних і поширених способів представлення та вирішення проблем розпізнавання зображень. Нейронні мережі дуже добре працюють у задачах розпізнавання образів, оскільки поєднують математичні та логічні обчислення. Нейронна мережа дозволяє обробляти велику кількість факторів, незалежно від їх походження, це стабільний універсальний алгоритм. Нейронні мережі дозволяють будувати залежності параметрів у вигляді полінома на основі навчальної вибірки, що дуже спрощує реалізацію розпізнавання об'єктів [6].

### Виклад основного матеріалу

Щоб знайти об'єкт у зображенні, база даних об'єктів повинна існувати. Наповнення бази планується за рахунок користувачів системи, які будуть відбирати об'єкти для подальшої анімації. Для таких об'єктів система повинна вибрати ознаки, класифікувати об'єкт і зберегти його в базі даних. База даних класифікованих об'єктів у майбутньому використовуватиметься для пошуку схожих об'єктів на зображенні та дозволить користувачеві автоматично вибрати об'єкти для подальшої анімації.

Тому визначення об'єкта на зображенні має бути стійким до повороту об'єкта, масштабу зображення та різноманітності ракурсів зображення. Вибираються такі ознаки об'єкта:

- Визначення площі та периметра;
- Визначення радіусів вписаного та описаного кіл;
- Означення сторін описаного прямокутника;
- Визначення кількості та взаємного розташування кутів;
- Побудова гістограми градієнтів об'єктів.

Стандартна формула для розрахунку площі об'єкта розраховується шляхом підрахунку кількості елементів, пов'язаних з об'єктом. У разі надходження зображень до системи площа може бути розрахована за кількістю пікселів у області зображення.

$$S = \sum_x \sum_y \begin{cases} 1, A(x, y) \in O \\ 0, A(x, y) \notin O \end{cases}$$

де  $O$  — множина масиву пікселів  $A(x, y)$ , що належать об'єкту [7].

Щоб знайти периметр об'єкта  $P$ , достатньо порахувати кількість пікселів, що належать контуру об'єкта. Для забезпечення інваріантності до масштабу об'єкта додається нормований знак  $U$ .

$$U = \frac{A}{P^2}$$

Визначення радіусів вписаного та описаного кіл. Спочатку потрібно визначити геометричний центр об'єкта.

$$C(x, y) = \left( \frac{\sum_x \sum_y x A(x, y)}{\sum_x \sum_y A(x, y)}, \frac{\sum_x \sum_y y A(x, y)}{\sum_x \sum_y A(x, y)} \right),$$

де  $x$  і  $y$  — номери рядків і стовпців усіх пікселів  $A(x, y)$ , що належать об'єкту.

На цьому етапі можна визначити радіуси вписаного та описаного кіл. Це робиться шляхом вибору мінімальної та максимальної довжини вектора, який починається в центрі зображення і закінчується в точках всередині периметра.

$$R_{max} = \max_{0 \leq i \leq N} \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2},$$

$$R_{min} = \min_{0 \leq i \leq N} \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2},$$

де  $x_c$  і  $y_c$  — координати центру,  $N$  — кількість пікселів по периметру,  $(x_i, y_i) \in P$  [8]. Нормований параметр  $R$  додається до формули.

$$R = \frac{R_{max}}{R_{min}}$$

Визначення сторін описаного прямокутника. Визначте максимальне та мінімальне значення абсцис та ординат зображення об'єкта  $x_{max}$  і  $x_{min}$ ,  $y_{max}$  і  $y_{min}$ , а потім визначте висоту  $H$  і довжину  $L$  прямокутника:

$$L = \max_{0 \leq i \leq N} x_i - \min_{0 \leq i \leq N} x_i;$$

$$H = \max_{0 \leq i \leq N} y_i - \min_{0 \leq i \leq N} y_i,$$

де  $N$  — кількість пікселів по периметру, а  $(x_i, y_i) \in P$  [9].

Визначення кількості та взаємного розташування кутів. Для цього оцініть відстань  $l$  між початковою і кінцевою точками фрагмента контуру.

$$l_i = x_i |x_{i-2} - x_{i+2}| + |y_{i-2} - y_{i+2}|.$$

Потім перевіряється умова  $l_i \leq N$ , де  $N$  — умовне порогове значення. Вам потрібно буде реалізувати його визначення на основі минулих властивостей об'єкта. Якщо умова виконується, то точка належить до множини кутових точок [10].

Побудова гістограми градієнта об'єкта. Основною ідеєю є припущення, що зовнішній вигляд і форму об'єктів в області цифрового зображення можна описати розподілом градієнтів інтенсивності. Розраховується напрямок градієнта кожного пікселя. Для двовимірної функції  $f(x,y)$  вектор градієнта має такий вигляд:

$$\nabla f = \left[ \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right].$$

Часткові похідні функції інтенсивності в координатах  $x$  і  $y$  є оцінками контрасту в напрямку відповідних осей координат [11]. В якості точної оцінки контрасту в напрямку відповідної осі координат можна усереднювати три різні оцінки контрасту навколо пікселя формул:

$$\nabla_x = \frac{\partial f}{\partial x} \approx \frac{1}{3} \left( \frac{f[x+1,y]-f[x-1,y]}{2} + \frac{f[x+1,y-1]-f[x-1,y-1]}{2} + \frac{f[x+1,y+1]-f[x-1,y+1]}{2} \right),$$

$$\nabla_y = \frac{\partial f}{\partial y} \approx \frac{1}{3} \left( \frac{f[x,y+1]-f[x,y-1]}{2} + \frac{f[x-1,y+1]-f[x-1,y-1]}{2} + \frac{f[x+1,y+1]-f[x+1,y-1]}{2} \right).$$

Для оцінки даних дотримуйтесь мінімізації зображення маски оператора контрасту об'єкта. Для зниження обчислювальної складності найкраще використовувати диференціально-вимірну маску [11].

Після підрахунку градієнта інтенсивності слід розділити зображення на комірку та побудувати гістограму об'єкта градієнтів для кожного кишені осередку  $G$  модуль гістограми відповідає градієнту інтенсивності в точці.

$$G = \sqrt{\nabla_x^2 + \nabla_y^2}.$$

Потім проводять згладжування гістограм у клітинках, щоб забезпечити інваріантність відносної зміни гніву та контрасту.

Виділення ознак об'єкта. Щоб знайти подібність між об'єктами, будемо обчислювати семантичну відстань між ознаками.

$$D_{ij} = \sqrt{\sum_{k=1}^N (X_{ik} - X_{jk})^2},$$

де  $D_{ij}$  - евклідова відстань між точками  $i$  і  $j$ , які характеризують одну ознаку різних об'єктів.  $N$  - найменша кількість значень, що враховуються для ознаки між двома об'єктами,  $k$  - порядковий номер значення ознаки.

Процедура визначення кількості кластерів складається з наступних кроків:

1. Запускається алгоритм  $k$ -середніх для  $K$  кластерів і відповідна внутрішня дисперсія  $\delta_k = \sum_{i=1}^N \sum_{x \in c_i} \|x - m_i\|^2$ .  $x$  - вектор, що характеризує об'єкт, що входить до кластера  $c_i$ , загальна відстань між об'єктами та центром кластера. Параметр  $m_i$  є центром поточного кластера. Для різних значень  $K$  будуються множини  $\delta_k$ .

2. Вибір ступеня перетворення  $Y = \frac{p}{2}$ .

3. Обчисліть стрибки за формулою  $J_k = \delta_k^{-Y} - \delta_{k-1}^{-Y}$ .

4. Для остаточної кількості кластерів виберіть значення, що дорівнює максимальному [12].

• Розпізнавання об'єктів. Архітектура згорткової нейронної мережі була обрана для пошуку об'єктів на зображенні з наступних причин:

- зображення мають великий розмір;
- велика кількість параметрів і класів об'єктів;
- незмінність до зміни масштабу зображення, кутів зйомки камери та інших геометричних спотворень вхідного сигналу [13].

Структура нейронної мережі:

• Вхідний шар. Буде містити вхідне значення пікселів зображення, розмірність шару [64x64x3].

• Згортковий шар. Розмір фільтра 7x7x3. Аналіз буде виконано на 5 об'єктах, тому шар складатиметься з 5 площин з розмірами [62x62x1].

• Шар підшару. Розмір маски 2x2. Він складається з 5 площин розміром [31x31x1].

• Вихідний шар. Розмір [1x1xN], де  $N$  - кількість кластерів у базі даних.

В якості функції активації обрано гіперболічний тангенс:

$$f(a) = A \cdot \tanh(S_a),$$

де  $f(a)$  - значення шуканого елемента,  $a$  - зважена сума сигналів попереднього шару,  $A$  - амплітуда цієї функції,  $S$  - її положення відносно опорної точки [14].

Ця функція має ряд переваг для вирішення проблеми:

- симетричні функції активації, такі як гіперболічний тангенс, забезпечують швидшу конвергенцію, ніж стандартна логістична функція;
- функція має неперервну першу похідну;

- функція має просту похідну, яку можна обчислити через її значущість, що забезпечує меншу обчислювальну складність.

### Результати

Для перевірки класифікації було використано 5 різних класів об'єктів. Зразки взяті з відкритого доступу проекту TensorFlow [15]. Штучна нейронна мережа була навчена на зразках зображень, кожен клас містив від 2528 до 16185 зображень розміром 64x64 пікселі. Потім для тестування було відібрано 1000 зображень об'єктів кожного класу. Результати наведені в таблиці 1.

Таблиця 1.

Результат класифікації та розпізнавання об'єктів

Назва класу	Кількість об'єктів у навчальній вибірці	Середній відсоток успішного визнання
Автомобілі	16185	83%
Квіти	3670	68%
Яблука	2528	85%
Листя	6732	71%
Коти	12361	85%

Отже, найнижчий результат був у класі «Квіти» через те, що навчальна вибірка була невеликою, а зображення об'єктів були квітами більш ніж одного виду. У вибірці класу «Автомобілі» були також автомобілі різних марок і типів кабін, але через велику кількість зображень у вибірці підсумковий результат тесту був вищим. Найвищий результат отримано для класу «Яблука». Це пояснюється тим, що зразок містив об'єкт, структура та форма якого мало відрізнялися.

### Висновки з даного дослідження і перспективи подальших розвідок у даному напрямі

У даній роботі представлено алгоритм, за допомогою якого можна визначити основні ознаки об'єктів, такі як площа та периметр, радіуси вписаного та описаного кіл, сторони описаного прямокутника, кількість та взаємне розташування кутів, градієнт гистограми об'єкта. На основі цих ознак проводити кластеризацію та класифікацію зображення. Оцінювалась успішність розпізнавання на основі згорткової нейронної мережі. За результатами можна зробити висновок, що чим менше інваріантність класу, тим більша точність розпізнавання. Кількість даних у навчальній вибірці мало впливає на точність алгоритму.

### Література

1. Tien D. B., Ching Y. S., Zi-Cai L., Yuan Y. T., "Computer Transformation of Digital Images and Patterns", p. 276, 1989.
2. Y.-Q. Wang, "An Analysis of Viola-Jones Face Detection Algorithm", IPOL Journal, 2013.
3. Khan H. Abdullah, M. Shamian Bin Zainal, «Efficient eyes and mouth detection algorithm using combination of viola jones and skin color pixel detection» International Journal of Engineering and Applied Sciences, № Vol. 3 № 4, 2013.
4. V. Gaede и O. Gunther, "Multidimensional Access Methods", ACM Computing Surveys, pp. 170-231, 1998.
5. S. Khan, H. Rahmani, Syed Afaq Ali Shah, M. Bennamoun, G. Medioni, S. Medioni, "A Guide to Convolutional Neural Networks for Computer Vision", Morgan & Claypool, p. 207, 2018.
6. Sibte ul Hussain, "Machine Learning Methods for Visual Object Detection". p. 160, 2012.
7. P. Arabie, L. J. Hubert, G. De Soete, "Clustering and Classification", p. 500, 1996.
8. D. Parks, "Object Detection and Analysis: A Coherency Filtering Approach", p. 172, 2008.
9. Yongqiang Z., Chen Y., Seong G. K., Quan P., Yongmei C., "Multi-band Polarization Imaging and Applications" 1st ed., p. 204, 2016.
10. Manikandan S., "Vision Based Assistive System for Label and Object Detection", p. 64, 2015.
11. Salma H., "Object Detection Using Histogram Of Gradients", p. 52, 2018.
12. Wu J., "Advances in K-means Clustering: A Data Mining Thinking", Springer Science & Business Media, p. 180, 2021.
13. J.Loy, "Neural Network Projects with Python: The ultimate guide to using Python to explore the true power of neural networks through six projects", Packt Publishing, p. 308, 2019.
14. Brannon W. C., "Object Detection in Low-spatial-resolution Aerial Imagery Using Convolutional Neural Networks", p. 79, 2019.
15. Dataset <https://knowyourdata-tfds.withgoogle.com>

### References

1. Tien D. B., Ching Y. S., Zi-Cai L., Yuan Y. T., "Computer Transformation of Digital Images and Patterns", p. 276, 1989.
2. Y.-Q. Wang, "An Analysis of Viola-Jones Face Detection Algorithm", IPOL Journal, 2013.
3. Khan H. Abdullah, M. Shamian Bin Zainal, «Efficient eyes and mouth detection algorithm using combination of viola jones and skin color pixel detection» International Journal of Engineering and Applied Sciences, № Vol. 3 № 4, 2013.
4. V. Gaede и O. Gunther, "Multidimensional Access Methods", ACM Computing Surveys, pp. 170-231, 1998.

- 
5. S. Khan, H. Rahmani, Syed Afaq Ali Shah, M. Bennamoun, G. Medioni, S. Medioni, "A Guide to Convolutional Neural Networks for Computer Vision", Morgan & Claypool, p. 207, 2018.
  6. Sibte ul Hussain, "Machine Learning Methods for Visual Object Detection". p. 160, 2012.
  7. P. Arabie, L. J. Hubert, G. De Soete, "Clustering and Classification", p. 500, 1996.
  8. D. Parks, "Object Detection and Analysis: A Coherency Filtering Approach", p. 172, 2008.
  9. Yongqiang Z., Chen Y., Seong G. K., Quan P., Yongmei C., "Multi-band Polarization Imaging and Applications" 1st ed., p. 204, 2016.
  10. Manikandan S., "Vision Based Assistive System for Label and Object Detection", p. 64, 2015.
  11. Salma H., "Object Detection Using Histogram Of Gradients", p. 52, 2018.
  12. Wu J., "Advances in K-means Clustering: A Data Mining Thinking", Springer Science & Business Media, p. 180, 2021.
  13. J.Loy, "Neural Network Projects with Python: The ultimate guide to using Python to explore the true power of neural networks through six projects", Packt Publishing, p. 308, 2019.
  14. Brannon W. C., "Object Detection in Low-spatial-resolution Aerial Imagery Using Convolutional Neural Networks", p. 79, 2019.
  15. Dataset <https://knowyourdata-tfds.withgoogle.com>