

VLM-відеоаналітика —

новий тренд на ринку відеоспостереження



Ще кілька років тому відеоаналітика була обмежена певним набором функцій. Створювати індивідуальні сценарії під конкретні задачі бізнесу чи державних структур було майже неможливо. Але стрімкий розвиток штучного інтелекту змінив правила гри. VLM-відеоаналітика стає новим технологічним трендом.

На яку відеоаналітику очікують замовники?

Майже кожна вертикаль ринку постійно шукає рішення з відеоаналітики. Безпечні міста очікують на відеоаналітику, що автоматично виявляє порушення правил перебування людей в громадських місцях та визначає нестандартну поведінку людей, і таку, що автоматично виявляє порушення правил дорожнього руху та парковок (включно з оплатою паркомісць). Промисловість шукає рішення з розпізнавання засобів індивідуального захисту (каска, маски, рукавички, пояси), що запобігає небезпечним інцидентам на виробництві. Банки бажають

отримати повний контроль за робочим часом працівників і якістю їх роботи, а також аналіз поведінки клієнтів у відділеннях. Ритейл бажає автоматизувати контроль за наповненістю полиць в магазинах і аналітику, яка б дозволила впливати на продажі. Замовники очікують, що все це допоможе покращити їхню операційну ефективність, звернути увагу операторів на нестандартні події, а також автоматизувати бізнес-процеси.

Цей список очікувань, хоча й демонструє високу зацікавленість, водночас створює значний виклик для ринку. Оскільки впровадження штучного інтелекту у відеоаналітику робить діапазон можливих сценаріїв майже

безмежним, це призводить до певного «технологічного розгублення» серед кінцевих замовників. Враховуючи, що переважна більшість замовників корпоративного рівня досі не має досвіду тестування та впровадження відеоаналітики, їхні розмови з інтеграторами, дистриб'юторами та вендорами відбуваються у площині емоцій та бажань, а не технічних параметрів, технологій і сукупної вартості рішень.

На що технічно здатна відеоаналітика визначається, на жаль, не бажаннями замовника, а поколіннями аналітик, які можуть бути застосовані для вирішення завдань. Покоління відеоаналітик ґрунтуються виключно на розвитку технологій.

Перше покоління відеоаналітики: Motion Detection

За більш ніж 25 років свого існування відеоаналітика пройшла кілька етапів розвитку. Перше покоління відеоаналітики було створено на початку 2000-х років і складалося з простих алгоритмічних правил: детекція руху (motion detection), детекція втрати сигналу від камери (network disconnect) та перетин віртуальних ліній (crossline). Ці правила переважно працювали на борту IP-камер професійних виробників. Ця аналітика була дуже чутливою до освітлення та зміни кадру (тіні, сонце) і тому мала часті хибні спрацювання.

Друге покоління відеоаналітики: Machine Learning

2-ге покоління відеоаналітики отримало розвиток у 2010-х років та використовувало принципи машинного навчання ML (Machine Learning), що дозволило швидше та точніше аналізувати поле зору камер, визначати та класифікувати об'єкти в кадрі (object detection): люди, автомобілі та предмети. Це дозволило створити цілу низку правил охоронної та бізнес-відеоаналітики, що значно покращило роботу операторів відеоспостереження. Правила відеоаналітики 2-го покоління:

- детекція об'єкта при перетині лінії (Virtual line/Cross line);
- детекція входу/виходу об'єкта з зони (Enter/Exit/Intrusion);
- детекція руху та знаходження людини в зоні певний час (Loitering);

- детекція руху об'єкта у певному напрямку (Directional detection/Tailgating);
- виявлення залишеного/зниклого об'єкта (Appear/Disappear);
- підрахунок людей (People counting);
- детекція зламу/перекриття поля зору камери (Tampering);
- детекція розфокусування (Defocus detection).

Третє покоління відеоаналітики: Deep Learning

3-тє покоління відеоаналітики з'явилося у 2015–2019 роках за рахунок технології глибокого навчання DL (Deep Learning), адже відеопотоки з IP-камер чи VMS оброблялися нейронними мережами, які набули здатності розпізнавати певні шаблони у величезних обсягах даних. DL-аналітика дозволила точніше класифікувати об'єкти, а нейронні мережі дали змогу вирішувати більш комплексні завдання порівняно з ML-відеоаналітикою (рис. 1).

До правил відеоаналітики 2-го покоління додалися нові правила відеоаналітики 3-го покоління:

- детекція та відстеження об'єкта в полі зору камери (Digital autotracking);
- зупинка автомобіля в зоні (Stopped vehicle);
- детекція туману (Fog detection);
- детекція облич (Face detection);
- розпізнавання облич (Face recognition);
- підрахунок людей в черзі (Queue);
- детекція скупчення людей (Crowd detection);
- теплові карти (Heatmap).

Різні розробники відеоаналітики пропонували свої варіації цих правил та їхніх назв, але суть залишалася незмінною: в основі роботи правил були чіткі математичні алгоритми, які інтегратор чи замовник не могли змінювати. Результат роботи відеоаналітики 3-го покоління дуже сильно залежав від геометрії сцени та роздільної здатності відеопотоку, який обробляли алгоритми. Наприклад, розробники відеоаналітики розпізнавання номерів гарантували 96–98% розпізнавання за умов дотримання інструкції (кути встановлення IP-камери, швидкість затвора, освітленість, тип матриці, роздільна здатність та інші параметри). Якщо встановлені на об'єкті IP-камери не задовольняли вимогам



Рис. 1. Порівняння DL та ML-відеоаналітики

Ліворуч: ML-аналітика (визначила багато зон, де нібито є рух, хоча рухається в кадрі лише людина). Праворуч: DL-аналітика ігнорує фон і виділяє лише людину, її обличчя та автомобілі в кадрі.

розробників відеоаналітики, обладнання потребувало заміни, що збільшувало бюджет її впровадження. Іншого шляху не було.

Четверте покоління відеоаналітики: AI-based video analytics

Розвитку 4-го покоління відеоаналітики (AI-based video analytics) сприяла пандемія COVID-19 і завдання розпізнавання людей в масках, дотримання соціальної дистанції та підрахунку людей у приміщенні. З 2020 року і по теперішній час маємо AI-відеоаналітику, що базується на покращеній технології глибокого навчання, адже воно стало можливим не лише на серверах, а і на борту IP-камер. AI-відеоаналітика здатна (рис. 2):

- здійснювати класифікацію об'єкта (людина/транспортний засіб);
- визначати параметри людини (стать, колір та сумка);
- визначати особливості обличчя (вік, стать, маска та окуляри);
- ідентифікувати транспортний засіб (автомобіль / автобус / вантажівка / мотоцикл / велосипед).

Завдяки впровадженню відеоаналітики 4-го покоління замовники отримали новий підхід до роботи з системою відеоспостереження та інтегрованими системами безпеки. Покращилась точність всіх існуючих правил аналітики, що зменшило хибні спрацювання та дозволило автоматизувати робочі місця операторів. Найбільші зміни торкнулися роботи операторів з відеоархівом. Завдяки метаданим з атрибутами оператори отримали можливість не просто «промотувати» години запису, а миттєво знаходити конкретних людей, автомобілі чи послідовність подій. В разі підвищились швидкість та точність розслідувань, експорт відеодоказів та загальна операційна ефективність.

Четверте покоління стало можливим завдяки широкому впровадженню відеокарт з GPU в серверах та покращенню архітектури процесорів камер, що дозволило замовникам обирати не тільки серверну, а й AI edge-аналітику на борту камер багатьох виробників.

Завдяки цьому стрімкому розвитку в 2021–2024 роках стало можливим більше, ніж відеоаналітика пропонувала замовникам за всі 25+ років свого існування.

Ось деякі приклади впроваджених рішень з відеоаналітикою 4-го покоління.

- Безпечні міста: порушення швидкісного режиму, контроль проїзду на червоне світло, пошук викрадених автомобілів, підрахунок кількості автомобілів на перехрестях.
- Промисловість: автоматизований в'їзд автотранспорту та прохід персоналу на територію, зважування продукції при завантаженні на транспортні засоби, контроль людей без касок та масок у промислових зонах.
- Ритейл: підрахунок відвідувачів, підрахунок людей в чергах, зони зацікавленості покупців, маршрути покупців, автоматичний виїзд автомобілів після оплати паркування.
- Банки: контроль касових операцій та транзакцій в банкоматах на відео, контроль розмов між персоналом та клієнтами, ідентифікація клієнтів за розпізнаним обличчям, автоматизований в'їзд банківського автотранспорту на територію.
- Державні проекти: контроль проходження людьми та автотранспортом державних кордонів, контроль лісових пожеж.

Чи стало можливим вирішити будь-які завдання замовника з відеоаналітикою 4-го покоління? Безперечно, ні.

З розвитком AI на запит замовників з'явився новий тренд — **персоналізована відеоаналітика** (Custom-based Video Analytics). Суть цього тренду полягає в тому, що завдяки deep learning і гнучким інструментам розробки, а також швидким математичним обчисленням стало можливим не просто запропонувати готові алгоритми чи правила, а створити унікальні, кастомізовані рішення під конкретний запит замовника, об'єкт чи бізнес-процес.

Якщо замовник хоче, **щоб «відеоаналітика все сама робила», це є запит на кастомізацію**, тобто на розроблення персональної відеоаналітики, і це означає, що у замовника є окремі бюджети та час впроваджувати на своєму об'єкті індивідуальну розробку.



Рис. 2. AI-відеоаналітика 4-го покоління: класифікація об'єктів та визначення атрибутів

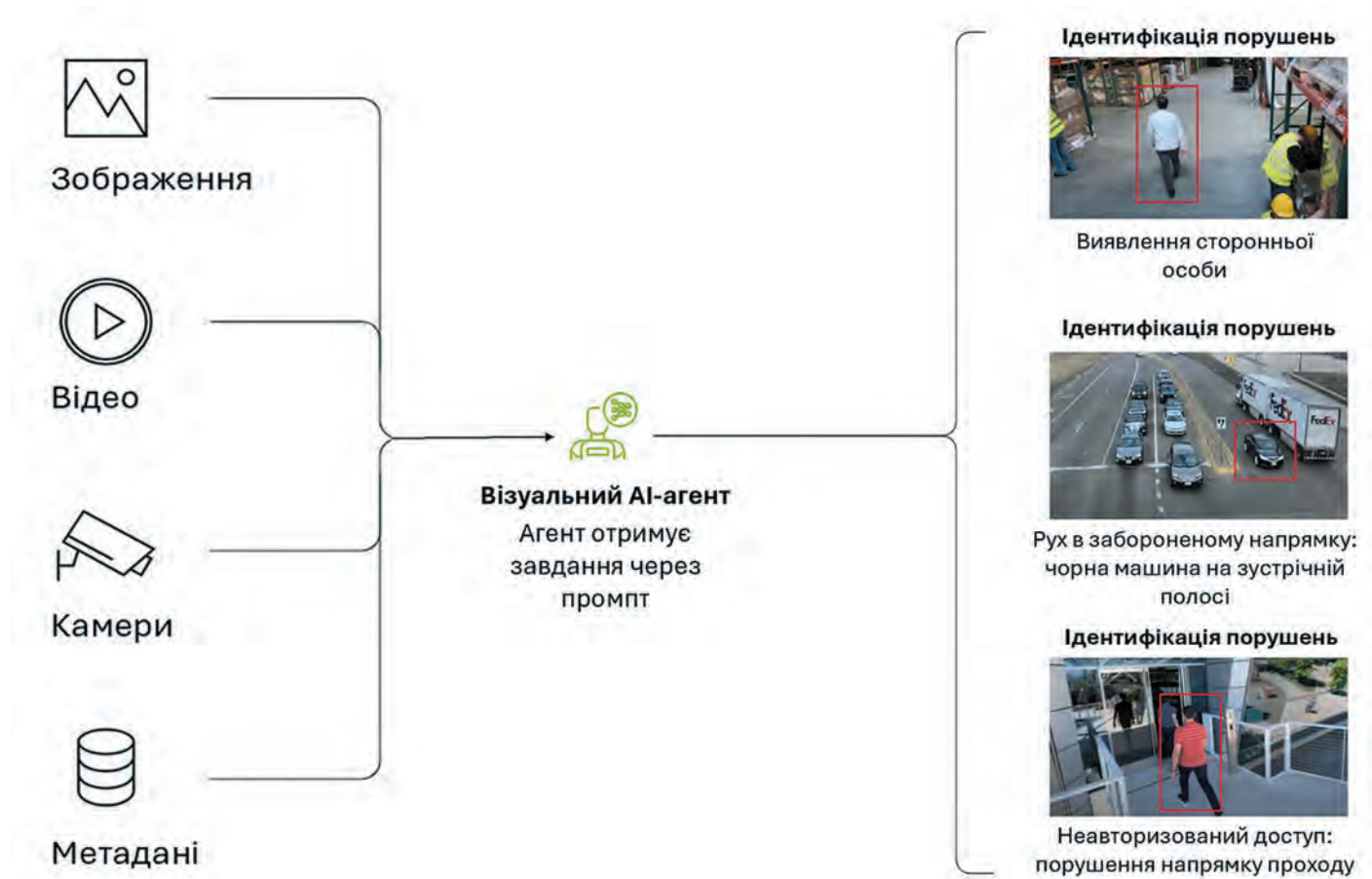


Рис. 3. Принцип роботи VLM-відеоаналітики (джерело: NVIDIA)

Гарною новиною є те, що за останні пару років розробники відеоаналітики усвідомили, що їх вже не десятки, а сотні, і боротьба за замовників зробила їх гнучкішими і готовими іти в кастомізацію. Якщо промислового замовнику дійсно треба виявляти захисні каски, рукавички, окуляри і він готовий виділити на це бюджети, розробники відеоаналітики готові працювати над навчанням нейронних мереж на базі конкретного об'єкта замовника. Це означає впровадження на об'єкті пілотного проекту тривалістю 2–3 місяці та залучення розробників до процесу покращення результатів. Щоб розробники не витрачали свій час марно, вони пропонують замовникам придбати готові правила відеоаналітики, і після цього другим етапом ідуть в кастомізацію, яка передбачає використання AI. Це завжди індивідуальний розрахунок вартості розробки залежно від обсягу роботи для розробників.

Цим методом успішно вирішують свої задачі замовники в Європі. Всі публічні кейси зі складними сценаріями, описи яких можна знайти у відкритому доступі — наприклад, відомі приклади нестандартної відеоаналітики (видача сповіщення з разі заповнення сміттєвих баків у місті, детекція сміття чи навіть розпізнавання типів військової техніки дронами) — це Custom-based AI Video Analytics, над впровадженням якої розробники працювали окремо та ретельно вчили нейромережі. Такі проекти індивідуальні і завжди впроваджуються за допомогою value-added дистриб'юторів, які забезпечують комунікацію всіх зацікавлених сторін, а також професійних інтеграторів, які здійснюють інсталяцію, налаштування та тестування кастомних рішень.

П'яте покоління відеоаналітики: VLM-відеоаналітика

Візуальні мовні моделі (VLM, Vision language models) — це найновіші розробки AI-індустрії, які вже стрімко починають впроваджуватися у сфері відеоаналітики. **VLM-моделі** — це мультимодальні генеративні моделі штучного інтелекту, здатні розуміти й обробляти відео, зображення та текст.

VLM — це система штучного інтелекту, побудована шляхом поєднання великої мовної моделі (LLM) із візуальним еncoderом, що надає мовній моделі здатність «бачити» (рис. 3).

Завдяки цьому VLM-моделі вміють аналізувати відео, розуміти та описувати його зміст, відповідати на запитання людини (комунікація нагадує вбудований ChatGPT) та створювати складні аналітичні правила без попереднього навчання. Можливості, швидкість та точність VLM-відеоаналітики перевищують можливості AI-відеоаналітики четвертого покоління, оскільки аналізується не просто те, що камера бачить, а те, як це побачене відповідає (чи збігається) зі змістом текстового запиту оператора. Завдяки такому принципу оператор може поставити складні завдання і отримати конкретні результати, ґрунтуючись на контекстному розумінні сцени (рис. 4). Запит: «Чи є щось нетипове в третьому ряду?», відповідь: «Полиця обвалилася, коробки впали з полиці о 15:30 та заблокували ряд».

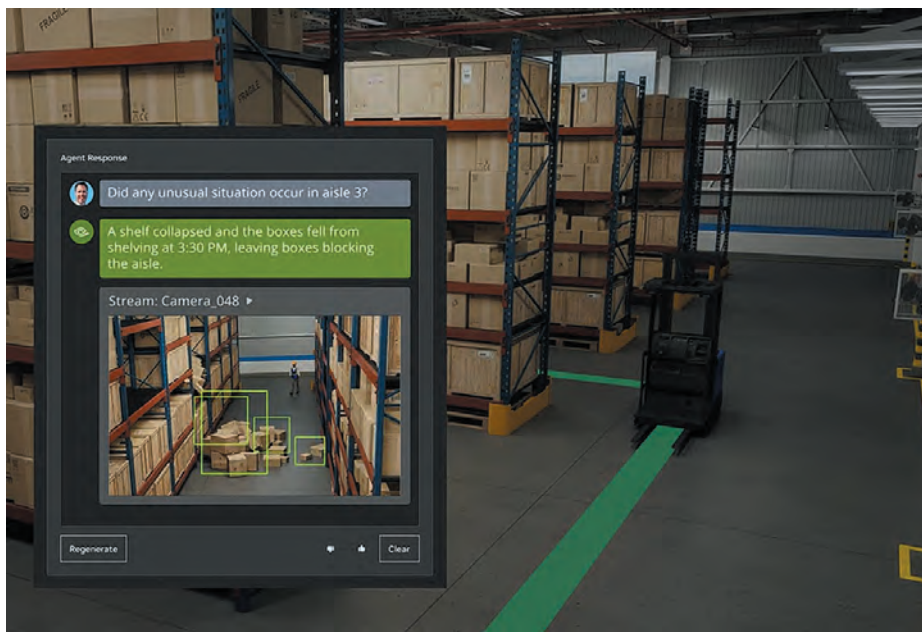


Рис. 4. Принцип комунікації оператора з VLM-відеоаналітикою (джерело: NVIDIA)

Головна умова наразі — оператор має вміти ефективно спілкуватися з AI, тобто повинен бути в змозі написати коректний текстовий запит (промпт) у чаті, щоб відеоаналітика могла його відпрацювати. Оскільки базові VLM-моделі навчалися переважно на англійських даних, найвища точність та надійність правил наразі забезпечується при використанні англійської мови.

Далі розберемо на реальних прикладах, як відеоаналітика може вирішити конкретні завдання замовників.

Відеоаналітика для виявлення засобів індивідуального захисту. Практичний розбір

Міжнародна організація праці (англ. International Labour Organization, ILO) має статистику, що щорічні світові економічні втрати від нещасних випадків на виробництві та професійних захворювань становлять приблизно 4% світового валового внутрішнього продукту (ВВП). Понад 3 мільйони працівників щороку гинуть внаслідок нещасних випадків на роботі та професійних захворювань. Крім того, близько 395 мільйонів працівників щороку отримують травми. Згідно з аналізом, нещасні випадки на виробництві є причиною 330 000 смертей.

У будівельній та виробничій сферах травми голови залишаються однією

з головних причин смертності та інвалідності. Хоча частка таких травм може варіюватися, вони є одними з найдорожчих у лікуванні. Падіння з висоти є однією з основних причин смертності на будівництві та на промислових об'єктах.

Промислові замовники хочуть підвищити рівень охорони праці та знизити кількість страхових виплат через виробничі травми. Контроль носіння захисних касок, страхувальних поясів та інших засобів індивідуального захисту (англ. Personal Protective Equipment, надалі PPE) є прямим способом запобігання цим інцидентам.

На практиці вирішення цього завдання реалізується через комплексний підхід: контроль перебування працівників у робочих зонах (це завдання систем контролю доступу) і контроль PPE. Коректна побудова системи контролю доступу забезпечить перебування в робочих зонах тільки тих робітників, які мають там бути, і контроль часу їх перебування. А завдання контролю носіння PPE замовники хочуть перекласти на відеоаналітику, адже всі робочі зони зазвичай вже охоплені системою відеоспостереження. Для замовника все виглядає просто і логічно: в усіх зонах встановлені IP-камери, є відео в режимі реального часу і відеоархів. Не вистачає тільки автоматизації процесу, тобто механізму отримання тривожних сповіщень, якщо в зоні підвищеної небезпеки робітник буде без

захисної каски, рукавичок, окулярів та страхувальних поясів. Звучить просто, але насправді це непросто, адже відеоаналітика — чіткий математичний алгоритм, побудований на Deep Learning.

Якість роботи алгоритму відеоаналітики — це зона відповідальності розробників відеоаналітики. Підемо вглиб алгоритму, щоб розібратися, як насправді це працює. IP-камера передає відеопотік і декілька разів за хвилину неймережа виявляє всіх людей у кадрі (Person Detection). Далі алгоритм зосереджується на області голови кожної людини, класифікує її атрибути і визначає: «Каска» (на основі форми, кольору та візуальних ознак, які неймережа запам'ятала під час навчання), або: «Голова без каски». Якщо система класифікує об'єкт як «Каска» — все добре, якщо ж фіксується «Голова без каски» — видається тривожне сповіщення. Математичний алгоритм не розуміє, що таке «каска» чи «безпека». Він просто порівнює набір візуальних патернів (колір, контур, текстуру) з базою знань, сформованою під час навчання. Точність роботи відеоаналітики прямо залежить від якості та різноманітності даних (касок у різних ракурсах, освітленні та на різних працівниках), на яких розробники тренували свою неймережу.

Зона відповідальності замовника — забезпечити технічні умови для коректної роботи відеоаналітики, а саме — забезпечити відеопотік з камер чи з VMS для відеоаналітики з достатньою кількістю пікселів, яку займає об'єкт (в нашому випадку — каска) у кадрі. Каска має висоту приблизно 30 см. У більшості професійних систем відеоаналітики для надійного виявлення каски вона повинна мати не менше 20×20 пікселів (рекомендований параметр — 30×30 пікселів або більше).

Кількість пікселів на метр PPM (англ. Pixels Per Meter) — це одиниця вимірювання, яка використовується для визначення деталізації зображення з IP-камери на заданій відстані (рис. 5).

Розробників, які пропонують детекцію касок, — десятки. Різниця, як завжди, приховується в деталях. Камери

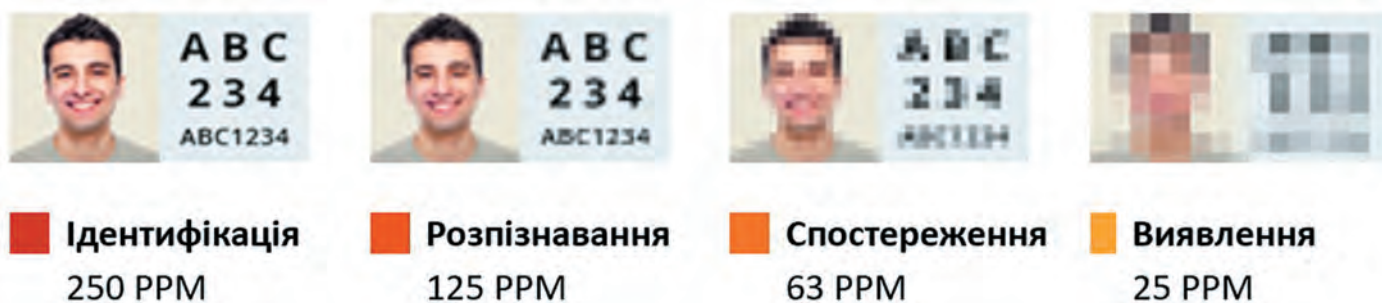


Рис. 5. Кількість пікселів на метр для ідентифікації, розпізнавання, спостереження та виявлення людини

з цією функцією пропонують деякі виробники IP-камер, але тільки на певних моделях. Оскільки у промислових замовників вже встановлені якісь IP-камери, для детекції касок у 99% випадках застосовують серверну відеоаналітику на базі окремого програмного забезпечення.

Змоделюймо типову ситуацію на об'єкті замовника для виявлення каски на голові людини: 2Мп-камера з об'єктивом 2.8 мм, встановлена на висоті 3.5 метри. Калькулятор нам показує, що відеопотік з цієї камери дозволить розпізнати наявність чи відсутність каски на голові людини на відстані від 1.5 метра (170 PPM) до 9 метрів (63 PPM). На більшій відстані каска може розпізнаватися некоректно чи не розпізнаватися взагалі (рис. 6).

Ці нескладні підрахунки дають всім зацікавленим сторонам фундаментальні математичні висновки.

Якщо замовник дійсно хоче впровадити детекцію касок, необхідно перевірити технічні умови впровадження такої відеоаналітики:

- порахувати PPM для камер в обраних зонах;
- перевірити придатність зображення від існуючих IP-камер вдень і вночі.

Далі рахується комерційна частина та сукупна вартість володіння проектом. Для цього необхідно:

- обрати 1–2 виробників відеоаналітики з правилом «детекція каски» і порахувати з ними вимоги до сервера відеоаналітики;
- провести пілот та перевірити, як працює відеоаналітика вдень, вночі та в умовах змінної освітленості;

- порівняти вартість/точність роботи різних виробників між собою;
- отримати консультації професійних інтеграторів щодо впровадження відеоаналітики: як окремого продукту чи разом з VMS;
- порахувати експлуатаційні витрати на відеоаналітику та VMS на наступні 3–5 років.

Цей техніко-економічний процес необхідно пройти замовнику, якщо він хоче вирішити лише одне практичне завдання — отримувати тривожні сповіщення у разі відсутності касок на головах його співробітників. Для забезпечення інших потреб у PPE-відеоаналітиці, включно з детекцією захисних рукавичок, окулярів та поясів, треба або звертатися до розробників по Custom-based AI Video Analytics, або впроваджувати VLM-відеоаналітику, що передбачає, що у замовників є відповідні бюджети.

VLM-моделі у реальному часі здатні аналізувати відеопотоки з IP-камер та виявляти події чи поведінкові аномалії на виробництві: людина без каски чи жилета, людина в забороненій зоні, людина наблизилася до механізмів та машин, виявлення перешкод у певних зонах тощо. Ці моделі дозволяють операторам створювати текстові правила, які відображають складні, контекстно-залежні людські наміри.

Наприклад (рис. 7), запит: «Визнач будь-які небезпечні дії та зафіксуй екологічні ризики». Відповідь: «Аналіз безпеки робочого майданчика вітряка. Огляд. Відео показує людину, яка працює на анемометрі вітротурбіни. Працівник під'єднаний до страхувального троса. Ключові висновки:

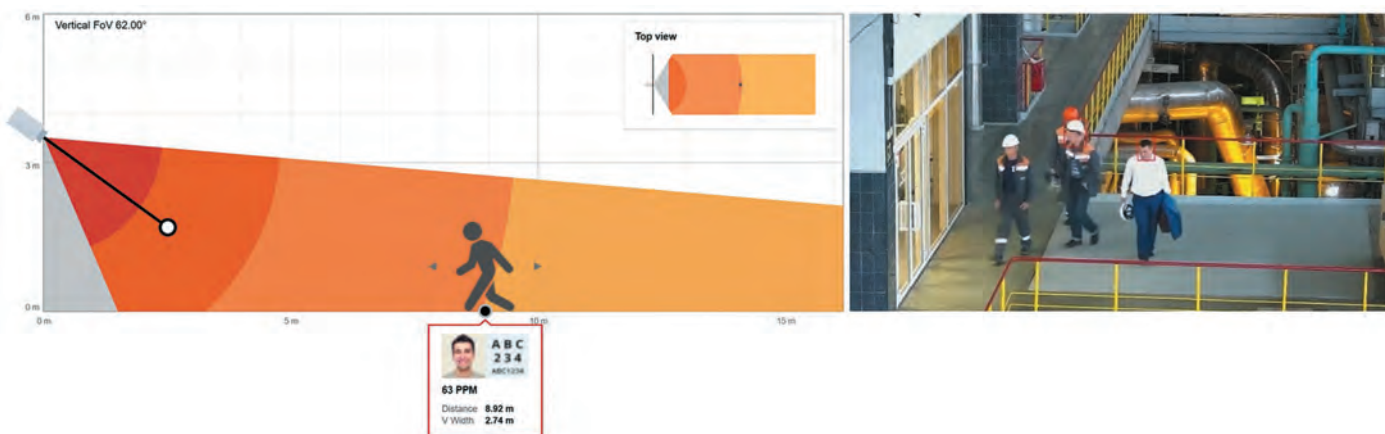


Рис. 6. Приклад детекції каски



Рис. 7. Приклад роботи VLM-відеоаналітики для промисловості (джерело: NVIDIA)

- використання страхувальної системи: працівник надійно закріплений за допомогою страхувального пояса;
- навколишнє середовище: відсутні негайні екологічні ризики, такі як екстремальні погодні умови;
- захист працівника: відсутні видимі порушення у використанні засобів індивідуального захисту».

Пройшовши шлях від простої детекції руху до складної DL-класифікації, відеоаналітика досягла етапу, де функціональність визначається не жорсткими алгоритмами, а інтелектуальним розумінням сцени. П'яте покоління, представлене VLM-моделями, здатне виконати ті завдання, які були недосяжні для всіх попередніх чотирьох поколінь разом узятих: перехід від виявлення об'єктів до розуміння намірів і контексту сцени.

Наразі очевидним є істотне відставання китайських виробників камер у сфері програмного забезпечення. Їхні пропозиції обмежуються вбудованою в камери відеоаналітикою другого покоління (ML) і лише на окремих моделях камер — обмеженими функціями третього покоління (DL), що робить їхню продукцію непридатною для впровадження високоточних кастомізованих рішень та нових VLM-систем.

У досить складних позиціях опинилися традиційні виробники закритих систем відеоспостереження Avigilon та Bosch, адже в них є тільки певний набір правил DL-відеоаналітики і немає кастомної чи VLM-відеоаналітики. Axis Communications має певну гнучкість за рахунок завантаження на борт камер аналітики своїх технологічних партнерів. Hanwha Vision пропонує варіативності: і набір аналітики на борту камер, і завантаження на борт камер аналітики своїх технологічних партнерів.

Серверна відеоаналітика може запропонувати значно більше, ніж камерна. Абсолютну перевагу отримали виробники відкритих платформ VMS, особливо Milestone Systems, яка має десятки технологічних партнерів-розробників відеоаналітики і спільний з NVIDIA інноваційний проект Hafnia.

Впровадження VLM-відеоаналітики, яке є найбільш наближеним до ідеального бачення замовників, вимагає відповідних інвестицій у програмно-апаратне забезпечення, адже вся професійна DL та VLM-відеоаналітика не зможе працювати на примітивних NVR та потребує окремих відеосерверів, а для зручності роботи з великими системами відеоспостереження — ще і переходу на єдину VMS-платформу відкритої архітектури. Тільки комплексний підхід забезпечить швидку та зручну обробку даних і впровадження AI у відеоспостереження.

Щоб коректно оцінити сукупну вартість володіння (TCO) та технологічну готовність об'єкта до відеоаналітики 4-го та 5-го поколінь, необхідно залучати професійних системних інтеграторів з відповідним практичним досвідом.

VLM-відеоаналітика — це не просто новий продукт, це нова філософія безпеки та операційної ефективності, що дозволяє кінцевому користувачеві вперше взяти керування аналітичними можливостями системи у свої руки, використовуючи найпростіший інтерфейс у світі — текст. Це крок до того, щоб відеосистема перетворилася з пасивної системи зберігання відеоданих на активного, інтелектуального помічника.

Альона ШВЕЦОВА,
незалежний експерт з систем безпеки,
[cctvmadonna](https://cctvmadonna.com)