УДК 681.3.07 DOI https://doi.org/10.32782/1995-0519.2024.1.35

ПРОГРАМНІ ПІДХОДИ ВИКОРИСТАННЯ ВЕЙВЛЕТ-ПЕРЕТВОРЕННЯ В ЗАДАЧАХ РОЗПІЗНАВАННЯ ТЕКСТУ ІЗ ЗОБРАЖЕНЬ

Олексій Переверзєв
аспірант кафедри акустичних та мультимедійних електронних систем
Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», просп. Берестейський, 37, Київ, Україна, 03056, mr.pereverziev@gmail.com
ORCID: 0000-0002-8208-3449

Кирило Трапезон
кандидат технічних наук, доцент, доцент кафедри акустичних та мультимедійних електронних систем
Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», просп. Берестейський, 37, Київ, Україна, 03056, kirill.trapezon@gmail.com
ORCID: 0000-0001-5873-9519

Цифрове зображення, як елемент візуального мультимедійного контенту, дуже часто використовується у програмних додатках аналізу на наявність текстової складової частини в рекламних носіях інформації, у медичній галузі для розпізнавання текстової частини за змістом у базах медичних звітів, в інтелектуальних системах безпеки руху на основі концепції інтернету речей. Водночас необхідно враховувати не лише точність розпізнавання тексту, але й час, який потрібний для аналізу та розпізнавання текстових символів. Визначено теоретичні передумови використання вейвлет-перетворення в завданнях розпізнавання тексту. Наведена структурна схема алгоритму виділення та розпізнавання тексту, на основі якої створено програмний алгоритм розпізнавання текстової інформації з поверхні зображення за допомогою підключених бібліотек pyCharm. Ключовою особливістю розробленого алгоритму є те, що він не передбачає підключення нейронної мережі, отже, у такому разі налаштування та подальший семантичний аналіз зображення не потребує багато часу. Знайдено, що розроблений алгоритм, через використання модуля оптичного розпізнавання OCR, некоректно працює із символами кирилиці, натомість літери латинського алфавіту визначає досить точно, навіть коли присутні зайві графічні елементи зображення, які не стосуються символів тексту та відповідного аналізу. Проведена перевірка розробленого алгоритму на двох зображеннях, які отримано різними способами, показано, що темне тло фотографії на першому етапі аналізу може сприйматись як частина можливого тексту. Водночас отримано підтвердження того, що дана особливість не призводить до спотворення власне текстової частини зображення, але є причиною того, що сама світлина дещо спотворюється за значеннями яскравості пікселів зображення.

Ключові слова: зображення, вейвлет, перетворення, синтез, розпізнавання, текст, алгоритм.

Вступ, актуальність роботи. Для цифрового зображення, як однієї з найважливіших складових частин мультимедійного контенту, характерна величезна кількість різних прикладних застосувань. Одним із таких застосувань є технологія розпізнавання й аналізу тексту на зображеннях. Ця технологія, наприклад, поширена в системах відеоаналітики [1], тобто в комплексах аналізу текстового контенту в рекламних відеофайлах, у медичній галузі – для розпізнавання тексту на зображеннях звітів, в інтелектуальних системах відеонагляду на основі концепції інтернету речей. В основі цієї технології закладено використання різних методів і підходів, серед яких сегментація [2] та методи на основі вейвлет-перетворення. Цим питанням присвячена величезна кількість проведених наукових досліджень. Робота [3] визначає вейвлет-перетворення для розв'язання завдань відновлення зображень. Запропонований алгоритм для реконструкції зображення дозволяє добре видаляти імпульсний шум і зменшувати ефект згладжування країв контурів об'єктів на відеопослідовностях.

У статті [4] розглянуто технологію оборотного нанесення водяних знаків (тексту) на зображення шляхом використання вейвлет-перетворення під час створення текстових документів медичних звітів. Це перетворення використовується для захисту вмісту зображення й уникнення різних способів його підробки. Для цього автори роботи пропонують шифрування знаків різними ключами. Основною метою, як можна визначити зі

статті, є розгляд алгоритму захищеного нанесення водяних знаків для відкриття можливості подальшого виявлення та локалізації можливих підроблених ділянок оригінального зображення.

Цікаве використання вейвлетів описано в роботі [5], де показано новий підхід до покращення якості зображень (надвисока роздільна здатність зображення), а саме, пропонується використовувати глибоку згорткову нейронну мережу, у структурі якої є бета-вейвлет. Так, для аналізу зображення пропонується обчислення різних аналітичних фільтрів бета-вейвлет для оцінки та відновлення відсутніх деталей під час збільшення роздільної здатності зображення. Варто також зазначити, що в роботі використовуються підходи апроксимації та наближень, які не зовсім детально проаналізовано з погляду точності та достовірності.

Ще одне практичне застосування вейвлет-перетворення можна помітити в дослідженні [6], де в межах розглянутої техніки відеоаналізу та дактилоскопії і з метою виявлення відеокопій пропонується використовувати набір вейвлет-фільтрів на основі поліфазного представлення сигналу та залучення алгоритму гравітаційного пошуку. Як приклад у роботі використовується вейвлет Даубекіеса. У підсумку запропоновано метод проєктування вейвлетів для завдань оброблення зображень.

Загальним питанням розпізнавання тексту на поверхні зображення присвячена робота [7]. У цьому дослідженні авторами наведено гібридну модель класифікації відеозображень для подальшого підвищення ефективності процедур виявлення та розпізнавання тексту на поверхні зображення. Для цього використовується модель ResNet50 з метою оцінювання інформації про рівень розподілу пікселів і мережа VGG16 для вивчення особливостей максимально стабільних екстремальних областей зображення. Додатково в наведеному підході використовується каскадна згорткова нейронна мережа. З отриманих результатів роботи на основі аналізу тестових зображень можна помітити прямий зв'язок якості розпізнавання тексту та рівня шуму зображення.

Огляду наявних методів детектування тексту на поверхні зображення присвячена праця [8]. У цій роботі автори підкреслюють важливість виділення семантичної інформації з кадру зображення і водночас наголошують, що виявлення текстової інформації на поверхні зображення з відеопослідовності є досить складною технічною процедурою. Чинниками, які можуть бути причиною невдалого розпізнавання, є низька роздільна здатність, складне тло, нерівномірний текст і зміни його розміру, кольору та стилю. Беручи до уваги об'ємність проведеного огляду, варто відзначити, що в роботі не проведено для визначених підходів на основі вейвлет-функцій узагальненої систематизації обмежень.

Огляд літературних джерел підтвердив актуальність використання вейвлет-перетворень у завданнях, які пов'язані з обробленням зображень. Водночас, на жаль, у розглянутих публікаціях не представлені ефективні алгоритми розпізнавання тексту на поверхні зображення без використання нейронних мереж. Тобто є нагальна необхідність розроблення простого програмного алгоритму виділення тексту із зображення, який би характеризувався малим часом аналізу зображення та водночас високою точністю результату розпізнавання.

Метою роботи є розроблення програмного алгоритму з виділення та подальшого розпізнавання текстових символів на поверхні зображення, який би, з одного боку, ураховував переваги вейвлет-перетворень, а з іншого – не використовував підключення нейронної мережі.

Для досягнення поставленої мети необхідно виконати такі завдання: розробити алгоритм розпізнавання тексту на поверхні зображення; провести перевірку точності цього алгоритму; визначити обмеження та недоліки запропонованого програмного алгоритму.

Матеріал, результати досліджень. Якщо кольорове зображення представити у вигляді сигналу функції, то ідея застосування вейвлет-перетворень полягає в розкладанні цього сигналу за базисом, який утворюється зсувами та копіями функції-прототипу в різних масштабах. Ці базисні функції і називаються вейвлетами, якщо вони можуть бути визначені на просторі $L^2(R)$ ($L^2(R)$ – простір комплексно значущих функцій $f(x)$ на прямій з обмеженою енергією), та мають ознаки коливання навколо осі абсцис і швидко сходяться до нуля за збільшення абсолютного значення аргументу [9]. Так процедура згортки сигналу з вейвлет-функцією дозволяє визначити характерні особливості цього сигналу в околі локалізації цього вейвлета. Базисну функцію перетворення можна визначити так [9; 10]:

$$F_{a,b} = \frac{1}{\sqrt{a}} F\left(\frac{x-b}{a}\right), \quad a \in R^+; \; b \in R;$$

тоді вейвлет-перетворення матиме вигляд:

$$W(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} F\left(\frac{x-b}{a}\right) \cdot f(x)\,dx.$$

Параметр b визначає розташування вейвлета в часі, a – параметр масштабу. Щоб можна було отримати зворотне вейвлет-перетворення, базисна функція $F(x)$ має задовольняти енергетичній умові:

$$\int_0^\infty \frac{|\Phi(\omega)|^2}{\omega} d\omega < \infty,$$

де $\Phi(\omega)$ – перетворення Фур'є від функції $F(x)$.

Для побудови алгоритму розпізнавання тексту на поверхні зображення використаємо один із найпростіших вейвлетів – вейвлет Хаара (рис. 1) [11; 12].

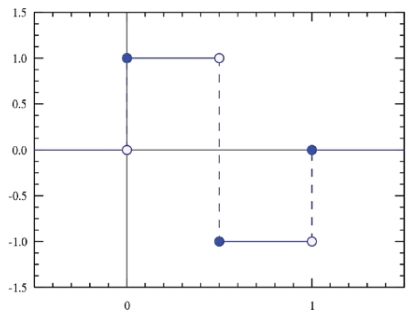

Рис. 1. Графічне зображення функції вейвлета Хаара [12]

Для побудови програмного алгоритму розпізнавання тексту на зображенні з використанням вейвлет-перетворення використаємо середовище pyCharm і відповідні бібліотеки. Підключимо бібліотеку pytesseract. Окрім цього, застосуємо для завантаженого зображення (рис. 2) порогову обрізку для виділення тексту з фону. Блок-схема алгоритму виділення та розпізнавання тексту показана на рисунку 3. На основі створеної схеми з рисунку 3 сформуємо програмний код у середовищі pyCharm. Варто зазначити, що застосування функції порогової обрізки для тексту на зображення буде проведено відповідно до вимог, які розміщено в роботі [13].

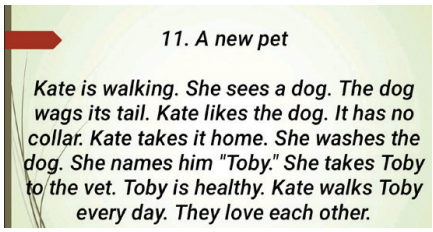

Рис. 2. Графічне зображення тексту

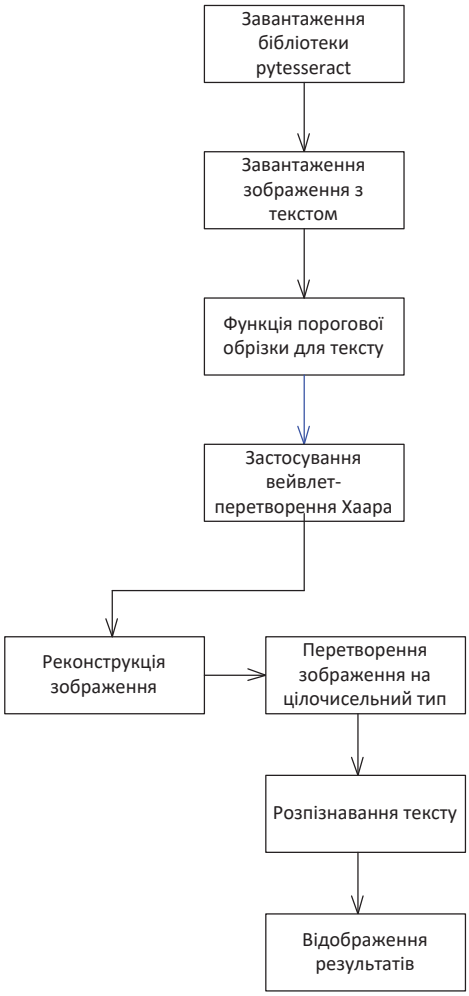

Рис. 3. Структурна схема алгоритму виділення та розпізнавання тексту

Зазначимо, що після завантаження бібліотеки pytesseract необхідно налаштувати на комп'ютері Tesseract OCR – модуль оптичного розпізнавання. Далі наведено програмний код алгоритму:

```
import cv2
import numpy as np
import matplotlib.pyplot as plt
import pywt
import pytesseract

# Встановлення шляху до Tesseract OCR
pytesseract.pytesseract.tesseract_cmd = r'C:\Program Files\Tesseract-OCR\tesseract.exe'

# Завантаження зображення
image = cv2.imread('3.jpg', 0) # Зчитування у відтінках сірого

# Застосування порогової обрізки для виділення тексту
```

```
_, thresholded_image = cv2.threshold(image, 0, 255, cv2.THRESH_BINARY + cv2.THRESH_OTSU)

# Вейвлет-перетворення зображення (використовуємо вейвлет Хаара)
coeffs = pywt.dwt2(thresholded_image, 'haar')

# Реконструкція зображення з коефіцієнтів
reconstructed_image = pywt.idwt2(coeffs, 'haar')

# Перетворення зображення на цілочисельний тип
reconstructed_image = np.uint8(reconstructed_image)

# Розпізнавання тексту за допомогою pytesseract
recognized_text = pytesseract.image_to_string(reconstructed_image)

# Відображення результатів і розпізнаного тексту
plt.figure(figsize=(12, 6))

plt.subplot(231)
plt.title('Original Image')
plt.imshow(image, cmap='gray')

plt.subplot(232)
plt.title('Thresholded Image')
plt.imshow(thresholded_image, cmap='gray')

plt.subplot(233)
plt.title('Reconstructed Image')
plt.imshow(reconstructed_image, cmap='gray')

plt.tight_layout()
plt.show()
# Виведення розпізнаного тексту

print("Recognized Text:")
print(recognized_text)
```

Результати досліджень. Отже, на основі сформованого коду програми завантажимо зображення з текстом, яке показано на рисунку 2. Наведений код реалізує порогову обрізку для виділення тексту на зображенні, використовує метод вейвлет-перетворення Хаара для отримання частотних коефіцієнтів, формує реконструкцію зображення з текстом з подальшим застосуванням оптичного розпізнавання символів за допомогою бібліотеки pytesseract. Зазначимо, що порогова обрізка зображення на основі метода Оцца потрібна для визначення лише областей із текстом і формування так званих «зайвих» фонових пікселів. Ці пікселі далі просто ігноруються під час аналізу поверхні зображення, тим самим можна досягти вищої точності розпізнавання символів тексту. На рисунку 4 наведено порівняльні результати роботи алгоритму.

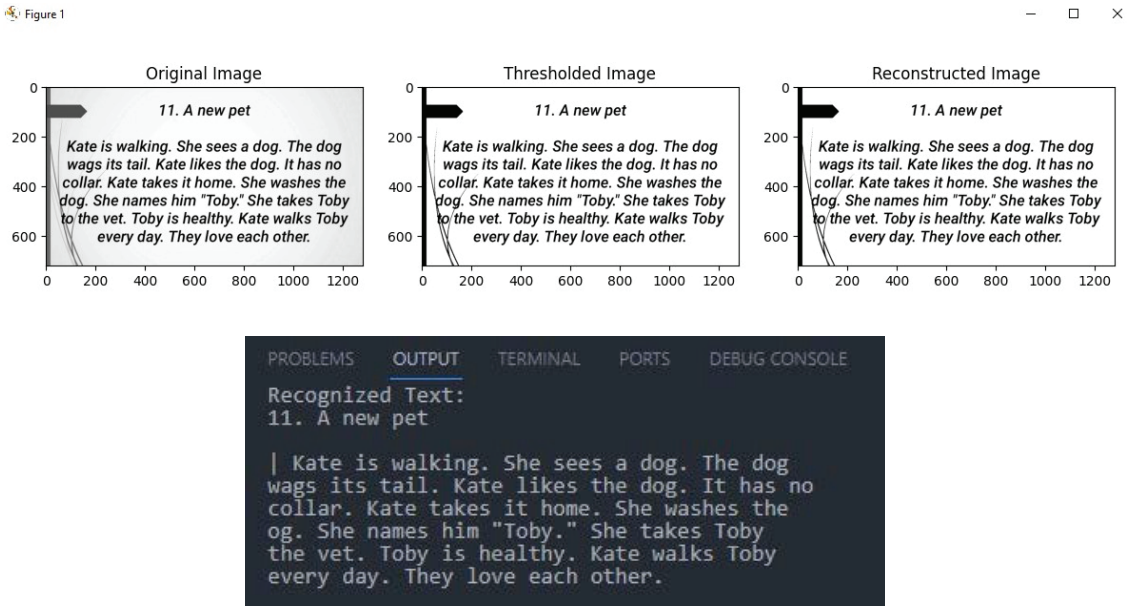

Рис. 4. Результати алгоритму

На основі отриманих результатів варто зазначити такі особливості роботи алгоритму. Незважаючи на те, що оригінальне зображення з текстом є кольоровим, сам алгоритм працює із зображенням для аналізу, як із чорно-білим. Це зумовлено тим, що процедура порогової обрізки зображення переводить саме зображення в чорно-біле. Окрім цього, з рисунка 4 можна переконатись, що алгоритм розпізнавання тексту стійкий до фонових елементів. Тобто обрамлення зображення має рисунок у формі кривих ліній, які перетинають в оригінальному зображенні слова “dog”, “collar” та “to”. Як можна переконатись на отриманому результаті, алгоритм не сприймає ці додаткові елементи і не спотворює відтворений текст. Для додаткової перевірки візьмемо скрин із текстом із сайту і перевіримо, як це зображення буде проаналізовано. Результати роботи алгоритму показано на рисунках 5 та 6. Тут можна відмітити цікаву особливість, що фото, яке розташовано на скрині, через алгоритм зазнає деяких спотворень і зміни яскравості. Це, напевно, зумовлено тим, що запропонований алгоритм під час виділення фонових пікселів сприймає темні частини світлини на першому етапі як можливі частини тексту. І ця особливість підкреслює, що розпізнавання символів проходить циклічні етапи перевірки з відкиданням зайвих елементів зображення.

На рисунку 6 наведено результат роботи алгоритму в терміналі середовища pyCharm.

На рисунку 6 можна побачити, що алгоритм розпізнавання некоректно працює з деякими символами, насамперед це стосується символів на кшталт «лапки». Окрім цього, створений алгоритм має і свої обмеження. Насамперед це стосується того, що він некоректно проводить розпізнавання літер нелатинського алфавіту. Вочевидь, це зумовлено особливостями роботи модуля оптичного розпізнавання OCR.

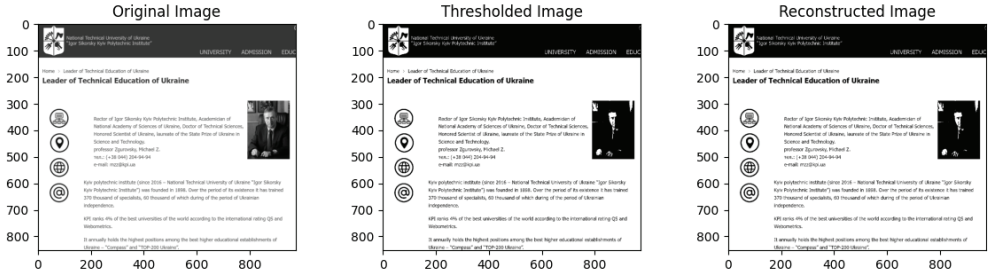

Рис. 5. Результати алгоритму

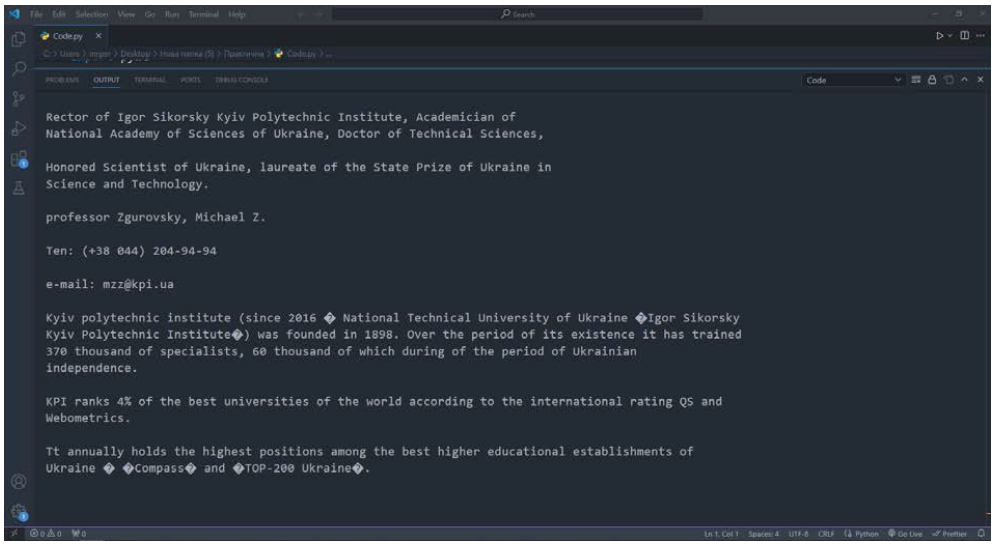

Рис. 6. Відтворений текст у терміналі програмного середовища pyCharm

Висновки. У роботі розглянуто питання розпізнавання тексту з поверхні зображення на основі використання найпростішого вейвлет-перетворення Хаара у програмному середовищі pyCharm. Визначено теоретичні передумови використання вейвлет-перетворення в завданнях розпізнавання тексту. У створенні програмного алгоритму та відповідної структурної схеми враховано, що процедура згортки сигналу з вейвлет-функцією дозволяє визначити характерні особливості цього сигналу в околі локалізації цього вейвлета. Знайдено, що розроблений алгоритм некоректно працює із символами кирилиці, натомість літери латинського алфавіту визначає досить точно, навіть за наявності зайвих елементів зображення, які не стосуються символів тексту та відповідного аналізу. Проведена перевірка розробленого алгоритму на двох зображеннях, які отримано різними способами, і показано, що темне тло фотографії на першому етапі аналізу може сприйматись як частина можливого тексту. Водночас отримано підтвердження того, що дана особливість не призводить до спотворення текстової частини, але є причиною того, що сама фотографія дещо спотворюється за значеннями яскравості пікселів зображення.

ЛІТЕРАТУРА

1. On the use of semantic technologies for video analytics / L. Greco et al. *J Ambient Intell Human Comput.* 2021. № 12. P. 567–587. https://doi.org/ 10.1007/ s12652-020-02021-y.
2. A systematic review of deep learning frameworks for moving object segmentation / D. Gupta et al. *Multimed Tools Appl*. 2023. https://doi.org/ 10.1007/s11042-023-16417-3.
3. Jiang L., Yin H. Total generalized variation and wavelet transform for impulsive image restoration. *SIViP.* 2022. № 16. P. 773–781. https://doi.org/ 10.1007/ s11760-021-02017-x.
4. A reversible watermarking for image content authentication based on wavelet transform / D. Li et al. *SIViP.* 2024. https://doi.org/ 10.1007/s11760-023-02950-z.
5. Image super resolution boosting using beta wavelet / N. Chaibi et al. *SIViP.* 2023. https://doi.org/ 10.1007/ s11760-023-02887-3.
6. Preetha S., Bindu V. A Wavelet Optimized Video Copy Detection Using Content Fingerprinting. *J Sign Process Syst.* 2023. № 95. P. 363–377. https://doi.org/ 10.1007/ s11265-022-01830-y.
7. A deep action-oriented video image classification system for text detection and recognition / A. Chaudhuri et al. *SN Appl. Sci.* 2021. № 3. P. 838. https://doi.org/ 10.1007/ s42452-021-04821-z.
8. Decade research on text detection in images/ videos : a review / V. Manjunath Aradhya et al. *Evol. Intel.* 2021. № 14. P. 405–431. https://doi.org/ 10.1007/ s12065-019-00248-z.
9. Mallat S. A theory for multiresolutional signal decomposition: the wavelet representation. *IEEE Trans. Pattern Analysis and Machine Intelligence.* 1989. № 7. P. 674–693.
10. Daubechies I. The wavelet transform, time-frequency localization and signal analysis. *IEEE Trans. Inf. Theory.* 1990. № 36. 5. P. 961–1005. https://doi.org/10.1109/18.57199.
11. Lal S., Kumari P. Wavelet Approximation of Functions Belonging to Generalized Lipschitz Class by Extended Haar Wavelet Series and its Application in the Solution of Bessel Differential Equation Using Haar Operational Matrix. *Int. J. Appl. Comput. Mat.* 2019. № 5. P. 67. https://doi.org/10.1007/s40819-019-0646-7.
12. Ramakrishnan V., Pete D. Haar Wavelet-Based Fusion of Multiple Exposure Images for High Dynamic Range Imaging. *SN COMPUT. SCI.* 2022. № 3. P. 129. https://doi.org/10.1007/s42979-021-01010-y.
13. Remote sensing image denoising based on improved semi-soft threshold / S. Lei et al. *SIViP.* 2021. № 15. P. 73–81. https://doi.org/ 10.1007/s11760-020-01722-3.

SOFTWARE APPROACHES TO THE USE OF WAVELET TRANSFORM IN TEXT RECOGNITION TASKS FROM IMAGES

Oleksii Pereverziev

Postgraduate Student at the Department of Acoustic and Multimedia Electronic Systems

National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", 37 Beresteiskyi ave., Kyiv, Ukraine, 03056, mr.pereverziev@gmail.com

ORCID: 0000-0002-8208-3449

Kirill Trapezon

Candidate of Technical Sciences, Associate Professor, Associate Professor of the Department of Acoustic and Multimedia Electronic Systems

National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Beresteiskyi ave., 37, Kyiv, Ukraine, 03056, kirill.trapezon@gmail.com

ORCID: 0000-0001-5873-9519

The aim of the work is to develop a software algorithm for the extraction and subsequent recognition of text characters on the image surface, which, on the one hand, would take into account the advantages of wavelet transforms and, on the other hand, would not use the connection of a neural network. The article presents analytical dependencies for determining the wavelet transform that can be used for image analysis. The practical part includes program code and an algorithm for recognizing text characters based on the libraries of the pyCharm software environment. The developed structural scheme of the algorithm for detecting and recognizing text characters allows cataloging image databases that are identified only on the basis of the textual component of information. The limitations of the developed algorithm are determined and the correctness of its operation is checked on two images obtained in different ways. The developed software algorithm for detecting and recognizing textual information allows you to quickly and accurately extract text from an image without using neural networks, only based on the unique features of wavelet functions.

The presented results of the study, due to the obtained confirmation of the quality of the algorithm, can be used in the future to analyze image databases in various applications of technical vision systems, in particular, in solving problems of analyzing images with text labels in electronic video surveillance systems. The algorithm of actions for constructing a software algorithm for detecting and recognizing text characters using the libraries of the pyCharm environment and the OCR module is presented. The limitations of the developed algorithm are determined and practical evidence of the effective use of wavelet functions in the tasks of recognizing textual information as part of a color image is obtained.

Key words: image, wavelet, transformation, synthesis, recognition, text, algorithm.

REFERENCES

1. Greco, L., Ritrovato, P., & Vento, M. (2021). On the use of semantic technologies for video analytics. *J Ambient Intell Human Comput, 12*, 567–587. https://doi.org/10.1007/s12652-020-02021-y.

2. Gupta, D., Kumar, M., & Chaudhary, S. (2023). A systematic review of deep learning frameworks for moving object segmentation. *Multimed Tools Appl*. https://doi.org/10.1007/s11042-023-16417-3.

3. Jiang, L., & Yin, H. (2022). Total generalized variation and wavelet transform for impulsive image restoration. *SIViP, 16*, 773–781. https://doi.org/10.1007/s11760-021-02017-x.

4. Li, D., Dai, X., Gui, J., et al. (2024). A reversible watermarking for image content authentication based on wavelet transform. *SIViP*. https://doi.org/10.1007/s11760-023-02950-z.

5. Chaibi, N., Ben Aoun, N., Eladel, A., et al. (2023). Image super resolution boosting using beta wavelet. *SIViP*. https://doi.org/10.1007/s11760-023-02887-3.

6. Preetha, S., & Bindu, V. (2023). A Wavelet Optimized Video Copy Detection Using Content Fingerprinting. *J Sign Process Syst, 95*, 363–377. https://doi.org/10.1007/s11265-022-01830-y.

7. Chaudhuri, A., Shivakumara, P., Chowdhury, P., et al. (2021). A deep action-oriented video image classification system for text detection and recognition. *SN Appl. Sci., 3*, 838. https://doi.org/10.1007/s42452-021-04821-z.

8. Manjunath Aradhya, V., Basavaraju, H., & Guru, D. (2021). Decade research on text detection in images/videos: a review. *Evol. Intel., 14*, 405–431. https://doi.org/10.1007/s12065-019-00248-z.

9. Mallat, S. (1989). A theory for multiresolutional signal decomposition: the wavelet representation. *IEEE Trans. Pattern Analysis and Machine Intelligence, 7*, 674–693.

10. Daubechies, I. (1990). The wavelet transform, time-frequency localization and signal analysis. *IEEE Trans. Inf. Theory, 36* (5), 961–1005. https://doi.org/10.1109/18.57199.

11. Lal, S., & Kumari, P. (2019). Wavelet Approximation of Functions Belonging to Generalized Lipschitz Class by Extended Haar Wavelet Series and its Application in the Solution of Bessel Differential Equation Using Haar Operational Matrix. *Int. J. Appl. Comput. Mat, 5*, 67. https://doi.org/10.1007/s40819-019-0646-7.

12. Ramakrishnan, V., & Pete, D. (2022). Haar Wavelet-Based Fusion of Multiple Exposure Images for High Dynamic Range Imaging. *SN COMPUT. SCI, 3*, 129. https://doi.org/10.1007/s42979-021-01010-y.

13. Lei, S., Lu, M., Lin, J., et al. (2021). Remote sensing image denoising based on improved semi-soft threshold. *SIViP, 15*, 73–81. https://doi.org/10.1007/s11760-020-01722-3.

Стаття надійшла 13.02.2024