



МЕТОД КОНВЕРСІЇ ВІЗУАЛЬНИХ МЕДІА-РЕСУРСІВ НА ОСНОВІ ПРОСТОРОВОЇ АГРЕГАЦІЇ ТА БАГАТОМАСШТАБНОЇ ВЕРИФІКАЦІЇ

Левикін І.В., д.т.н., професор, кафедра МСТ, ХНУРЕ

Шимко Д.І., аспірант, кафедра МСТ, ХНУРЕ

Abstract. *This paper proposes a method for converting visual media resources to minimize computational redundancy and API latency during OCR processing. The approach integrates spatial data aggregation with multi-scale verification across a four-stage pipeline. Implementation results demonstrate a significant reduction in network transactions and enhanced semantic extraction accuracy without requiring additional server-side hardware.*

Keywords: *visual media conversion, spatial aggregation, multi-scale verification, inference optimization, optical character recognition.*

Автоматизована конверсія динамічного візуального контенту у структуровані масиви даних є важливою задачею для сучасних систем збору візуальної телеметрії [1]. Масштабування таких систем ускладнюється надмірним обчислювальним навантаженням під час взаємодії з хмарними сервісами розпізнавання. Оптимізація вартості обробки даних є загальносистемним викликом у сфері розробки систем на базі моделей генеративного штучного інтелекту, де одним із методів такої оптимізації є мінімізація кількості запитів до хмарних API [2]. Безперервна передача кадрів до зовнішніх API оптичного розпізнавання символів (OCR) спричиняє лінійне зростання мережових затримок та витрат. Додатковим обмежувальним фактором є природа алгоритмів комп'ютерного зору, що призводить до хибних спрацьовувань та пропусків даних. Вирішення цієї проблеми вимагає мінімізації кількості мережових запитів без зниження точності розпізнавання.

Метою дослідження є розробка методу конверсії візуальних медіа-ресурсів для зниження обчислювальної надмірності шляхом просторової агрегації цільових візуальних фрагментів та підвищення достовірності результатів за допомогою багатомасштабної верифікації.

Запропонований метод обробки складається з чотирьох етапів. На першому етапі здійснюється первинна валідація вхідного відеопотоку. Застосування механізмів адаптивної фільтрації та відкладеного аналізу дозволяє ізолювати та відхилити нецільові кадри до початку ресурсоємної обробки [3]. Алгоритм перевіряє зображення на відповідність заданим колірним діапазнам та геометричним параметрам об'єктів. Кадри, що не проходять цю перевірку, відхиляються на рівні локального клієнта.

На другому етапі застосовується просторова агрегація розрізнених цільових областей. Для усунення надлишкових транзакцій відфільтровані фрагменти компонуються в єдине зображення, що забезпечує передачу необхідного контексту одним мережовим запитом. Формування композитних зображень (packed canvas) мінімізує затримки та операційні витрати систем відеоаналітики під час розпізнавання [4].



Третій етап реалізує багатомасштабну генерацію даних. Складене зображення дублюється з різними коефіцієнтами масштабування для створення піксельних копій з альтернативною роздільною здатністю. Застосування алгоритмів багатомасштабного аналізу зменшує кількість помилок OCR та підвищує точність розпізнавання тексту [5].

Четвертий етап забезпечує просторове та семантичне зіставлення результатів. Модуль верифікації виконує перехресне зіставлення масивів даних, отриманих з різних масштабованих копій. Він обчислює площу взаємного перекриття розпізнаних областей відносно їхньої загальної площі та перевіряє збіг координат відповідних обмежувальних рамок. Це забезпечує верифікацію істинних спрацьовувань та автоматичне відфільтровування хибних розпізнавань із нетиповими просторовими характеристиками.

Апробацію розробленого методу здійснено під час створення клієнтського застосунку The Finals Stats (<https://finals-stats.com>). Впровадження просторової агрегації забезпечило кратне зменшення кількості звернень до зовнішніх API, оптимізувавши витрати на обробку. Водночас багатомасштабна верифікація виконує функцію фільтра хибних спрацьовувань машинного розпізнавання. Перенесення основної обробки на сервер та виконання локально лише геометричної перевірки запобігає перевантаженню пристрою клієнта. Запропонована архітектура є масштабованою і може бути адаптована для суміжних завдань моніторингу потокового відео в реальному часі.

Список літератури

1. Левикін, І.В., & Шимко, Д.І. (2025). Методи автоматизованої конверсії відеоконтенту в структурований медіаматеріал. Поліграфічні, мультимедійні та web-технології. Інновації та розвиток: монографія. (с. 161-185). Харків: ТОВ «Друкарня Мадрид».
2. Prokhorov, O., Shymko, D., Kuzminska, O., Chukhray, A., Shatalov, O., & Kholodniak, O. (2025). A system for generating chatbots to support learning in the field of exact sciences using generative artificial intelligence models. *Radioelectronic and Computer Systems*, (2), 22-44. <https://doi.org/10.32620/reks.2025.2.02>.
3. Шимко, Д.І., & Левикін, І.В. (2026). Методи мінімізації обчислювальної надмірності при автоматизації конверсії потокових медіа-ресурсів. *Радіоелектроніка та молодь у XXI столітті. Т. 2.* (с. 259-261).
4. Jeon, C., Kim, S., Yi, J., & Lee, Y. (2024). Mondrian: On-device high-performance video analytics with compressive packed inference. *arXiv*. <https://arxiv.org/abs/2403.07598>.
5. Geng, T. (2024). Transforming scene text detection and recognition: A multi-scale end-to-end approach with transformer framework. *IEEE Access*. 12, 40582-40596. <https://doi.org/10.1109/ACCESS.2024.3375497>.