DOI: 10.18372/2310-5461.66.20246
УДК 004.056.53

Ю. Ю. Сокиран
Київський національний університет
імені Тараса Шевченка
orcid.org/0009-0004-7041-2307
email: sokyran@knu.ua

І. І. Пархоменко, канд. техн. наук, доцент
Київський національний університет
імені Тараса Шевченка
orcid.org/0000-0001-6889-9284
e-mail: ivan.parkhomenko@knu.ua

ЗАСТОСУВАННЯ СЕМАНТИЧНОГО ТА ОБ'ЄКТНОГО ПОШУКУ ДЛЯ ОПТИМІЗАЦІЇ ПОШУКУ ПО МЕДІА ЗАДЛЯ *OSINT* РОЗСЛІДУВАНЬ

Вступ

У сучасному інформаційному просторі стає все більше даних. Це не тільки текстові дані, але й аудіо, фото та різного роду відео. Більше того, динаміку росту кількості та розміру цих даних тільки збільшується [1], адже після того як великі мовні моделі (англ. *large language model, LLM*) та сервіси, що пропонують додаткові послуги із роботи з ними, набули популярності серед усього людства, створювати або генерувати ці дані стало дуже легко у великій кількості. При цьому якість цих даних не обов'язково покращилась. Окрім цього, все більшої популярності набувають соціальні мережі, користувачі яких публікують все більше медіа. Враховуючи ці обставини, у сфері розвідки на основі відкритих джерел (англ. *Open source intelligence, OSINT*) важливо мати змогу шукати інформацію достатньо швидко та ефективно відфільтровувати дійсно релевантні та інформативні варіанти.

У сучасній обробці відео значне місце займає розробка ефективних методів кількісного вимірювання семантичної різниці між кадрами. Одним із перспективних підходів є використання векторних представлень (ембедінгів), отриманих за допомогою таких моделей, як *CLIP*, та обчислення косинусної схожості між ними для визначення міжкадрової відстані. Зважаючи на стрімке зростання обсягів відеоданих, розуміння існуючих методів порівняння відеокадрів та виявлення ключових моментів є надзвичайно важливим подальшого розвитку подібних метрик.

З урахуванням зазначеного уявляється доцільним проаналізувати цю ситуацію та визначити напрями покращення ефективності пошуку інформації по відкритих джерелах.

Аналіз останніх досліджень і публікацій

У галузі порівняння відеокадрів виділяється кілька ключових напрямків досліджень. За даними IEEE, сучасні методи семантичного індексування відео спираються на великі набори даних та інноваційні підходи до виявлення та індексування семантичних концептів [2]. Це дозволяє значно підвищити точність пошуку в порівнянні з традиційними методами.

Модель *CLIP* (*Contrastive Language-Image Pre-training*), представлена *OpenAI* у 2021 році, здійснила революцію в галузі мультимодального аналізу [3]. Дослідження *Szymon Palucha* підкреслює, що *CLIP* став основним компонентом багатьох мультимодальних систем штучного інтелекту [4]. Модель спирається на контрастивне навчання для створення єдиного векторного простору для текстів та зображень, що дозволяє ефективно виконувати крос-модальні задачі пошуку.

У сфері відео-ембедингів останні дослідження зосереджені на інтеграції зовнішніх знань, обробці неповних даних, захопленні просторово-часової динаміки [5, 3]. Ці підходи забезпечують покращену продуктивність у таких задачах, як рекомендація відео, класифікація та пошук відео на основі контенту.

Перцептивне хешування залишається важливим методом для ідентифікації візуально подібних кадрів. Дослідження 2021 року в *EURASIP Journal on Image and Video Processing* висвітлює методи перцептивного хешування відео з максимізованою стійкістю для автентифікації відеоконтенту [6]. Ці методи забезпечують збалансування між чутливістю до маніпуляцій із вмістом та стійкістю до операцій, що зберігають перцептивний зміст.

У галузі виявлення об'єктів спостерігається значний прогрес. За даними ScienceDirect, сучасні методи глибинного навчання для виявлення об'єктів не лише розпізнають категорії об'єктів, але й точно прогнозують їх розташування за допомогою обмежувальних рамок [7]. Ці методи дозволяють аналізувати значно більше семантичної інформації порівняно з традиційними підходами на основі пікселів або ознак.

Особливо важливим для *OSINT*-розслідувань є ефективна обробка великих обсягів даних. Дослідник Brandon Castellano розробив інструмент, який прискорює та автоматизує перегляд відео, виділяючи кадри, де починається нова сцена, і дозволяючи швидко аналізувати ключові частини відео [8]. Такі інструменти стають незамінними для *OSINT*-аналітиків, які працюють із великими обсягами відеоматеріалів.

Інтеграція семантичних ембедингів із методами виявлення об'єктів представляє перспективний напрямок досліджень, що дозволяє поєднати переваги обох підходів. Цей комбінований підхід особливо цінний для *OSINT*-розслідувань, де важливо не лише виявити об'єкти, але й зрозуміти їх контекст та взаємозв'язки.

Постановка завдання

Незважаючи на стрімкий розвиток технологій обробки та аналізу великих даних, зростання обсягів різнорідної інформації (текстової, аудіо, фото та відео) створює значні виклики в сфері розвідки на основі відкритих джерел (*OSINT*). Відтак, існує нагальна потреба у розробці комплексних методів та інструментів, які дозволять підвищити ефективність пошуку, фільтрації та захисту інформації у відкритих джерелах в умовах інформаційного перевантаження та зростаючих кіберзагроз.

Мета статті

Метою статті є розробка та обґрунтування комплексного підходу до оптимізації пошуку в медіа-контенті для *OSINT* розслідувань шляхом інтеграції методів семантичного та об'єктного аналізу, що забезпечує ефективне виявлення ключових відеокадрів і підвищення точності фільтрації релевантної інформації даних.

Основна частина

Порівняння відеокадрів – це один із методів фільтрації інформації відео-контенту. Відео являє собою ряд картинок, тому можна аналізувати кожну і залишати тільки необхідні, а саме відфільтровувати ті, які не мають нової інформації, повторюють вже відоме. Такий підхід дозволяє суттєво зменшити обсяг даних для обробки та зберігання. Крім того, ефективність порівняння відеокадрів залежить від вибраного алгоритму та порогових значень схожості зображень.

Методи порівняння на рівні пікселів – ці методи порівняння аналізують безпосередньо значення кожного пікселя в кадрах. Ці методи швидкі, але чутливі до шумів, змін освітлення та незначних зміщень камери.

• Покадрове розрізнення: одним із базових методів порівняння відеокадрів є покадрове розрізнення, яке полягає у простому відніманні значення пікселів попереднього кадру від поточного [9]. Цей підхід ґрунтується на припущенні, що камера залишається нерухомою; у випадку руху камери потрібні складніші методи, такі як компенсація руху камери [9]. Покадрове розрізнення ефективно виявляє рух, оскільки зміни в інтенсивності пікселів між послідовними кадрами часто свідчать про переміщення об'єктів. Для підвищення точності виявлення руху, особливо в умовах шумів та змін освітлення, застосовують перетворення кадрів у відтінки сірого, медіанне розмиття для усунення шумів та заповнення прогалин у масці руху. Крім того, для виявлення небажаних дефектів, спотворення чи небажаних елементів у відео (все це ще називають артефактами) можна використовувати покадрове розрізнення шляхом віднімання відповідних кадрів з двох різних відеопотоків після їхнього можливого масштабування. Попри свою простоту, покадрове розрізнення має обмеження, зокрема чутливість до тіней та шумів, що може призвести до хибних спрацювань [9].

• Побітове порівняння: Якщо два відеофайли мають абсолютно ідентичний формат та тип, найпростішим та найнадійнішим способом їхнього порівняння є побітове порівняння, яке перевіряє кожен байт файлу на ідентичність [10]. У криміналістиці для визначення ідентичності окремих зображень у відеозаписі використовується побітове порівняння за допомогою спеціалізованого програмного забезпечення, такого як *X-Ways Forensics* [11]. Побітове порівняння є надзвичайно чутливим до будь-яких змін у файлі, включаючи метадані або відмінності у стисненні, що робить його корисним для перевірки цілісності файлу або виявлення точних копій, але непридатним для порівняння візуально схожого контенту, що може відрізнятися форматом або стисненням [10].

Методи порівняння на основі ознак – це, наприклад, колірні гістограми. Мірою схожості між відеокадрами може слугувати порівняння колірних гістограм, зокрема блокових колірних гістограм [12]. Схожість між двома кадрами визначається на основі абсолютної різниці відповідних

значень гістограм для кожного блоку, нормалізованої на загальну кількість пікселів та рівнів гістограми. Колірні гістограми відображають розподіл кольорів у зображенні, і їхнє порівняння може вказувати на загальну колірну схожість між кадрами, навіть якщо просторове розташування пікселів відрізняється. Блокові гістограми дозволяють фіксувати локальні варіації кольорів та є більш стійкими до незначних просторових зміщень або рухів об'єктів [12]. Порівняння колірних гістограм є ефективним для кадрів зі схожою колірною композицією, але не враховує текстуру, форму або семантичний зміст [12].

Процес пошуку зображень на основі колірних гістограм складається з трьох основних етапів:

1. Етап наповнення бази даних:

- Завантажуються всі зображення з бази даних;
- Виділяються окремі колірні канали (*R, G, B*);
- Обчислюються гістограми для кожного колірного каналу. Для кожного пікселя визначається інтенсивність кольору (значення від 0 до 255). Підраховується кількість пікселів з кожним значенням інтенсивності. Результат представляється у вигляді графіка або масиву значень;
- Гістограми об'єднуються в єдиний вектор ознак;
- Вектори зберігаються для подальшого порівняння;

2. Етап запиту:

- Завантажується зображення для пошуку.
- Воно обробляється тим самим способом, що й зображення в базі даних.
- Генерується вектор колірної гістограми для зображення запиту.

3. Етап пошуку:

- Вектор запиту порівнюється з усіма збереженими векторами.
- Обчислюються показники схожості.
- Результати ранжуються за показниками схожості.
- Повертаються K найбільш схожих зображень.

Показники схожості обчислюються за допомогою косинусної подібності:

$$\cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2 \sum_{i=1}^{n} B_i^2}}. \qquad (1)$$

Ця технологія дозволяє знаходити візуально схожі зображення на основі їхнього колірного профілю, без необхідності ручної анотації метаданих.

Перцептивне хешування є ефективним способом ідентифікації візуально схожих відеокадрів, навіть якщо вони не є точними копіями на рівні пікселів [13]. Цей метод передбачає генерування компактного хеш-значення (відбитка) на основі візуального вмісту кадру. Щоб уникнути проблем з логотипами та рекламними оголошеннями, які часто розташовуються у верхній та нижній частинах екрана, спочатку ізолюється середня третина кожного відеокадру. Потім до цієї області застосовується алгоритм перцептивного хешування. Схожість між двома кадрами визначається шляхом порівняння їхніх хеш-значень, часто за допомогою відстані Геммінга. Якщо різниця між хешами не перевищує певного порогу (наприклад, 25 %), кадри вважаються схожими [13]. Перцептивне хешування застосовується в різних сферах, включаючи онлайн-безпеку (виявлення шкідливого контенту), захист авторських прав (пошук схожих відео з водяними знаками або модифікаціями) та цифрову криміналістику (ідентифікація схожих зображень у базах даних). Хоча перцептивне хешування є ефективним для виявлення візуально схожого контенту, його чутливість до семантичних відмінностей може бути обмеженою [13].

Цей підхід можна використовувати, щоб знайти фотографію людини у відео, але з урахуванням деяких важливих міркувань. Це добре працює якщо шукаємо дуже схоже зображення або кадр людини (та сама поза, подібне освітлення), якщо людина з'являється в середній частині кадру без всяких водяних знаків або модифікацій та якщо зовнішній вигляд людини не змінився кардинально (той самий одяг, зачіска тощо).

Також є обмеження для виявлення особи:

• Варіанти пози: люди змінюють положення та пози на відео.

• Відмінності в масштабі: людина може виглядати більшою або меншою в різних кадрах.

• Варіанти обрамлення: особа може не постійно з'являтися в середній третині.

• Зміни освітлення: різне освітлення може вплинути на хеш сприйняття.

Процес пошуку представлено на рис. 1.

Для визначення схожості між відеокадрами на більш високому семантичному рівні використовується виявлення об'єктів, наприклад, за допомогою моделі *YOLO* (*You Only Look Once*) [14]. Основна ідея полягає в тому, що відео, присвячені схожим темам, ймовірно, містять схожі кадри.

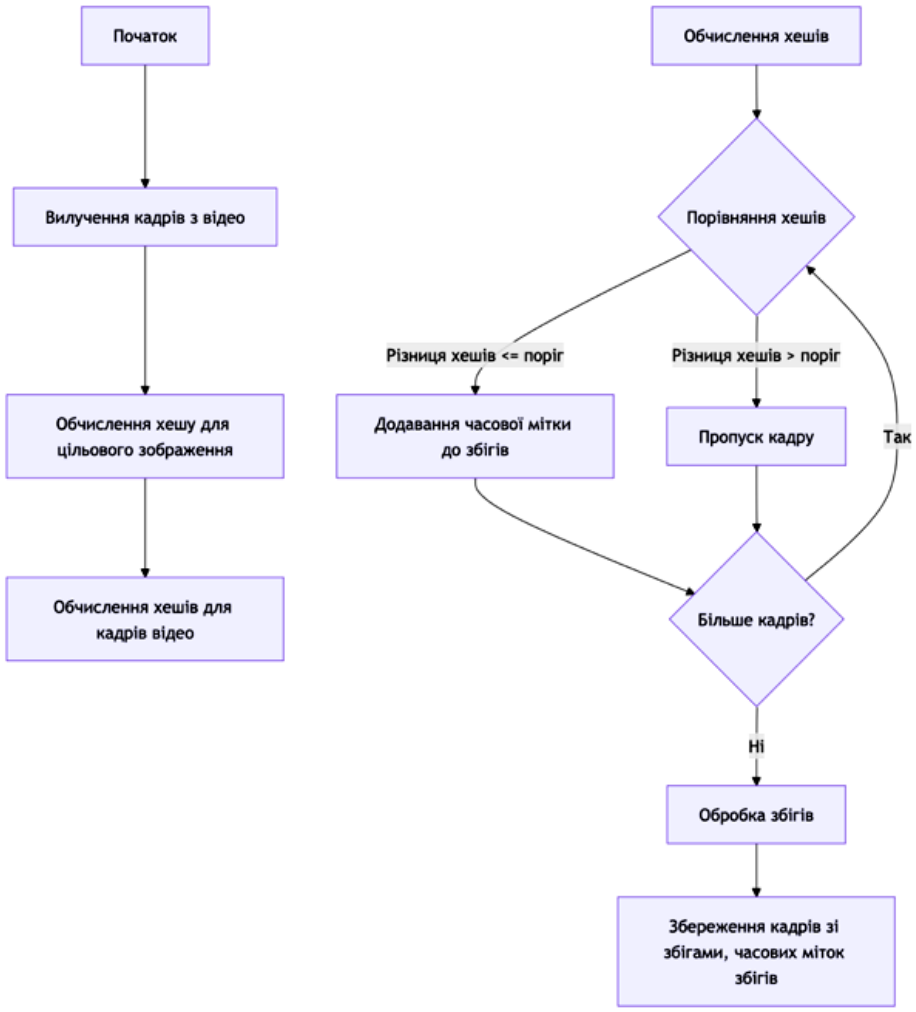

Рис. 1. Процес пошуку зображення серед відео кадрів за перцептивним хешем

Підхід із використанням моделі *YOLO* фокусується на виявленні ключових кадрів на основі наявності об'єктів у них. *YOLO* є швидким та відносно точним детектором об'єктів, який ідентифікує об'єкти, їхні розташування та розміри на кожному кадрі. Ключові кадри можуть бути вилучені шляхом сегментації відео на основі змін у положеннях або кількості об'єктів між кадрами. Пошук по відео потім визначається шляхом порівняння наборів ключових кадрів на основі виявлених об'єктів. Наприклад, для сегментації можна використовувати метод на основі *IoU* (*Intersection Over Union*), який порівнює перекриття обмежувальних рамок об'єктів між послідовними кадрами, або метод на основі частоти, який групує кадри за кількістю певних об'єктів (наприклад, людей) [14]. Для аналізу схожості кадрів після вилучення ключових кадрів також може використовуватися косинусна схожість на векторах ознак, отриманих за допомогою таких моделей, як *Vision Transformer* (*ViT*) [15]. Порівняння на основі виявлення об'єктів є більш семантично обґрунтованим, ніж порівняння на рівні пікселів, але його ефективність залежить від точності детектора об'єктів [14].

Донавчання, також відоме як тонке налаштування (*fine-tuning*), в контексті машинного навчання, зокрема для моделей виявлення об'єктів, таких як *YOLO*, передбачає використання попередньо навченої моделі, яка вже засвоїла загальні візуальні закономірності на великому наборі даних, і її подальше навчання на меншому, більш специфічному наборі даних [16]. Метою є адаптація моделі до нової задачі або набору об'єктів, які не були представлені в оригінальному навчальному наборі даних. Донавчання має кілька важливих переваг: воно вимагає менше навчальних даних, використовує знання, отримані попередньо навченою моделлю, забезпечує швидше навчання та часто призводить до кращої продуктивності [16].

Загальний процес донавчання *YOLO* включає кілька ключових етапів [16]:

• Збір та анотація даних: першим кроком є збір набору зображень, що містять конкретні об'єкти, які необхідно розпізнавати (наприклад, різні моделі танків). Кожне зображення повинно

бути ретельно анотовано, об'єкти повинні бути обведені обмежувальними рамками, і кожній рамці має бути присвоєно відповідну мітку класу (наприклад, "T-72", "*M1 Abrams*"). Збалансований набір даних з варіаціями освітлення, кутів огляду та фонів сприятиме створенню більш надійної моделі.

• Підготовка набору даних: зібрані та анотовані дані необхідно організувати у форматі, який розуміє *YOLO*. Зазвичай це передбачає наявність папок із зображеннями та відповідних текстових файлів з інформацією про анотації у певному форматі (індекс класу, нормалізовані координати центру, ширина та висота обмежувальної рамки).

• Вибір моделі: обирається попередньо навчена модель *YOLO*, яка буде слугувати основою для донавчання. Для користувачів з обмеженими ресурсами часто рекомендуються менші та швидші варіанти моделей.

• Конфігурація: за потреби, файл конфігурації моделі може бути скоригований для відображення кількості класів у новому наборі даних (наприклад, якщо потрібно розпізнавати три різні моделі танків, кількість класів має бути встановлено на три).

• Навчання: запускається процес донавчання з використанням відповідного скрипта навчання та налаштуванням гіперпараметрів, таких як швидкість навчання, розмір пакета та кількість епох.

• Оцінка: після завершення навчання оцінюється продуктивність донавченої моделі з використанням таких метрик, як середня точність (*mean Average Precision, mAP*) [17].

Сучасним та ефективним методом порівняння відеокадрів є використання семантичних ембедінгів, отриманих за допомогою таких моделей, як *CLIP* [18]. Ці ембедінги представляють високорозмірні вектори, що кодують семантичний зміст зображень та тексту в спільному просторі. Для порівняння двох відеокадрів спочатку отримують їхні *CLIP*-ембедінги, а потім обчислюють косинусну схожість між цими векторами [18]. Чим вище значення косинусної схожості, тим більша семантична схожість між кадрами. Цей підхід є стійким до варіацій візуального стилю, роздільної здатності та навіть часткових перекриттів, оскільки він фокусується на глибинному значенні вмісту кадру [19]. *CLIP*-ембедінги також можуть використовуватися для класифікації відеокадрів шляхом порівняння їхньої схожості з ембедінгами текстових запитів або міток [20]. Крім того, існують підходи до навчання часових ембедінгів відеокадрів, де ембедінг кадру залежить від його часового контексту (попередніх та наступних кадрів), що дозволяє фіксувати семантичну схожість на основі часової структури відео [19]. Такі моделі, як *Amazon Nova* та *Titan*, також використовуються для створення ембедінгів відеокадрів, аудіосегментів та транскрипцій для семантичного пошуку відео [18]. Google також розробив мультимодальні ембедінги, які безпосередньо підтримують відео, на відміну від CLIP, який зазвичай обробляє відео покадрово [21].

Використовуючи методи на основі виявлення об'єктів та на основі семантичних ембедінгів можемо покращити метод визначення ключових кадрів.

Поєднання семантичного аналізу відстані між кадрами та розпізнавання об'єктів за допомогою *YOLO* працюватиме значно краще, ніж кожен з цих підходів окремо. Ці методи доповнюють один одного, створюючи більш надійну та гнучку систему аналізу відео. Семантична відстань вловлює загальні візуальні та змістові зміни в кадрі, зосереджуючись на композиції та загальній сцені. Вона ефективно виявляє суттєві зміни в контексті відео навіть тоді, коли конкретні об'єкти залишаються незмінними. Натомість алгоритм *YOLO* концентрується на виявленні конкретних об'єктів, їх кількості, розташуванні та класифікації. Ця взаємодоповнюваність дозволяє системі охопити набагато більше аспектів аналізу, ніж будь-який окремий метод. Компенсація обмежень є ключовою перевагою комбінованого підходу. Семантична відстань може пропустити важливі зміни у випадку, коли загальна композиція кадру залишається подібною, але з'являються нові важливі об'єкти. Об'єктне розпізнавання ж має схильність пропускати значні зміни сцени, якщо набір об'єктів залишається незмінним. Об'єднуючи ці методи, ми створюємо систему, яка здатна долати обмеження кожного окремого підходу.

Зменшення помилкових спрацювань досягається завдяки зваженому поєднанню результатів обох методів. Семантичний аналіз може надмірно реагувати на незначні зміни освітлення чи ракурсу, які насправді не є змістовно важливими. Одночасно об'єктний аналіз може містити помилки розпізнавання або хибно класифікувати важливість об'єктів. Комбінування цих підходів і встановлення відповідних порогів дозволяє досягти збалансованого рішення, зменшуючи кількість хибних спрацювань.

У випадку зміни сцени при збереженні тих самих об'єктів, *YOLO* самостійно не виявить зміну, оскільки розпізнає ті ж самі об'єкти. Однак семантична відстань зареєструє зміну загального контексту та композиції. І навпаки, коли в кадрі з'являється важливий об'єкт без значної зміни загальної сцени, семантична відстань може не помітити цієї зміни, але *YOLO* зафіксує появу нового значущого об'єкта.

Надійність при різних умовах зйомки є ще однією суттєвою перевагою. При поганому освітленні або при наявності візуальних шумів ефективність *YOLO* може знижуватися, але семантичний аналіз здатний все ще виявляти значні зміни. Натомість при схожих візуально, але різних за змістом сценах, *YOLO* допоможе ідентифікувати ключові об'єкти, які відрізняють ці сцени одна від одної.

На рис. 2 представлена структуру даного підходу. Отримуємо на вхід відео та вилучаємо з нього кадри із певною частотою (10 кадрів для коротких відео, 30 і більше для більш довгих задля прискорення процесу). Далі ці кадри обробляємо за допомогою семантичного та об'єктного аналізу.

Найбільшою перевагою комбінованого підходу є гнучкість налаштування для різних типів відео. Для документальних фільмів можна збільшити вагу розпізнавання об'єктів, щоб краще виявляти появу важливих персонажів або предметів. Для художніх фільмів доцільно підвищити вагу семантичної відстані, щоб краще вловлювати зміни настрою та атмосфери. Можна також встановлювати пріоритети для певних класів об'єктів залежно від контексту аналізу, наприклад, надаючи перевагу розпізнаванню людей у документальних фільмах про соціальні теми.

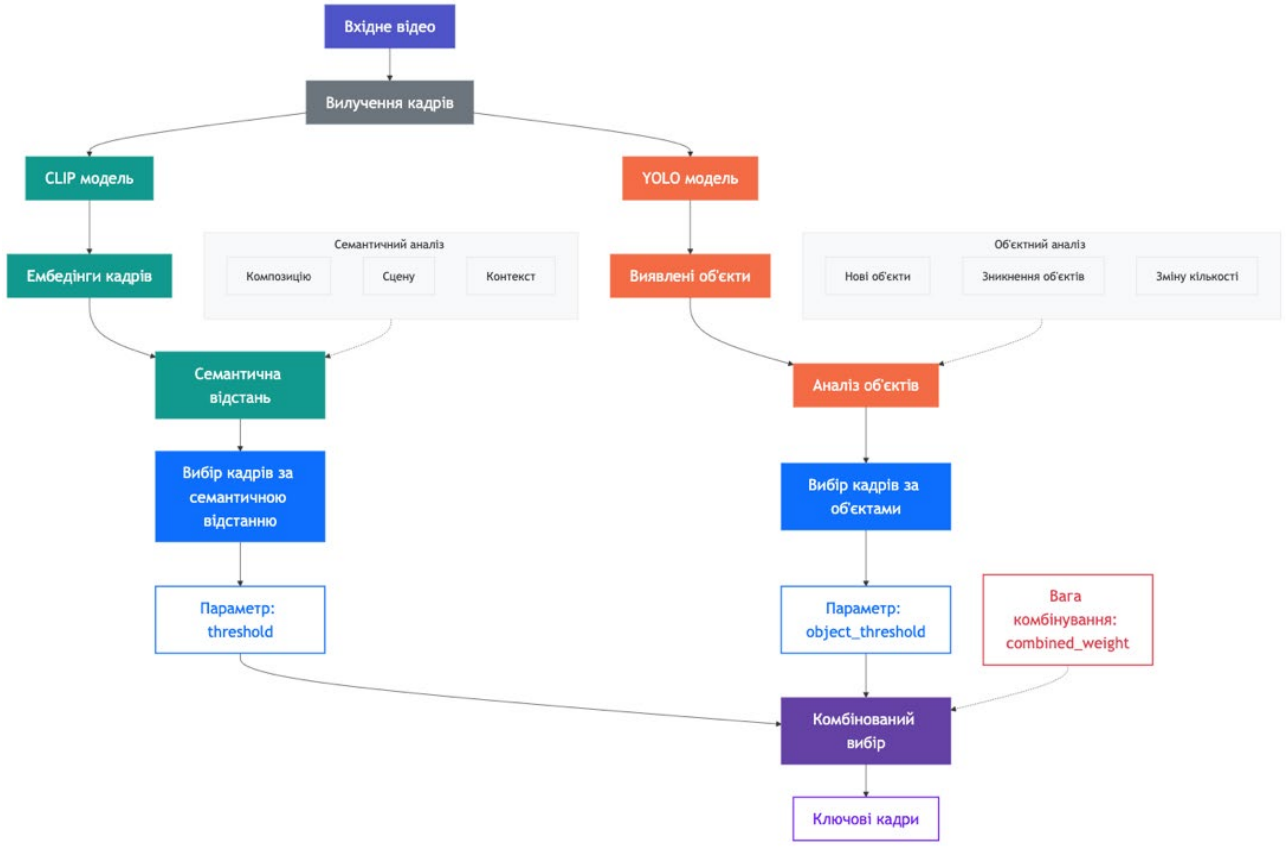

Рис. 2. Структура підходу із використанням семантичного та об'єктного аналізу

Процес оцінки важливості кадру ґрунтується на п'яти ключових критеріях:

1. Поява нових класів об'єктів – коли в кадрі з'являються об'єкти нових типів, яких не було в попередньому кадрі.

2. Зникнення класів об'єктів – коли певні типи об'єктів перестають бути присутніми в кадрі.

3. Зміна кількості об'єктів – суттєве збільшення або зменшення кількості об'єктів одного класу.

4. Поява пріоритетних об'єктів – виявлення об'єктів із заздалегідь визначених важливих класів (наприклад, людей або транспортних засобів).

5. Рівень впевненості у виявленні – висока впевненість моделі щодо виявлених об'єктів.

Кожен критерій нормалізується до значення між 0 і 1, а потім зважується відповідно до важливості:

- поява нових класів: ~30 %;
- зникнення класів: ~15 %;
- зміна кількості: ~15 %;
- пріоритетні об'єкти: ~30 %;
- рівень впевненості: ~10 %.

Фінальна оцінка обчислюється як зважена сума всіх компонентів. Кадри, оцінка яких перевищує встановлений поріг, вважаються ключовими.

Для підвищення ефективності об'єктна оцінка часто комбінується з оцінкою семантичної відстані між кадрами через формулу:

$$C = (W_s \times D_s) + (w_o \times O) \quad (2)$$

де C – комбінована оцінка; w_s – вага семантичної оцінки; D_s – семантична відстань; w_o – вага об'єктної оцінки; O – об'єктна оцінка.

Цей підхід забезпечує збалансоване виявлення ключових моментів у відео на основі як змісту сцени, так і конкретних об'єктів.

У підсумку, такий комбінований підхід створює більш надійну, адаптивну систему, яка забезпечує глибше розуміння змісту відео та ефективніше виявлення справді важливих кадрів у широкому спектрі відеоматеріалів. Ця синергія двох методів дозволяє охопити як загальні, так і специфічні аспекти візуального наративу, що робить аналіз відео більш комплексним і точним.

Висновки

У роботі розглянуто проблеми ефективного пошуку та фільтрації інформації у відеоконтенті в контексті *OSINT* в умовах зростання обсягу та різноманітності медіаданих. Проаналізовано існуючі методи порівняння відеокадрів: на рівні пікселів, на основі ознак, перцептивне хешування, виявлення об'єктів та семантичні ембедінги, визначено їх переваги та обмеження.

Запропоновано комбінований підхід, що інтегрує семантичний та об'єктний аналіз, який демонструє більшу ефективність порівняно з використанням окремих методів. Розроблено метод оцінки важливості кадру на основі п'яти ключових критеріїв, а також формулу для обчислення комбінованої оцінки як зваженої суми семантичної відстані та об'єктної оцінки.

Перспективними напрямами подальших досліджень є розробка методів автоматичного налаштування вагових коефіцієнтів для різних типів медіаконтенту, інтеграція аудіоаналізу для мультимодального пошуку, а також дослідження можливостей донавчання моделей виявлення об'єктів для конкретних задач *OSINT*-розслідувань. Окремий інтерес становить розробка ефективних методів індексації та зберігання векторних представлень медіаконтенту для забезпечення швидкого пошуку за семантичною схожістю у великих обсягах даних.

ЛІТЕРАТУРА

[1] Dupré M. H. (2024). Huge Proportion of Internet Is AI-Generated Slime, Researchers Find. https://futurism.com/the-byte/internet-ai-generated-slime

[2] Shih-Fu M. Recent Advances and Challenges of Semantic Image/Video Search. IEEE International Conference.

[3] Sutskever I. Learning Transferable Visual Models From Natural Language Supervision, 2021.

[4] Palucha S. Understanding OpenAI's CLIP model, 2024. https://medium.com/@paluchasz/understanding-openais-clip-model-6b52bade3fa3

[5] Deng J. A Unified Model for Video Understanding and Knowledge Embedding with Heterogeneous Knowledge Graph Dataset, 2023.

[6] Ma Q. X. L. Perceptual hashing method for video content authentication with maximized robustness, 2021.

[7] Alaa T. Video Summarization Techniques: A Comprehensive Review. In Scitepress, 2024.

[8] Castellano B. Python and OpenCV-based scene cut/transition detection program & library, 2025. https://github.com/Breakthrough/PySceneDetect?tab=readme-ov-file

[9] Berrios I. Introduction to Motion Detection, (2023). https://medium.com/@itberrios6/introduction-to-motion-detection-part-1-e031b0bb9bb2

[10] Iqbal S. Proving Reliability of Image Processing Techniques in Digital Forensics Applications. Security and Communication Networks, 2022. https://doi.org/10.1155/2022/1322264

[11] X-Ways. (n.d.). X-Ways Forensics: Integrated Computer Forensics Software. https://www.x-ways.net/forensics/index-m.html

[12] Tan, Y. (1999). A Framework for Measuring Video Similarity and Its Application to Video Query by Example. Department of Electrical Engineering Princeton University.

[13] Sommerville G. Automatically compare two videos to find common content. Amazon, 2021. https://aws.amazon.com/blogs/media/metfc-automatically-compare-two-videos-to-find-common-content/.

[14] Kender R. L. a. P. J. Determining Video Similarity With Object Detection, 2021.

[15] Luo Z. Frame Comparison and Frame Clustering with Vision Transformer and K-Means on COVID-19 News Videos from Different Affinity Groups. Columbia University, 2023.

[16] Kwasny R. Fine-tuning YOLO in a Cafe: Custom Object Detection, 2024. https://rafalkwasny.com/yolo-object-detection

[17] Alkandary A. S. K. A Comparative Study of YOLO Series (v3–v10) with DeepSORT and StrongSORT: A Real-Time Tracking Performance Study, 2025. https://doi.org/10.3390/electronics14050876

[18] Daniel M. W. A. Video semantic search with AI on AWS. Amazon, 2025. https://aws.amazon.com/blogs/media/video-semantic-search-with-ai-on-aws/.

[19] Ramanathan V. Learning Temporal Embeddings for Complex Video Analysis, 2015. https://doi.org/10.48550/arXiv.1505.00315
[20] Radford A. Learning Transferable Visual Models From Natural Language Supervision, 2021. https://doi.org/10.48550/arXiv.2103.00020
[21] Pendlebury B. O. M. Exploring Video Search with OpenOrigins. Datastax, 2024. https://www.datastax.com/blog/video-search-with-openorigins-frame-search-versus-multi-modal-embeddings

Сокиран Ю. Ю., Пархоменко І. І.

ЗАСТОСУВАННЯ СЕМАНТИЧНОГО ТА ОБ'ЄКТНОГО ПОШУКУ ДЛЯ ОПТИМІЗАЦІЇ ПОШУКУ ПО МЕДІА ЗАДЛЯ OSINT РОЗСЛІДУВАНЬ

Сучасний інформаційний простір характеризується експоненціальним зростанням обсягів різнорідних медіаданих. Це особливо актуально після широкого впровадження великих мовних моделей та підвищення активності користувачів соціальних мереж. Цей процес зростання створює принципово нові виклики для фахівців у сфері розвідки на основі відкритих джерел (OSINT), які повинні оперативно обробляти та аналізувати величезні масиви текстової, аудіо, фото та відеоінформації. Водночас якість цих даних не завжди покращується пропорційно їхній кількості, що вимагає розробки нових підходів до ефективної фільтрації та пошуку релевантної інформації в умовах інформаційного перевантаження. Одним із інструментів, які можуть допомогти в цьому, є комп'ютерний зір, який дозволяє автоматизувати процеси аналізу візуального контенту та виявлення ключових елементів у медіаданих. Застосування методів комп'ютерного зору в поєднанні з сучасними алгоритмами машинного навчання відкриває нові можливості для підвищення ефективності OSINT-розслідувань. У статті розглянуто проблеми пошуку та фільтрації інформації в контексті розвідки на основі відкритих джерел в умовах зростання обсягів різнорідних даних. Проаналізовано існуючі методи порівняння відеокадрів, включаючи методи на рівні пікселів, порівняння на основі ознак, перцептивне хешування, використання алгоритмів виявлення об'єктів та семантичних ембедінгів. Запропоновано комбінований підхід, який інтегрує семантичний та об'єктний аналіз для ефективного виявлення ключових кадрів у відеоматеріалах. Розроблено метод оцінки важливості кадру на основі п'яти критеріїв: поява нових класів об'єктів, зникнення класів об'єктів, зміна кількості об'єктів, поява пріоритетних об'єктів та рівень впевненості у виявленні. Комбінована оцінка обчислюється як зважена сума семантичної відстані та об'єктної оцінки, що забезпечує збалансоване виявлення ключових моментів у відео. Такий підхід створює адаптивну систему аналізу, яка забезпечує глибше розуміння змісту відеоматеріалів для OSINT-розслідувань, ефективно відфільтровуючи нерелевантні дані.

Ключові слова: кібербезпека; *OSINT*; розвідка з відкритих джерел; пошук у медіа; аналіз відео; семантичний пошук; об'єктний пошук; виявлення об'єктів; перцептивне хешування; фільтрація інформації.

Sokyran Yu., Parkhomenko I.

APPLICATION OF SEMANTIC AND OBJECT SEARCH FOR OPTIMIZING MEDIA SEARCH IN OSINT INVESTIGATIONS

The modern information space is characterized by growth in volumes of diverse media data, which has become particularly acute after the widespread adoption of large language models and increased user activity on social media. This process of growth creates fundamentally new challenges for professionals in the field of Open Source Intelligence (OSINT), who must promptly process and analyze big arrays of textual, audio, photo, and video information. At the same time, the quality of this data does not always improve proportionally to its quantity. This creates need for the development of new approaches for effective filtering and searching of relevant information under conditions of information overload. One of the tools that can help in this regard is computer vision. Computer vision allows for the automation of visual content analysis processes and the identification of key elements in media data. The application of computer vision methods combined with modern machine learning algorithms opens new possibilities for improving the efficiency of OSINT investigations, so they must be studied further. The article examines the challenges of searching and filtering information in the context of Open Source Intelligence amid increasing volumes of data. Existing methods of video frame comparison are analyzed, including pixel-level methods, feature-based comparison, perceptual hashing, object detection algorithms, and semantic embeddings. A combined approach that integrates semantic and object analysis for effective key frame detection in video materials is proposed. A method for evaluating frame importance has been developed based on five criteria: appearance of new object classes, disappearance of object classes, changes in object quantity, appearance of priority objects, and detection confidence level. The combined score is calculated as a weighted sum of semantic distance and object evaluation, providing balanced detection of key video moments. This approach creates an adaptive analysis system that ensures a deeper understanding of video content for OSINT investigations, effectively filtering out irrelevant data.

Keywords: cybersecurity; OSINT; open source intelligence; media search; video analysis; semantic search; object search; object detection; perceptual hashing; information filtering.

Стаття надійшла до редакції 01.06.2025 р.
Прийнято до друку 11.06.2025 р.