

БАКАЛАВРСЬКА РОБОТА

БР. ІІ - 85.00.00.000 ІЗ

Група ІІ-22-3

Тріщ Тарас

2026

Івано-Франківський національний технічний університет нафти і газу

Факультет інформаційних технологій

Кафедра інженерії програмного забезпечення

Тріщ Тарас Вікторович

(прізвище, ім'я, по батькові)

УДК 004.942

(індекс)

БАКАЛАВРСЬКА РОБОТА

Застосування штучного інтелекту для автоматичного монтажу та редагування відео-контенту

(назва роботи)

Інженерія програмного забезпечення

(назва освітньої програми)

121 - Інженерія програмного забезпечення

(шифр і назва спеціальності)

Робота містить результати власних досліджень, використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело:

Здобувач освітнього ступеня Тріщ Т.В.
(підпис, ініціали та прізвище здобувача)

Науковий керівник Процюк Галина Ярославівна, асистент
(підпис, прізвище, ім'я, по батькові, науковий ступінь, вчене звання керівника)

Допущено до захисту
Завідувач кафедри

доц. Бандура В.В.
(посада) (підпис) (дата) (ініціали та прізвище)

Івано-Франківський національний технічний університет нафти і газу
Інститут, факультет інформаційних технологій
Кафедра інженерії програмного забезпечення
Ступінь вищої освіти бакалавр
Спеціальність 121 – Інженерія програмного забезпечення

ЗАТВЕРДЖУЮ:

Зав. кафедрою

ІІЗ

доцент.

В.В. Бандура

“ ____ ” _____ 2026 р.

ЗАВДАННЯ

НА БАКАЛАВРСЬКУ РОБОТУ СТУДЕНТОВІ

Тріщ Тарас Вікторович

(прізвище, ім'я, по-батькові)

1. Тема проекту (роботи) "Застосування штучного інтелекту для автоматичного монтажу та редагування відео-контенту"

керівник проекту (роботи) Процюк Галина Ярославівна, асистент

затверджені наказом вищого навчального закладу від “ 21 ” квітня 2026 р. № 179/7

2. Строк подання студентом проекту (роботи) 10 червня 2026 р.

3. Вихідні дані до проекту (роботи) Результати і матеріали отримані під час проходження переду-дипломної практики

4. Зміст розрахунково - пояснювальної записки(перелік питань, які потрібно розробити)

1. Аналіз вимог до програмного забезпечення

2. Аналіз вимог і постановка задачі

3. Проектування системи

4. Реалізація проекту

5. Тестування та впровадження

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень)

1 Приклад діаграми CNN для класифікації зображень (рис. 2.1, ст. 26)

2 3D-куб і 2D-многовиміри швів (рис. 2.8, ст. 38)

4. Огляд відео з кількома режимами перегляду (рис. 2.11, ст. 43)

4 Інструмент м'які ножиці (рис. 3.14, ст. 53)

5 Text2animation для вибору кадру (рис. 2.14, ст. 62)

6. Консультанти розділів проекту (роботи)

Розділ	Консультант	Підпис, дата	
		Завдання видав	Завдання прийняв

2. Дата видачі завдання 2026 р.

Керівник _____
(підпис)

Завдання прийняв до виконання _____
(підпис)

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів дипломного проекту (роботи)	Строк виконання етапів проекту	Примітка
1	Визначення та обґрунтування теми роботи	17.01.2026	виконано
2	Огляд існуючих концепцій, рішень та сервісів в даній області	21.02.2026	виконано
3	Побудова моделі або алгоритму власного рішення	01.03.2026	виконано
4	Документування реалізації власного оригінального рішення вибраними засобами	21.04.2026	виконано
5	Оформлення пояснювальної записки бакалаврської роботи	02.05.2026	виконано

Студент _____ Тріш Т.В.
(підпис)

Керівник роботи _____ Процюк Г.Я.
(підпис)

АНОТАЦІЯ

Бакалаврська робота містить 91 сторінок, 24 рисунки, 2 таблиці, список використаних джерел із 40 найменувань.

Метою роботи є дослідження методів і технологій застосування штучного інтелекту для автоматичного монтажу та редагування відеоконтенту.

Об'єкт дослідження: процеси автоматизованого монтажу та редагування відеоконтенту із використанням технологій штучного інтелекту.

Предмет дослідження: методи машинного навчання, нейронні мережі, алгоритми комп'ютерного бачення та програмні засоби автоматичного аналізу, генерації та редагування відеоматеріалів.

Результати дослідження: проведено аналіз сучасних технологій штучного інтелекту у сфері відеомонтажу, досліджено методи автоматичного редагування відео, а також реалізовано підхід до автоматизації процесів.

У першому розділі розглянуто теоретичні основи застосування штучного інтелекту у відеомонтажі та редагуванні контенту, а також позитивний і негативний вплив ШІ на контент-індустрію.

У другому розділі досліджено методи та інструменти інтелектуального монтажу відео, зокрема згорткові нейронні мережі, технології комп'ютерного бачення, методи автоматичного редагування та набори даних.

У третьому розділі розглянуто реалізацію системи автоматичного монтажу відео, а також проведено оцінку ефективності роботи системи й аналіз перспектив розвитку технологій штучного інтелекту у сфері.

Висновок: використання технологій штучного інтелекту для автоматичного монтажу та редагування відеоконтенту дозволяє підвищити швидкість і якість обробки відеоматеріалів, автоматизувати рутинні процеси, оптимізувати витрати часу та покращити ефективність.

КЛЮЧОВІ СЛОВА: ШТУЧНИЙ ІНТЕЛЕКТ, ВІДЕОМОНТАЖ, НЕЙРОННІ МЕРЕЖІ, АВТОМАТИЧНЕ РЕДАГУВАННЯ ВІДЕО, МАШИННЕ НАВЧАННЯ, ВІДЕОКОНТЕНТ, ГЕНЕРАЦІЯ ВІДЕО, DEEP LEARNING.

ANNOTATION

Bachelor's thesis of 91 pages, 24 figures, 2 tables, and a reference list containing 40 sources.

The aim of the work is to study methods and technologies of applying artificial intelligence for automatic video editing and montage.

Research object: processes of automated video editing and montage using artificial intelligence technologies.

Research subject: machine learning methods, neural networks, computer vision algorithms, and software tools for automatic analysis, generation, and editing of video materials.

Research results: an analysis of modern artificial intelligence technologies in video editing was conducted, methods of automatic video processing were investigated and an approach to intelligent automation of video production processes.

The first section describes the theoretical foundations of applying artificial intelligence in video editing and content production, the principles of intelligent systems, as well as the positive and negative impact of AI on the content industry.

The second section analyzes methods and tools for intelligent video montage, including convolutional neural networks, computer vision technologies, automatic editing methods, and datasets for training AI models.

The third section covers the implementation of an automatic video editing system, well as the evaluation of system performance and analysis of future prospects for artificial intelligence technologies in video production.

Conclusion: the use of artificial intelligence technologies for automatic video editing and montage improves the speed and quality of video processing, automates routine operations, optimizes time consumption, and increases the efficiency of multimedia content creation.

KEY WORDS: ARTIFICIAL INTELLIGENCE, VIDEO EDITING, NEURAL NETWORKS, AUTOMATIC VIDEO EDITING, MACHINE LEARNING, VIDEO CONTENT, VIDEO GENERATION, DEEP LEARNING

ЗМІСТ

ВСТУП	8
РОЗДІЛ 1. ТЕОРЕТИЧНІ ОСНОВИ ЗАСТОСУВАННЯ ШТУЧНОГО ІНТЕЛЕКТУ У ВІДЕОМОНТАЖІ ТА РЕДАГУВАННІ КОНТЕНТУ ...	11
1.1 Основні поняття та принципи роботи систем штучного інтелекту у відео обробці	11
1.2 Вплив штучного інтелекту на автоматизацію створення та редагування відео контенту	16
1.3 Проблеми, ризики та етичні аспекти використання штучного інтелекту у контент-індустрії	21
1.4 Висновки по розділу	25
РОЗДІЛ 2. МЕТОДИ ТА ІНСТРУМЕНТИ ІНТЕЛЕКТУАЛЬНОГО МОНТАЖУ ВІДЕО	26
2.1 Методи машинного навчання та згорткові нейронні мережі для аналізу відеоданих	26
2.2 Інструменти автоматичного редагування, генерації та маніпуляції відео контентом	35
2.3 Набори даних, метрики оцінювання та методи тестування якості відео монтажу	49
2.4 Висновки по розділу	54
РОЗДІЛ 3. РЕАЛІЗАЦІЯ СИСТЕМИ АВТОМАТИЧНОГО МОНТАЖУ ТА ОЦІНКА ЇЇ ЕФЕКТИВНОСТІ	55
3.1 Інтелектуальна автоматизація відеовиробництва та сценарії використання системи	55
3.2 Реалізація модулів віртуального та автоматизованого відео монтажу	65
3.3 Аналіз ефективності роботи системи та перспективи розвитку технологій ШІ у відеопродакшені	76
3.4 Висновки по розділу	89
ВИСНОВКИ	90
СПИСОК ПОСИЛАНЬ НА ДЖЕРЕЛА	92

					БР.ІІ –85.00.00.000 ІІЗ			
Змн.	Арк.	№ докум.	Підпис	Дата	Застосування штучного інтелекту для автоматичного монтажу та редагування відео-контенту	Літ.	Арк.	Аркушів
Розроб.		Тріш Т.В.						
Перевір.		Процюк Г.Я.					7	
Реценз.		Ваврик Т.О.				ІФНТУНГ ІІ-22-3		
Н. Контр.		Піх М.М.						
Затверд.		Бандура В. В.			Пояснювальна записка			

ВСТУП

У сучасному цифровому середовищі відеоконтент став одним із найпоширеніших способів передачі інформації, комунікації та створення мультимедійного продукту. Стрімкий розвиток соціальних мереж, відеохостингів, потокових платформ і цифрових медіа значно підвищив попит на якісний відеоконтент, що потребує швидкого та ефективного монтажу, редагування й обробки. Традиційні методи відеомонтажу вимагають значних часових і людських ресурсів, а також високого рівня професійної підготовки. У зв'язку з цим особливої актуальності набуває застосування технологій штучного інтелекту для автоматизації процесів створення, аналізу та редагування відеоматеріалів.

Актуальність теми зумовлена необхідністю підвищення ефективності процесів відеовиробництва, скорочення часу обробки мультимедійного контенту та автоматизації рутинних операцій монтажу. Сучасні системи штучного інтелекту дозволяють автоматично аналізувати сцени, визначати ключові фрагменти відео, виконувати синхронізацію аудіо та відео, застосовувати ефекти, здійснювати розпізнавання об'єктів і створювати готовий відеоконтент із мінімальним втручанням користувача. Використання методів машинного навчання, комп'ютерного бачення та нейронних мереж відкриває нові можливості для створення інтелектуальних систем автоматичного монтажу, здатних значно оптимізувати процеси відеопродакшену.

Метою роботи є дослідження та аналіз методів застосування штучного інтелекту для автоматичного монтажу та редагування відеоконтенту, а також розробка підходів до створення інтелектуальної системи автоматизації відеовиробництва.

Завданнями дослідження є аналіз предметної області та сучасних рішень у сфері автоматичного відеомонтажу, дослідження методів машинного навчання та комп'ютерного бачення для обробки відеоданих, вивчення згорткових нейронних мереж і алгоритмів автоматичного редагування, аналіз наборів даних та метрик оцінювання якості відеоконтенту, проектування структури

					БР.ІІІ - 85.00.00.000 ПЗ	Арк.
						9
Змн.	Арк.	№ докум.	Підпис	Дата		

інтелектуальної системи відеомонтажу, реалізація основних функціональних модулів, тестування системи та оцінка її ефективності.

Об'єктом дослідження є процеси автоматичного монтажу та редагування відеоконтенту із використанням технологій штучного інтелекту.

Предметом дослідження є методи, алгоритми та програмні засоби автоматизованої обробки відео, зокрема нейронні мережі, методи комп'ютерного бачення, інструменти машинного навчання та технології інтелектуального аналізу мультимедійного контенту.

Методи дослідження включають аналіз наукових джерел і сучасних технологій у сфері штучного інтелекту, порівняльний аналіз алгоритмів автоматичного відеомонтажу, моделювання процесів обробки відеоданих, експериментальні методи тестування програмних компонентів, а також оцінювання ефективності інтелектуальних систем за допомогою відповідних метрик якості.

Розроблена система автоматичного монтажу відеоконтенту передбачатиме можливість аналізу відеофрагментів, автоматичного визначення ключових сцен, синхронізації аудіо та відео, застосування ефектів і створення готового мультимедійного продукту. Особлива увага приділятиметься точності роботи алгоритмів, швидкості обробки даних, масштабованості системи та зручності використання. Очікується, що впровадження технологій штучного інтелекту у сферу відеомонтажу дозволить підвищити продуктивність створення контенту, зменшити витрати часу на редагування та забезпечити новий рівень автоматизації мультимедійного виробництва.

Бакалаврська робота містить 91 сторінок, 24 рисунки, 2 таблиці, 3 розділи, список використаних джерел із 40 найменувань.

					БР.ІІІ - 85.00.00.000 ПЗ	Арк.
						10
Змн.	Арк.	№ докум.	Підпис	Дата		

РОЗДІЛ 1. ТЕОРЕТИЧНІ ОСНОВИ ЗАСТОСУВАННЯ ШТУЧНОГО ІНТЕЛЕКТУ У ВІДЕОМОНТАЖІ ТА РЕДАГУВАННІ КОНТЕНТУ

1.1 Основні поняття та принципи роботи систем штучного інтелекту у відеообробці

Штучний інтелект (ШІ) став невід'ємною частиною сучасної технологічної ери, впливаючи на різні сфери, включаючи виробництво та монтаж аудіовізуального контенту. Використання ШІ в цьому контексті дозволяє створювати нові художні та розважальні продукти, одночасно спрощуючи рутинні та трудомісткі процеси. Наприклад, алгоритми ШІ можуть генерувати музику або анімацію на основі наданих даних, значно розширюючи можливості сучасного контенту та пропонуючи нові перспективи для креативної індустрії. Дослідження у сфері генерації музики для танцювальних проектів демонструє здатність ШІ створювати складні мультимедійні проекти, автоматично вибираючи музику для різних інструментів на основі танцювальних рухів. Це відкриває нові горизонти для мультимедійного мистецтва, зокрема в кіно та музиці.

Штучний інтелект також сприяє розвитку мультимодального розпізнавання емоцій, яке може обробляти аудіовізуальну інформацію та виявляти емоційні стани на основі динамічних реакцій [1]. Це покращує користувацький досвід, особливо в аудіовізуальному контенті, де реакції глядача чи слухача відіграють вирішальну роль у створенні глибшого емоційного впливу. Крім того, дослідження показують, що ШІ впливає на процеси розробки контенту та захисту інтелектуальної власності. Наприклад, досліджують вплив технологій ШІ на маркетингові стратегії та цифрові платформи, зосереджуючись на тому, як створення контенту на основі ШІ кидає виклик традиційним системам інтелектуальної власності. Хоча їхнє дослідження визнає важливість ШІ в креативному секторі, воно підкреслює, що можливості ШІ у виробництві та розповсюдженні контенту вимагають перегляду законів про інтелектуальну власність, щоб врахувати нові форми цифрових медіа. Дослідження підкреслює зростаючу потребу в

					БР.ІП - 85.00.00.000 ПЗ	Арк.
						11
Змн.	Арк.	№ докум.	Підпис	Дата		

комплексному регулюванні в таких сферах, як авторське право та ліцензування в контексті розповсюдження контенту на основі ІІІ.

У роботі обговорюється, як штучний інтелект використовується для спрощення певних галузей кінематографії, таких як створення реалістичних візуальних ефектів та анімації.

Протягом останнього десятиліття штучний інтелект став ключовим агентом трансформації в багатьох медіаіндустріях. В аудіовізуальному виробництві інструменти на основі штучного інтелекту швидко еволюціонували від периферійних удосконалень до центральних елементів творчого робочого процесу [2]. Цей зсув очевидний у сфері відеомонтажу, де програмне забезпечення на основі штучного інтелекту змінює не лише те, як збирається контент, але й те, як сама кінематографічна мова задумується та практикується. Однак, оскільки сучасні платформи монтажу інтегрують алгоритми машинного навчання, здатні до автоматичного виявлення сцен, монтажу на основі ритму та емоційно керованого секвенування, колись ручний процес монтажу стає все більш автоматизованим. Ці події піднімають фундаментальні питання щодо креативності та авторства, прийняття рішень та естетичної діяльності в епоху комп'ютерного кіно.

Маючи досвід роботи режисера та монтажера на різних етапах монтажу - від телесеріалів до інституційних, експериментальних та авторських художніх і документальних фільмів - а також навчання та дослідження в цій галузі, представлена тут робота поєднує практичний досвід та теоретичні питання. Спираючись на фундаментальних теоретиків монтажу, а також сучасних авторів, аналіз розглядає еволюцію монтажу поряд з еволюцією медіатехнологій. Особлива увага приділяється інтерфейсам та можливостям сучасних платформ монтажу на основі штучного інтелекту, які не лише підтримують, але й формують рішення користувачів. Методологія, що використовується в цьому розділі, буде зосереджена на контент-аналізі оглядів інструментів монтажу на основі штучного інтелекту на професійних каналах YouTube, використовуючи вибірку професійних та напівпрофесійних каналів YouTube, зосереджених на програмному забезпеченні для відеомонтажу та тенденціях.

					БР.ІІІ - 85.00.00.000 ПЗ	Арк.
						12
Змн.	Арк.	№ докум.	Підпис	Дата		

Зрештою, цей підрозділ розглядає монтаж як частину ширшої трансформації медіаграматики в цифрову епоху [3]. Він також проблематизує монтаж як мінливу, узгоджену практику – таку, в якій людські та машинні внески стикаються, сходяться та співтворюються. Результатом є не лише технічний зсув, але й культурний та концептуальний: автоматизація монтажу може одночасно обмежувати та розширювати можливості кінематографічного виробництва у XXI столітті.

Інтелектуальна кінематографія (ІК) використовує штучний інтелект (ШІ) для допомоги у виконанні завдань, пов'язаних з камерою, протягом усього робочого процесу медіавиробництва. Для передпродакшну це стосується планування зйомок, наприклад, за допомогою сценарію або введення даних користувачем.

Інтелектуальна кінематографія впливає на світ кіно, ігор, телевізійного мовлення та потокового передавання. Симбіотичний зв'язок між дослідженнями та відеовиробництвом історично був важливим для розвитку відеорозваг. Однак, з розвитком цього зв'язку, типи контенту та стилі виробництва все більше спиралися на різноманітний пул досліджень у багатьох галузях. Наприклад, є роботи, які автоматизують процеси ручного захоплення, а також роботи, які розуміють та покращують занурення аудиторії. Таке різноманіття та плинність діяльності ускладнюють встановлення єдиного визначення та опису ландшафту досліджень ІК. Отже, існує очевидна відсутність та потреба як у загальних оглядових статтях, так і в більш спеціалізованих статтях. У цій роботі ми намагаємося подолати ці проблеми, визначивши та представивши структурований огляд стану та майбутнього досліджень ІК [4]. Однак наше визначення ІК зосереджено саме на отриманні реального контенту. Ми виключаємо дослідження, спрямовані на допомогу в постпродакшні або автоматизації створення контенту (наприклад, з використанням моделей дифузії) з двох основних причин. По-перше, дослідження штучного інтелекту, пов'язані з постпродакшном, вже добре підтверджені оглядовими статтями. По-друге, штучний інтелект у постпродакшні зазвичай має справу з віртуальними камерами, об'єктами та світами, які зазвичай не мають значення для отримання реального контенту.

					БР.ІІІ - 85.00.00.000 ПЗ	Арк.
						13
Змн.	Арк.	№ докум.	Підпис	Дата		

У цій роботі класифікуються за *виробничим середовищем*. Подібно до визначення художнього середовища, тут дослідження групуються за відповідними інструментами та практиками, що використовуються в різних типах виробництва, а не за метою. Наприклад, прямі трансляції спортивних подій та подій мають багато спільних кінематографічних практик, тому вони згруповані під одним заголовком «*Виробництво в прямому ефірі*». Це дозволяє нам: (1) додатково розділити роботи на підкатегорії відповідно до пов'язаних галузей досліджень штучного інтелекту та забезпечити цілісне уявлення про ландшафт інтелектуальної комунікації, зберігаючи при цьому актуальність для кожного виробничого середовища; та (2) включити теми досліджень, які, хоча й не спрямовані на креативне відеовиробництво, мають до нього безпосереднє відношення. Теми, що розглядаються для кожного середовища, викладені нижче та візуалізовані на рисунку 1.1.

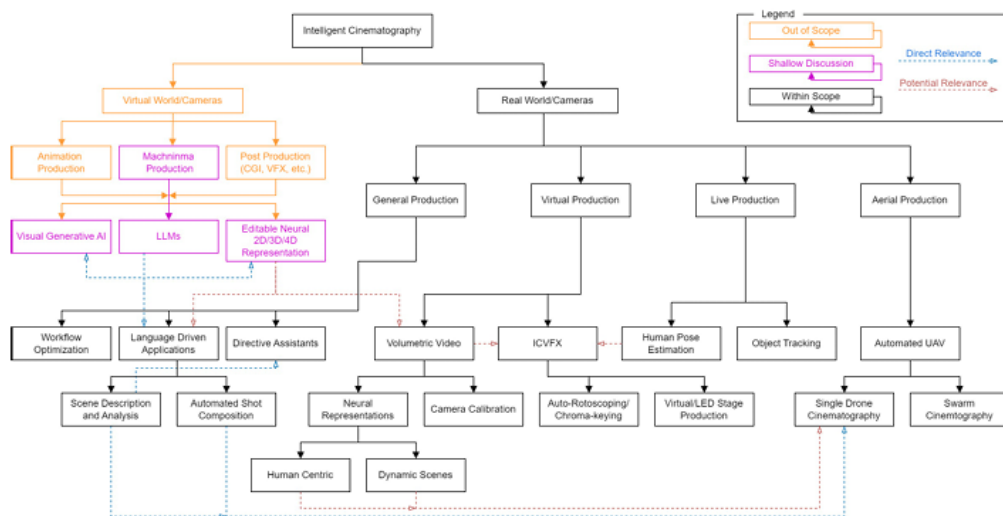


Рисунок 1.1 - Класифікація досліджень у галузі інформаційних технологій, що охоплює роботи як у рамках

Такий підхід мінімізує потребу в трудомісткому ручному редагуванні в цих сферах, що дозволяє ефективніше здійснювати виробничі процеси в анімаційному секторі. За даними, штучний інтелект широко застосовується в креативних індустріях, особливо у створенні та редагуванні медіаконтенту. Крім

того, досліджують інтеграцію штучного інтелекту в освіту, зосереджуючись на тому, як штучний інтелект підтримує адаптивність та вдосконалює платформи дистанційного навчання. Дослідження висвітлює еволюцію «розумної освіти» та створення інтерактивних віртуальних середовищ, що сприяють новим способам залучення студентів до академічних досліджень. Примітно, що віртуальні простори також розвиваються завдяки штучному інтелекту, сприяючи інтерактивним дослідницьким та навчальним практикам [5]. Це дозволяє навчальним закладам створювати віртуальні музеї та інтерактивні простори, що підвищує залучення студентів до академічних досліджень.

Ще одним важливим застосуванням штучного інтелекту є сприяння перекладу аудіовізуального контенту, адаптуючи його до нових культурних контекстів за допомогою машинного перекладу. Це значно спрощує розповсюдження медіаконтенту серед різномовної аудиторії, особливо у глобальних медіакорпораціях. За даними, штучний інтелект сприяє розвитку інноваційних освітніх технологій, зокрема, забезпечуючи більш персоналізований навчальний досвід, який можна адаптувати до індивідуальних потреб, з акцентом на творчі дисципліни. З огляду на швидкий розвиток штучного інтелекту та його вплив на створення креативного контенту, виникає низка етичних та правових проблем. Дослідження зміни, які генеративний штучний інтелект привносить у художню освіту, виступаючи за нові методи навчання та інтеграцію штучного інтелекту в навчальні програми для підготовки студентів до розвитку художніх практик в епоху штучного інтелекту. Тим дослідження використання алгоритмів нечіткої логіки у створенні музики, зосереджуючись на тому, як технології штучного інтелекту дозволяють створювати нові музичні композиції за допомогою алгоритмічних процесів [6]. Досліджується обмеження моделей штучного інтелекту, які можна пояснити, зосереджуючись на етичних проблемах, що виникають через відсутність прозорості штучного інтелекту в процесах прийняття рішень та складність розуміння обґрунтування результатів, згенерованих машиною.

Таким чином, це дослідження має на меті дослідити, як штучний інтелект трансформує процеси створення та редагування аудіовізуального контенту, а

					БР.ІІІ - 85.00.00.000 ПЗ	Арк.
						15
Змн.	Арк.	№ докум.	Підпис	Дата		

також визначити ключові тенденції, виклики та етичні міркування, з якими стикається галузь на тлі її швидкого розвитку. Аналіз охоплює різні застосування штучного інтелекту в креативних сферах, включаючи створення музики, анімацію та переклад контенту, щоб окреслити основні перспективи, обмеження та етичні проблеми, пов'язані з цією технологією.

1.2 Вплив штучного інтелекту на автоматизацію створення та редагування відеоконтенту

Використання штучного інтелекту (ШІ) у створенні та редагуванні аудіо-візуального контенту стає дедалі популярнішим завдяки його величезному потенціалу для оптимізації робочих процесів та зниження виробничих витрат. Інструменти на базі ШІ пропонують інноваційні підходи до автоматизації та вдосконалення всіх етапів виробництва контенту – від планування до редагування. У цьому контексті ШІ слугує універсальним інструментом, що підвищує ефективність та спрощує виконання проєктів різної складності.

Одним із ключових досягнень штучного інтелекту є його здатність оптимізувати процес відеомонтажу. Традиційно відеомонтаж передбачає тривалу ручну роботу, включаючи аналіз записаного матеріалу, вибір найкращих моментів, налаштування корекції кольору та синхронізацію аудіо з візуальними елементами. Сучасні алгоритми штучного інтелекту не тільки виконують ці завдання швидше, але й з більшою точністю, пропонуючи автоматичну ідентифікацію ключових моментів у відео. Наприклад, програмне забезпечення на базі штучного інтелекту, таке як Adobe Premiere Pro з функцією Sensei, може автоматично розпізнавати важливі сцени та стабілізувати зображення [7]. Це не лише скорочує час, витрачений на технічні операції, але й дозволяє творцям зосередитися на художніх рішеннях. Порівняння традиційних підходів до відеомонтажу та методів за допомогою штучного інтелекту наведено в таблиці 1.1.

					БР.ІП - 85.00.00.000 ПЗ	Арк.
						16
Змн.	Арк.	№ докум.	Підпис	Дата		

Таблиця 1.1

Порівняльний аналіз відеовиробництва за допомогою штучного інтелекту та традиційного відеовиробництва.

Характеристика	Традиційне відеовиробництво	AI-відеовиробництво
Час	Від тижнів до місяців	Від годин до днів
Вартість	Висока	Низька
Творчий контроль	Непослідовний	Зберігає контекст
Масштабованість	Обмежена	Без зусиль
Налаштування (Customization)	Ручне	Автоматизоване

Наведені дані ілюструють значні переваги, які пропонує штучний інтелект у відеовиробництві. У той час як традиційний відеомонтаж вимагає значного часу та ресурсів, штучний інтелект дозволяє виконувати подібні завдання за значно коротші періоди, іноді скорочуючи тижні роботи до кількох годин. Наприклад, алгоритми обробки відео, такі як DaVinci Resolve, з інтегрованим модулем штучного інтелекту дозволяють автоматично коригувати колір, аналізувати сцени та налаштовувати тональність кадру до єдиного стандарту [8]. Порівняно з ручним налаштуванням, яке вимагає ретельного налаштування для кожного кадру, така автоматизація зменшує навантаження редактора на 50-70%.

Ці аспекти підкреслюють важливість інтелектуальних алгоритмів не лише для технічних удосконалень, але й для впровадження інновацій, що спрощують складні завдання, такі як створення реалістичних візуальних ефектів або анімації. Це особливо очевидно в технологіях, пов'язаних з глобальними мережами (GAN), які дозволяють створювати дуже реалістичні ефекти без необхідності використання складного обладнання або тривалих обчислень [9]. Наприклад, GAN можуть автоматично генерувати складні 3D-об'єкти, текстури або навіть природні явища, такі як дощ, дим чи вибухи, з високим ступенем деталізації. Це має значний вплив на розвиток кіно- та телевізійного виробництва, полегшуючи роботу над сценами, що включають складні візуальні ефекти. Крім того, інструменти на основі штучного інтелекту дозволяють моделювати обличчя, додавати анімацію або змінювати емоції персонажів у режимі реального часу. Віртуальні актори, створені за допомогою таких технологій, виглядають цілком

					БР.ІІІ - 85.00.00.000 ПЗ	Арк.
						17
Змн.	Арк.	№ докум.	Підпис	Дата		

реалістично, зберігаючи природні вирази обличчя, рухи та навіть голоси. Наприклад, у кіно ці технології дозволяють відтворювати історичних постатей або розробляти абсолютно нових персонажів для науково-фантастичних фільмів, зменшуючи витрати, пов'язані зі складним гримом або спец ефектами [10]. Сучасні анімаційні студії використовують ці інструменти для створення реалістичних персонажів або навіть повністю комп'ютерно-згенерованих сцен, які майже не відрізняються від реальних кадрів.

Алгоритми глибокого навчання також відіграють вирішальну роль у відновленні та покращенні аудіовізуальних матеріалів. Зокрема, технологія надвисокої роздільної здатності дозволяє конвертувати відео низької роздільної здатності у формати HD або навіть 4K. Штучний інтелект автоматично налаштовує кольори, заповнює відсутні деталі, зменшує цифровий шум і покращує загальну чіткість зображення. Ці можливості широко використовуються як для відновлення старих фільмів, так і для підготовки сучасного контенту, знятого в складних умовах. Наприклад, технології, розроблені NVIDIA, можуть не тільки відновлювати втрачені фрагменти відео, але й створювати нові, реалістичні версії на основі архівних матеріалів, що може революціонізувати збереження кінематографічної спадщини. Ці алгоритми також можуть автоматично видаляти небажаний шум, регулювати рівні гучності або синхронізувати аудіо з відеоматеріалами [11]. Це особливо важливо під час роботи з матеріалами, що містять багатошарові аудіодоріжки, такими як фільми або реклама. Завдяки цим інструментам зменшується потреба в тривалому ручному редагуванні звуку, а результати часто перевершують ті, що досягаються за допомогою ручної обробки. Наприклад, сучасні алгоритми обробки звуку можуть автоматично усувати ефекти луни або фоновий шум вітру, що раніше вимагало залучення спеціаліста.

Для створення трейлерів або рекламних відеороликів штучний інтелект може швидко аналізувати відзнятий матеріал, щоб визначити найдинамічніші сцени або моменти, які найкраще передають суть фільму. Алгоритми можуть автоматично виділяти ключові елементи, такі як динаміка рухів, взаємодія персонажів та важливі сюжетні лінії. Це призводить до створення високоефективного

					БР.ІП - 85.00.00.000 ПЗ	Арк.
						18
Змн.	Арк.	№ докум.	Підпис	Дата		

контенту, який зберігає всю важливу інформацію для аудиторії без зайвих елементів. Ще одним значним досягненням штучного інтелекту є інтеграція технологій комп'ютерного зору, які спрощують такі завдання, як кадрування, розпізнавання облич та виявлення об'єктів. Наприклад, під час процесу монтажу такі ефекти, як розмиття фону або виділення центрального об'єкта, можуть застосовуватися автоматично [12]. Ці інструменти дозволяють швидко досягати кінематографічної якості навіть без спеціалізованих навичок відеомонтажу. Більше того, автоматизація рутинних завдань звільняє більше можливостей для творчості. Монтажери та режисери можуть зосередитися на концептуальних та художніх рішеннях, довіряючи технічні аспекти алгоритмам на основі штучного інтелекту. Наприклад, генерація попередніх версій фільмів або кліпів за допомогою штучного інтелекту з можливістю швидкої їх зміни дозволяє творцям експериментувати з різними варіаціями сценарію, композиціями кадрів або колірними палітрами.

Окрім технічних застосувань, ШІ пропонує нові творчі можливості. Наприклад, технології перетворення тексту на зображення на основі ШІ дозволяють створювати розкадровки або сценарії на основі початкових ідей клієнтів [13]. Це спрощує ранні етапи розробки проекту, дозволяючи швидше експериментувати з різними концепціями. Крім того, інтегровані ШІ алгоритми прогнозування тенденцій аналізують уподобання аудиторії та рекомендують стилістичні або наративні корективи для підвищення популярності контенту. Завдяки цим технологіям аудіовізуальний контент може не лише відповідати поточним потребам, але й адаптуватися до майбутніх очікувань, що робить ШІ незамінним інструментом у цій галузі. Крім того, ШІ здатний самостійно генерувати сценарії на основі зібраних даних. Використовуючи інструменти, що обробляють великі обсяги інформації, ШІ може аналізувати інтереси аудиторії через соціальні мережі, пошукові запити та інші джерела. Це дозволяє виявляти трендові теми та жанри, а також прогнозувати майбутні тенденції. Наприклад, платформи, що використовують технології ШІ, можуть автоматично генерувати сценарії, що відображають теми, що найбільше резонують із сучасною аудиторією.

					БР.ІІ - 85.00.00.000 ПЗ	Арк.
						19
Змн.	Арк.	№ докум.	Підпис	Дата		

Ще одна сфера, де ШІ демонструє свою ефективність, - це адаптація контенту для різних платформ. Його алгоритми можуть автоматично налаштовувати відеоформати відповідно до вимог платформи, таких як вертикальний формат 9:16 для TikTok або квадратний формат для Instagram. Крім того, штучний інтелект може автоматично обрізати тривалість відео, зберігаючи його основне повідомлення, та оптимізувати відео для різних пристроїв, знижуючи роздільну здатність для малопотужного обладнання. Редактори, інтегровані зі штучним інтелектом, такі як Magisto або Vimeo Create, також дозволяють автоматично додавати субтитри кількома мовами, покращуючи доступність контенту для міжнародної аудиторії – особливо цінна функція. Штучний інтелект Sensei в Adobe Premiere Pro оптимізує відеомонтаж, виявляючи ключові моменти в невідредагованому фільмі, значно скорочуючи тривалість монтажу та дозволяючи редакторам зосередитися на творчому виборі. Дослідження підтверджує функцію ШІ в автоматизації повторюваних операцій, таких як ідентифікація сцени та синхронізація звуку, тим самим оптимізуючи робочі процеси та зменшуючи виробничі витрати. DaVinci Resolve використовує ШІ для оптимізації кольорокорекції шляхом автономного аналізу фільму та зміни колірних тонів, тим самим покращуючи візуальну якість та пришвидшуючи постпродакшн [14]. Дослідження, підкреслюють ефективність корекції кольору на основі ШІ у покращенні візуальної узгодженості, сприяючи досягненню кращих результатів з меншими ручними зусиллями.

GAN революціонізували візуальні ефекти, створюючи реалістичні 3D-моделі, атмосферні явища та складні симуляції, такі як вибухи. Алгоритми глибокого навчання NVIDIA використовують підходи з надвисокою роздільною здатністю для покращення та масштабування застарілих аудіовізуальних джерел до форматів високої чіткості. Ці технології дозволяють невеликим компаніям створювати високоякісну анімацію та ефекти за зниженими цінами. GAN сприяють створенню складніших сценаріїв та інноваційних творчих можливостей з меншими ресурсами, тим самим розширюючи можливості для експериментів. Штучний інтелект все частіше використовується для налаштування аудіовізуальних

					БР.ІІІ - 85.00.00.000 ПЗ	Арк.
						20
Змн.	Арк.	№ докум.	Підпис	Дата		

матеріалів за допомогою обробки природної мови та машинного навчання [15]. Штучний інтелект автоматизує створення сценаріїв, адаптивну розробку наративу та пропозиції контенту на основі аналітики аудиторії, підвищуючи залученість глядачів. Підкреслюють роль штучного інтелекту у створенні інтерактивних, персоналізованих наративів, що підвищує релевантність контенту. Крім того, прогнозна аналітика дозволяє авторам адаптувати матеріал відповідно до поведінки аудиторії, тим самим покращуючи утримання та задоволення.

ШІ автоматизує написання сценаріїв та візуальні ефекти, сприяючи інтерактивним наративам, які скорочують час виробництва. У маркетингу фільми, створені ШІ, адаптуються до певних демографічних груп, тим самим підвищуючи ефективність кампанії. Стверджують, що ШІ є доповнюючим наративом, автоматизуючи виробництво та розширюючи залучення аудиторії за допомогою персоналізованих пропозицій, а також допомагаючи постачальникам навчального контенту, адаптуючи фільми до потреб учнів та покращуючи доступність.

Ці можливості ілюструють, як штучний інтелект трансформує традиційні процеси створення аудіовізуального контенту, поєднуючи автоматизацію з креативністю та відкриваючи нові можливості для галузі. Завдяки інтеграції інтелектуальних алгоритмів створення, редагування та оптимізація контенту стають не лише ефективнішими, але й більш адаптивними, дозволяючи продюсерам адаптувати свою роботу до вимог сучасного ринку. Штучний інтелект не просто прискорює рутинні завдання, він також сприяє інноваціям, розширюючи межі творчих можливостей.

1.3 Проблеми, ризики та етичні аспекти використання штучного інтелекту у контент-індустрії

Незважаючи на численні переваги, такі як автоматизація рутинних завдань, підвищення ефективності та зниження витрат, існують також значні проблеми, пов'язані з творчим процесом роботи з контентом. Хоча ШІ може значно спростити виробничі процеси, він також може призвести до зниження рівня

					БР.ІП - 85.00.00.000 ПЗ	Арк.
						21
Змн.	Арк.	№ докум.	Підпис	Дата		

унікальності.

Одним із важливих аспектів є зміна підходів до контролю якості аудіовізуального контенту. ІІІ здатний швидко аналізувати великі обсяги матеріалу та визначати, які елементи потребують виправлення, проте автоматизовані системи не завжди враховують усі контекстуальні нюанси [16]. Це особливо стосується емоційної складової контенту, глибини комунікації з аудиторією та соціальної значущості. Замість створення контенту, який відповідає індивідуальним потребам різних цільових аудиторій, ІІІ може створювати більш стандартизований, безособовий матеріал, який не враховує глибші аспекти сприйняття та емоційної взаємодії з глядачем. Це може призвести до стандартизації контенту, який втратить унікальність та індивідуальність, властиві традиційним методам створення.

Алгоритми штучного інтелекту, попри свою ефективність, часто залежать від попередньо встановлених шаблонів або патернів, що може обмежувати креативність та оригінальність згенерованого матеріалу [17]. У контексті музики, згенерованої штучним інтелектом, алгоритми часто створюють треки, використовуючи заздалегідь визначені формули або набори даних, що призводить до творів, позбавлених емоційної глибини та особистого підходу, які може надати композитор-людина. У відеовиробництві штучний інтелект може стандартизувати такі процедури, як кольорокорекція та відеомонтаж, що призводить до узгодженої естетики в багатьох проектах. Це підвищує ефективність, але також ставить під загрозу унікальність кожного твору, ускладнюючи здатність аудиторії розрізняти підходи багатьох авторів.

Крім того, матеріали, створені штучним інтелектом, можуть підривати культурне розмаїття. Алгоритми можуть недостатньо враховувати регіональні культурні відмінності, які є вирішальними для залучення різної аудиторії. Наприклад, у міжнародних медіа-організаціях методи штучного інтелекту, що використовуються для перекладу контенту або субтитрування, можуть нехтувати культурними ідіомами чи емоційними нюансами, які зазвичай передав би людський перекладач. Як наслідок, інформація може ставати дедалі гомогенізованішою, що призводить до меншої кількості можливостей для появи місцевих виразів або

					БР.ІІІ - 85.00.00.000 ПЗ	Арк.
						22
Змн.	Арк.	№ докум.	Підпис	Дата		

різних культурних точок зору. Отже, хоча штучний інтелект пропонує нові перспективи для персоналізації контенту, він також породжує гендерні побоювання щодо того, що креативність дедалі більше обмежується алгоритмічними обмеженнями, що призводить до появи матеріалу, який є менш унікальним та більш універсально стандартизованим.

Крім того, використання технологій штучного інтелекту може призвести до зниження якості контенту, що створюється, особливо коли використовуються автоматизовані системи редагування. Наприклад, алгоритми можуть мати труднощі з розпізнаванням культурних контекстів або тонких нюансів, що призводить до втрати точності та надійності. Алгоритми штучного інтелекту можуть генерувати контент, який виглядає переконливим, але не завжди відповідає фактичним фактам або авторитетним джерелам, що ставить під сумнів достовірність таких матеріалів. Оскільки автоматизовані системи можуть базуватися на великих наборах даних, але не завжди правильно інтерпретують їхній зміст, це може призвести до неправильних тлумачень або неправдивих заяв. Це створює загрозу, коли штучний інтелект активно використовується для створення новин, документальних фільмів або інших важливих форм контенту, де точність і правдивість інформації є критично важливими. Під час роботи зі штучним інтелектом часто виникають труднощі у досягненні бажаного рівня реалізму, особливо у складних сценах, де потрібно враховувати фізичні характеристики об'єктів, такі як їхня взаємодія з навколишнім середовищем, рух або гравітація [18]. Програми на основі штучного інтелекту можуть генерувати дуже хороші результати в простих ситуаціях, але під час роботи зі складними композиціями можуть виникати непередбачувані результати, що погіршує якість контенту та робить його менш природним або реалістичним [19]. У цьому контексті результати дослідження демонструють думки експертів щодо контенту, який був згенерований за допомогою генеративних алгоритмів.

Хоча ШІ може значно пришвидшити процес створення контенту, якість цього контенту все ще не може конкурувати з матеріалами, створеними людиною. Автоматизовані системи часто створюють контент, який є більш

					БР.ІП - 85.00.00.000 ПЗ	Арк.
						23
Змн.	Арк.	№ докум.	Підпис	Дата		

стандартизованим і менш персоналізованим, що знижує його привабливість для цільової аудиторії. Однак, завдяки швидкості та ефективності обробки великих обсягів даних, ШІ може бути корисним інструментом для допомоги у створенні контенту та виконанні рутинних завдань, таких як редагування або автоматична генерація шаблонів текстів. Водночас, коли ШІ використовується в більш складних, креативних сферах контенту, таких як написання сценаріїв або розробка оригінальних відео, його роль залишається обмеженою. У зв'язку з цим використання штучного інтелекту може створювати обмеження для творчості, нав'язуючи шаблони та алгоритмічні обмеження. Алгоритми часто нав'язують певні стандарти для створення контенту, що ускладнює застосування оригінальних, ідеальних підходів [20]. Це призводить до зниження гнучкості та унікальності створеного контенту, роблячи його менш адаптованим до унікальних потреб аудиторії. В результаті контент може втратити свою привабливість та індивідуальність.

Варто також зазначити, що штучний інтелект може впливати на якість звуку, що супроводжує аудіовізуальний контент. Алгоритми, які автоматично генерують музичні треки або звукові ефекти, часто не враховують складність композиції та культурне значення звукових елементів. Це може призвести до монотонності та стандартизації звукового дизайну, який не відображає естетичні чи емоційні аспекти, необхідні для високоякісного контенту. Замість створення унікальних звукових ландшафтів, автоматизовані системи часто відтворюють типові шаблони, які можуть здаватися надмірно механічними та позбавленими оригінальності. Ще одним наслідком розвитку штучного інтелекту в контент-індустрії є поширення технологій дипфейків, які дозволяють створювати реалістичні фальшиві матеріали [21]. Ці технології, використовуючи алгоритми штучного інтелекту, можуть змінювати відео та аудіо, додаючи або змінюючи вирази обличчя, голоси та поведінку людей на екрані. Результатом може бути створення цілком переконливих матеріалів, які обманюють глядача та можуть мати руйнівний вплив на репутацію окремих осіб, організацій або навіть цілих країн. Наприклад, у 2020 році були відомі випадки використання дипфейків для створення фальшивих політичних заяв, які могли впливати на громадську думку та

					БР.ІП - 85.00.00.000 ПЗ	Арк.
						24
Змн.	Арк.	№ докум.	Підпис	Дата		

політичні процеси в країнах, навіть без їхнього відома чи згоди. Технології дідфейків не лише маніпулюють інформацією, але й відкривають можливості для створення фейкових новин. Вони можуть бути використані для політичних маніпуляцій, поширення дезінформації або навіть шахрайства. Ці технології не лише створюють підроблені зображення чи відео, але й можуть генерувати людські голоси, що дозволяє імітувати телефонні розмови. Такі дії можуть мати серйозні наслідки для довіри до джерел інформації та журналістської практики, оскільки стає все важче відокремити правдиві матеріали від фальшивих.

Таким чином, оскільки ШІ активно використовується для редагування та створення аудіовізуальних матеріалів, виникають проблеми, пов'язані з втратою індивідуальності та творчого підходу до контенту. Автоматизовані системи не завжди враховують нюанси, які роблять контент унікальним та привабливим для глядачів [22]. Слід особливо зазначити, що технології дідфейків відкривають нові можливості для створення фейкових матеріалів, ставлячи під сумнів достовірність представленої інформації та ставлячи під загрозу довіру до використання ШІ у створенні контенту.

1.4 Висновки по розділу

У першому розділі було досліджено теоретичні основи застосування штучного інтелекту у відеомонтажі та редагуванні контенту. Розглянуто принципи роботи систем штучного інтелекту у сфері обробки відео, а також визначено їх роль у автоматизації створення мультимедійного контенту. Проаналізовано сучасні підходи до автоматичного монтажу, класифікації та обробки відеоматеріалів із використанням алгоритмів машинного навчання. Окрему увагу приділено проблемам, ризикам та етичним аспектам використання ШІ у контент-індустрії, зокрема питанням авторського права, достовірності контенту та впливу автоматизації на творчі процеси. У результаті встановлено, що використання ШІ суттєво підвищує швидкість та ефективність відеовиробництва, проте потребує дотримання етичних і технологічних обмежень.

					БР.ІІ - 85.00.00.000 ІЗ	Арк.
						25
Змн.	Арк.	№ докум.	Підпис	Дата		

РОЗДІЛ 2. МЕТОДИ ТА ІНСТРУМЕНТИ ІНТЕЛЕКТУАЛЬНОГО МОНТАЖУ ВІДЕО

2.1 Методи машинного навчання та згорткові нейронні мережі для аналізу відеоданих

Згорткові нейронні мережі (ЗНМ) є основною технологією, що лежить в основі ІК. ЗНМ – це клас нейронних мереж, які виявляють ознаки типу зображення шляхом фільтрації через околицю пікселів з навчальними коефіцієнтами фільтра, щоб вивчити загальні закономірності в наборі навчальних зображень [23]. У контексті ІК ЗНМ забезпечують базові архітектури для численних завдань, таких як візуальний генеративний ШІ, виявлення об'єктів та аерокінематографія.

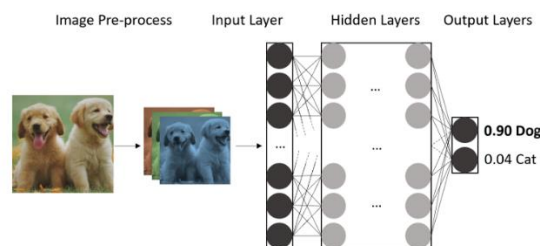


Рисунок 2.1 - Приклад діаграми CNN для класифікації зображень

Поширеним застосуванням CNN є класифікація зображень; базова структура показана на рисунку 2.1. Спочатку зображення подається на перший шар через N каналів, де кожен канал представляє різне представлення зображення. Наприклад, зображення RGB матиме три значення (червоний, зелений та синій) для кожного пікселя, таким чином $N = 3$. Вихідні дані цього шару потім надходять до серії прихованих шарів. Вихідні дані кожного шару називаються *картами ознак*. Останній прихований шар перетворює ці карти ознак на вектори ознак, щоб вихідний шар міг інтерпретувати ймовірність класу, використовуючи вектор ознак. Потім зображення класифікується, якщо його ймовірність потрапляє в

заздалегідь визначений довірчий інтервал. У випадку з рисунком 2.1, якщо наш довірчий інтервал становить 0,9, ми класифікуємо зображення як таке, що містить собак.

Приховані шари CNN можна гнучко налаштовувати. Чотири зазвичай використовуються компоненти: згортковий шар, шар функції активації, шар об'єднання та повнозв'язний шар:

1. Згортковий шар застосовує ядро (фільтр) розміром $n \times m$ до кожної околиці пікселя з вхідного каналу. Для роботи з пікселями по периметру зображення ми зазвичай застосовуємо доповнення (тип доповнення може змінюватися) [23]. Розширені згортки або деформовані згортки також використовуються для модифікації структури ядра для особливих випадків. Крім того, для пропуску деяких пікселів можна застосувати крок, тобто фільтр застосовується до пікселів, розташованих на s відстанях один від одного ($s = 1$ означає, що всі пікселі фільтруються, а $s = 2$ означає, що береться кожен другий піксель у рядку та стовпці).

2. Функція активації визначає, як ознаки шару перетворюються на вихід. Щоб використовувати функцію активації в прихованому шарі, вона повинна бути диференційованою та нелінійною, інакше ми не змогли б обчислити градієнти параметрів під час зворотного проходу через мережу. Шари з випрямленою лінійною одиницею (ReLU) та шари з витікаючими ReLU часто використовуються, оскільки вони є диференційованою нелінійною функцією, з додатковою перевагою, що вони не схильні до зникнення градієнтів (де градієнти параметрів стають занадто малими, щоб внести регресивну зміну в параметри моделі). Існує кілька типів функцій активації.

3. Шар об'єднання ефективно зменшує кількість вибірок ознак, зберігаючи важливіші ознаки з набору карт ознак.

4. Повністю зв'язаний шар - це шар, де кожен вузол залежить від усіх виходів попереднього шару. Це дозволяє мережі запроваджувати широкий спектр залежностей між параметрами.

У багатьох випадках в ІК існує потреба не лише в автоматичному розпізнанні

					БР.ІП - 85.00.00.000 ПЗ	Арк.
						27
Змн.	Арк.	№ докум.	Підпис	Дата		

наванні об'єктів, але й у визначенні їх місцезнаходження на зображенні або у відеопослідовності. Наприклад, у прямій трансляції футбольних матчів існує потреба класифікувати пози гравців з метою автоматизації керування камерою. Це може допомогти на наступних етапах гри чи події, що призведе до кращої організації кадрів. Враховуючи різноманітність фонових перешкод (наприклад, від натовпу, трави, динамічних рекламних банерів тощо), виявлення об'єктів створює значні труднощі. Розташування можна визначити або за допомогою контурування краю об'єкта на рівні пікселів, або за допомогою обмежувальної рамки. Перший варіант включає класифікацію та сегментацію на рівні пікселів, які разом називають семантичною сегментацією. На противагу цьому, другий варіант включає гілку регресії для оцінки чотирьох кутів обмежувальної рамки поряд з гілкою класифікації. Виявлення об'єктів за допомогою обмежувальних рамок швидше, ніж семантична сегментація, і широко використовується в програмах реального часу, таких як спортивні трансляції. Наприклад, моделі «Живеш тільки раз» (YOLO) для виявлення об'єктів та сегментації зображень потребують лише одного проходу мережі для визначення місцезнаходження [24]. Крім того, маскуванню зображень може допомогти YOLO зменшити стаціонарний шум, виключаючи відомі області пікселів, які не містять відповідної інформації під час виведення. Реальний випадок використання цього продемонстрували дослідники Каунаського технологічного університету, які застосували онлайн-виявлення та сегментацію об'єктів у режимі реального часу як засіб видалення відволікаючих факторів в камері для трансляції футбольного матчу.

YOLO працює шляхом поділу цільового зображення/кадру на N комірок розміром $s \times s$; вони називаються залишковими блоками. Потім, використовуючи регресію обмежувального прямокутника, модель оцінює, які залишкові блоки пов'язані з якими обмежувальними прямокутниками. Також прогнозується ймовірність присутності об'єкта в залишковому блоці. Використовуючи немаксимальне придушення, модель здатна зменшити вплив низьких показників ймовірності та визначити обмежувальні прямокутники з найвищою впевненістю; як показано на рисунку 2.2. Після цього, для вимірювання точності ми беремо перетин

					БР.ІІІ - 85.00.00.000 ПЗ	Арк.
						28
Змн.	Арк.	№ докум.	Підпис	Дата		

по об'єднанню (IoU) обмежувальної рамки істини на основі даних та оціненої рамки, щоб визначити, наскільки близько до правильної обмежувальної рамки знаходиться прогноз. Це забезпечує функцію втрат для використання під час оптимізації моделі.

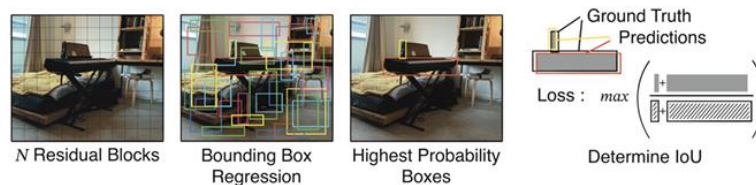


Рисунок 2.2 - Візуалізація процесу YOLO на прикладі визначення показника IoU.

Реальні об'єкти позначені сірими прямокутниками, а прогнозовані - різнокольоровими прямокутниками. Кілька версій YOLO використовувалися в ІК, який використовувався в КТУ для виявлення та сегментації дистракторів для футбольних ігор. YOLOv2 підтримує прогнозування фіксованої кількості об'єктів в одному блоці, що призводить до покращеного виявлення менших та згрупованих об'єктів. YOLOv3 базується на цьому, вводячи модель (з вищою складністю), яка може виявляти менші об'єкти та зберігати дрібні деталі. YOLOv4 (запропонований різними авторами) доповнює YOLOv3 перехресною міні-пакетною нормалізацією для вищої точності та зваженими залишковими зв'язками для кращої конвергенції під час навчання. Щодо ІК, ми виявили, що YOLO застосовується в прямих трансляціях та аерокінематографія.

Оцінка пози камери приблизно визначає орієнтацію, траєкторію та рух (зовнішні фактори) камери в тривимірному просторі, що, наприклад, корисно як основа для зменшення розмиття руху, спричиненого рухом камери. Точне відстеження камери також дозволяє автоматизоване виявлення об'єктів, виявлення пози людини та тривимірне захоплення (фотограмметрія). Для цього існують численні (комерційні) рішення, що включають додаткове обладнання (наприклад, LIDAR або GPS) для локального та глобального позиціонування. Хоча

					БР.ІП - 85.00.00.000 ПЗ	Арк.
						29
Змн.	Арк.	№ докум.	Підпис	Дата		

також проводяться дослідження рішень на основі відео для випадків, коли фізичні системи позиціонування недоступні.

Існує три основні підходи до оцінки пози: (1) візуальна одометрія (VO), (2) візуальна одночасна локалізація та картографування (vSLAM) та (3) структура з Рух (SfM), усі з яких мають спільні компоненти. Метою VO є поступове відновлення шляху камери, оптимізуючи поточну позу, враховуючи попередній набір поз (коригування групи вікон). Ми вважаємо VO недалекоглядним рішенням, оскільки воно має проблеми зі зв'язуванням поз, які перетинаються з раніше відвіданими місцями, тобто новий набір кадрів буде розглядатися як окремі місця [25]. Таким чином, vSLAM вивчає додаткову карту глобальної локалізації, а також успадковує мету VO щодо оптимізації узгодженості локальної траєкторії. Обмеженням використання vSLAM є те, що існує обмежена кількість камер та коригувань групи вікон, які можна обробити. Підходи SfM подолали це, використовуючи вимірювання, зроблені для кожної точки зору/кадру, щоб підвищити надійність та уникнути дегенеративних випадків. Це популярний спосіб попередньої обробки 2D-зображень для 3D-моделювання.

Крім того, обговорення цієї теми включають роботи, які порівнюють методи vSLAM та SfM для високодинамічних сценаріїв, які оцінюють методи vSLAM та SfM у період з 2010 по 2016 рік, які порівнюють сучасні загальні методи VO та vSLAM для робототехнічних контролерів (що застосовуються до автоматизованого керування камерами в ІС). Існує безліч варіантів використання в ІС: наприклад ми обговорюємо кілька випадків, коли це дослідження пов'язане з аерофотозйомкою [26]. Крім того, дослідження калібрування камери значною мірою підтримують роботу над автоматизованою 3D/об'ємною зйомкою, оскільки здатність позиціонувати зображення в 3D-просторі тісно пов'язана з якістю конвеєрів зворотного рендерингу. Це питання детальніше обговорюється в наступному підрозділі та повторно розглядається в основній частині.

Автоматизований 3D-контекст захоплення

Автоматизоване 3D-зйомка є популярною темою в дослідженнях ІС, як для аналізу, так і для вирішення оберненої графічної задачі. Тут метою є

					БР.ІІІ - 85.00.00.000 ПЗ	Арк.
						30
Змн.	Арк.	№ докум.	Підпис	Дата		

створення 3D-зображення за допомогою 2D-зображень та/або інших даних датчиків. На рисунку 2.3 ми ілюструємо спрощені представлення сучасних підходів до проблеми.



Рисунок 2.3 - Фотограмметрія спирається на дані датчиків (може включати камери) для оцінки представлення сцени у вигляді хмари точок

Популярним методом зйомки є лазерно-інформована фотограмметрія (наприклад, LiDAR), яка створює 3D-хмари точок на основі вимірювань глибини точки відносно положення камери. Тепер це можливо на недорогих платформах, таких як мобільний телефон, завдяки появі мобільних датчиків LiDAR-камер. Ці методи є перспективними для зйомки замкнутих просторів, проте залишаються проблеми щодо обробки інформації на великих відстанях, а також обмеження, пов'язані з шумними даними.

Поля нейронного випромінювання

Нещодавно для вирішення цієї проблеми були запроваджені методи нейронного поля випромінювання (NeRF), див. рис. 2.4.

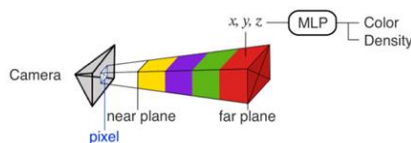


Рисунок 2.4 - Нейронні поля випромінювання вибирають точки в просторі для оцінки значення кольору та щільності для кожного зразка, проєктованого вздовж піксельного променя

Метою яких є моделювання процесів рендерингу, пов'язаних з автоматичним захопленням. На відміну від Photogrammetry, які представляють сцени як

					БР.ІП - 85.00.00.000 ПЗ	Арк.
						31
Змн.	Арк.	№ докум.	Підпис	Дата		

3D-структури (сітка, воксельна сітка, хмара точок), методи NeRF представляють властивості сцени як нейронні представлення, які дискретизуються для інтерполяції видів, визначених зовнішніми властивостями камери.

Маючи набір зображень, що описують реальне середовище, мережі NeRF можуть надійно вивчати візуальні особливості сцени.

NeRF є важливим проривом для кінематографістів і можуть бути адаптовані для низки кінематографічних завдань. Наприклад, рендеринг 3D з 2D-зображень може уникнути дорогих перезйомок, спричинених поганим освітленням, пейзажами, погодою або недоліками акторської гри. Кадри також можна переробляти в постпродакшні, а це означає, що можна змінювати тип кадру, тремтіння камери, тривалість кадру та фокус [27]. Хоча галузь досліджень NeRF є відносно новою, вона швидко розвивається і тепер може пропонувати високоякісні, компактні рішення, з обіцянкою випуску готових до виробництва NeRF найближчим часом.

Показано, що NeRF за наявності гарного наближення зовнішніх характеристик камери (положення та орієнтація зображення) та надійного семплера променів (для вибірки точок/об'ємів вздовж променя) як передумов, ми можемо моделювати 3D-сцени, що включають відбивні та прозорі матеріали, використовуючи багат шарову перцептронну (MLP) нейронну мережу рисунок 2.5.

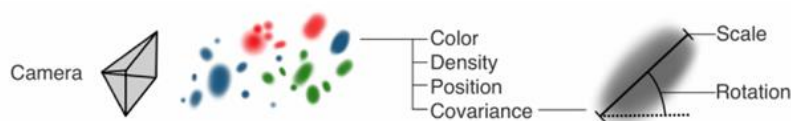


Рисунок 2.5 - Гаусівський сплетинг використовує хмару точок для оцінки положення зразків у просторі та використовує це для оцінки кольору, масштабу щільності та обертання кожної гаусової точки.

Це працює шляхом вибірки точки вздовж променя (що стосується заданого пікселя на навчальному зображенні) та використання глобального 3D-положення цієї точки $\{x, y, z\}$ та відносного напрямку перегляду точки $\{\theta, \phi\}$ як вхідні

					БР.ІІІ - 85.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		32

дані для MLP який повертає колір RGB та значення непрозорості, пов'язане з цією точкою в 3D-просторі. Акумулюючи багато точок вздовж променя, ми можемо застосувати метод рендерингу, подібний до альфа-змішування, для рендерингу кожного пікселя зображення. За допомогою функції втрат, що порівнює візуалізоване зображення з еталонним зображенням, ми можемо оптимізувати якість 3D-сцени, щоб можна було створювати нові, небачені раніше ракурси. Отже, дослідження NeRF підпадає під поняття синтезу нових ракурсів.

На відміну від оригінального NeRF, NeRF-SH використовує сферичні гармоніки для декодування візуальних ознак у значення кольорів RGB. Модель PlenOctree уточнює 3D-представлення октодерева для моделювання геометрії сцени. Хоча цей метод працює, навчання MLP займає багато часу, а якість часто низька та залежить від кількості навчальних зображень/видів 3D-сцени, які ми маємо. Згодом було запропоновано кілька підходів для пришвидшення обчислень NeRF: хеш-кодування з кількома роздільними здатностями, воксельні сітки та воксельні дерева. Це особливо актуально для дослідників, зацікавлених у динамічній парадигмі NeRF, оскільки складність моделі значно зростає при моделюванні 4D простору та часу [28]. Представлення PlenOctrees, проілюстроване на рисунку 2.7, базується на здатності NeRF функціонально представляти сцену, залежну від вигляду, шляхом (1) представлення структури сцени як розрідженого воксельного октодерева, а також (2) вибірка кольору шляхом залучення сферичних гармонік. (1) полегшує крок рендерингу за допомогою швидкодоступних представлень (октодерев). Тоді як (2) дозволяє нам моделювати залежність виду від листя октодерева, відображаючи залежні від виду значення кольору на поверхню сфери та прогножуючи коефіцієнти рівняння сферичної гармоніки (з фіксованим ступенем), щоб повернути значення кольору там, де промінь пікселя перетинає сферу. Згодом швидкість рендерингу та навчання значно оптимізуються, і можна стверджувати, що якість нового виду також покращилася. Одним застереженням є обсяг даних, який споживатиме представлення (у розмірі гігабайт), що, можливо, прийнятно для тих, хто має необхідне обладнання, проте існують більш доступні явні варіанти. Крім того, існують певні труднощі з

					БР.ІП - 85.00.00.000 ПЗ	Арк.
						33
Змн.	Арк.	№ докум.	Підпис	Дата		

візуалізацією необмежених сцен, оскільки воксельні сітки займають скінченний простір [29]. Долають це, використовуючи NeRF++ для рендерингу пейзажів за межами об'єкта. Це досягається шляхом моделювання переднього плану та фону як окремих компонентів.

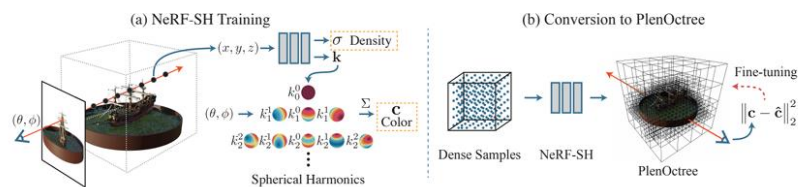


Рисунок 2.6 - Візуальне порівняння конвеєрів NeRF та PlenOctrees, що використовуються для нейронного рендерингу.

Гаусівське розбризкування. Запропоновано 3D гауссове сплетингове розтягування (GS) як засіб значного скорочення часу рендерингу за допомогою: (1) представлення гауссових «плям» у вигляді хмари точок з властивостями положення, коваріації, кольору та непрозорості, що дозволяє уникнути непотрібних обчислень порожнього простору, присутнього в моделях NeRF, та (2) *сплетингу тайлів* для швидшого рендерингу (NeRF зазвичай рендерить зі швидкістю < 1 FPS).

Розподіл тайлів пришвидшує рендеринг на наступних етапах. Ми підкреслюємо важливість підходу сортування (3), оскільки це обмежує NeRF у досягненні подібних обчислювальних цілей.

1. Розділити вигляд рендерингу на плитки 16×16
2. Відбракувати краплі з впевненістю $< 1\%$ перетину кожної усіченої кульки плитки або краплі, що виходять за межі ближньої та дальньої меж камери
3. Сортувати глибину блобів відносно кожної плитки за допомогою швидкого сортування Radix на графічному процесорі Merrill та Grimshaw; не по пікселях, як це зроблено в NeRF.
4. Відрендерити пікселі відносно відсортованих блобів для кожної плитки за допомогою α -змішування, доки сумарне α для кожного променя не стане рівним 1.

					БР.ІІІ - 85.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		34

Для навчання моделі GS сцена ініціалізується розрідженим набором хмар точок, які «ущільнюються» під час навчання шляхом клонування, масштабування та переміщення блобів, щоб вони відповідали геометрії, що відповідає набору навчального вигляду/зображення. Блоби, які є по суті прозорими (з низьким внеском), обрізаються.

Зрештою, GS може бути обраний замість NeRF через їхні покращені обчислювальні можливості та, в багатьох випадках, покращену продуктивність. Однак існуючі динамічні моделі GS набагато менш компактні, ніж альтернативи NeRF, і порівняно гірші в оцінці глибини через різницю в підходах до рендерингу на основі глибини. Крім того, оскільки GS є новішою розробкою, ніж NeRF, ми ще не побачили такої ж уваги до кінематографічних інструментів, таких як редагування сцен і камери або репрезентативні перетворення, такі як GS to mesh [30]. Однак швидкий розвиток цієї дослідницької галузі означає, що це лише питання часу, коли ці обмеження будуть усунені.

2.2 Інструменти автоматичного редагування, генерації та маніпуляції відеоконтентом

Існує кілька методів категоризації відеомонтажу. Редагування варіюється від рівня гучності та кадрів до рівня об'єктів. Редагування на рівні гучності вставляє або видаляє кліпи вхідного відео, вибирає набір кадрів або знімків або видаляє латки гучності. Редагування на рівні кадрів включає розфарбовування, передачу стилів тощо. Хоча модифікація об'єктів у кадрах переважно відбувається на рівні об'єктів, вона також залежить від типу вхідного відео, як обговорювалося вище. Тут ми класифікуємо алгоритми редагування відповідно до їх маніпуляцій: перенацілювання, яке змінює розмір і співвідношення відео, підсумовування, яке значно скорочує тривалість відео, та додавання спеціальних ефектів, які переважно змінюють вміст у кадрах.

Ретаргетинг відео

Різні дисплеї вимагають різних розмірів та співвідношень сторін одного

					БР.ІІІ - 85.00.00.000 ПЗ	Арк.
						35
Змн.	Арк.	№ докум.	Підпис	Дата		

й того ж відео, тому ретаргетинг стає практичною потребою. Маніпуляції ретаргетингу можна розділити на 5 класів: обрізання, масштабування, перегляд, різьблення по швах та деформація. На рисунку 2.7 показано відмінності між способами ретаргетингу.



Рисунок 2.7 - Перенацілювання широкоекранного запису на менші співвідношення сторін.

Перед обрізанням слід визначити ROI або карту помітності кожного кадру. Після визначення ROI, програма обрізання повинна згладити результати, враховуючи рух на екрані та віртуальну кінематографію. Обрізають кожен кадр та панорамують його до потрібного розміру. Він поєднує рух, розширену помітність та методи ідентифікації об'єктів зі статичних зображень, щоб визначити ROI, але не враховує залежність від темпу між кадрами протягом усього відео. Він визначає втрату перенацілювання як суму втрати інформації, масштабування, співвідношення сторін пікселів, вартості обрізання обличчя, вартості скупчення країв, вартості панорамування та обрізання та вартості підказок користувача, і використовує перебір, щоб знайти оптимальне рішення. Оптимізують шлях вікна обрізання за допомогою трьох основних операцій (панорамування, обрізання та масштабування) на основі зібраних даних відстеження погляду. ROI вибирається шляхом вивчення даних погляду глядачів за допомогою алгоритму RANSAC [31]. А для оцінки цього алгоритму порівнюються дані відстеження погляду оригінальних фільмів та перенацілених фільмів. Після оптимізації шляху вікна обрізання, додають деякі операції різьблення швів. Перенацілюють відео за допомогою відстеження погляду. Їхній робочий процес виглядає наступним чином. Спочатку він визначає, де знаходяться нові зрізи, використовуючи дані

погляду за допомогою динамічного програмування. Він оптимізує шлях вікна обрізання відповідно до принципів кінематографії та використовує $L(1)$ регуляризований опуклий оптимізаційний розв'язувач. Використовують імітатор плівки кожного кадру для керування вилученням кінематографічних ознак та створення карти важливості на стороні сервера. На стороні клієнта, програма обрізання перенацілює плівку на потрібний розмір. Кадри класифікуються відповідно до поведінки камери, а об'єкти в кожному кадрі додатково класифікуються за їхніми рухами [32]. Шукають оптимальний шлях вікна обрізання, який зберігає просторово-часову помітність та грані за допомогою методу Max-Flow/Min-Cut. Для створення карти помітності потрібно враховувати більше факторів: рівень зосередженої уваги, загальний бал помітності та упередженість через штраф за центр.

Масштабування – це збереження співвідношення ширини до висоти незмінним під час їхньої модифікації. Запропоновано перенацілювання відео шляхом сегментації відео на просторово-часові сітки. Вони використовують потік сітки для вибору ключових кадрів та зміни розміру потоків сітки в цих ключових кадрах. Розмір лівих кадрів змінюється шляхом простої інтерполяції їхнього вмісту сітки з двох найближчих перенацілених ключових кадрів. Цей алгоритм має низьку обчислювальну складність і може зберігати форми визначних об'єктів вздовж часової осі. Змінюючи розмір кожного кадру відео незалежно на основі їх об'єктів визначних об'єктів, а потім оптимізують їхні рухи для кожної лінії шляху оптичного потоку. Спочатку вирівнюють кадри в одній системі координат, оцінюючи рух камери, змінюючи розмір кожного кадру просторово та часово когерентно, а потім реконструюючи отримані кадри в початковій системі координат.

Деформація. Побудова відображення між кожною парою кадрів вихідного відео та цільового відео є деформацією [33]. Пропонується метод перенацілювання з деформацією в реальному часі та точністю до пікселів з 2D-варіантом EWA splatting. Перенацілюють відео в стислому домені для економії часу виконання. Відео спочатку частково декодується, обрізається за допомогою карти

					БР.ІІІ - 85.00.00.000 ПЗ	Арк.
						37
Змн.	Арк.	№ докум.	Підпис	Дата		

помітності, деформується на основі стовпцевої сітки до потрібного розміру та, нарешті, перекодується. Зосереджуються на усуненні артефактів тремтіння, одночасно враховуючи просторову та часову когерентність обчислює карту помітності на основі градієнта.

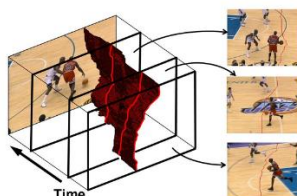


Рисунок 2.8 - 3D-куб і 2D-многовиміри швів

Величина, розпізнавання обличчя та розпізнавання руху, а також розглядає відображення між джерелом та перенаправленими пікселями як розріджену лінійну систему, розв'язану алгоритмом найменших квадратів. Карта помітності з інформацією про рух є причиною артефактів хвилювання та стискання. Щоб уникнути таких артефактів, вона враховує рух, обрізає вміст, що повторюється в часі, та деформує однорідні області, щоб замаскувати деформації та зберегти рух [34]. Пропонують інтерактивну систему перенацілювання, яка деформує відео за допомогою методу деформації координат середнього значення та уточнює результати, підсумовуючи часовий вихід на основі латки для усунення спотворень. Отже, цей алгоритм підтримує підсумовування відео, завершення та перетасування. Щоб уникнути повторних обчислень, Обчислюючи карту важливості для кожного пікселя кожного кадру, обчислюють кумулятивні карти стискання для напрямків x та y та зберігають їх. За будь-якої цільової роздільної здатності він може швидко деформувати відео відповідно до карт стискання. Деформують стерео 3D-відео з обмеженнями нерівності. Намагаючись деформувати відео з кількома режимами перегляду, додавши глибини.

Різьблення швів. Різьблення швів зменшує або розширює розмір зображення шляхом різьблення або вставки швів. Для 1D шва функція енергії зображення:

					БР.ІП - 85.00.00.000 ПЗ	Арк.
						38
Змн.	Арк.	№ докум.	Підпис	Дата		

$$e_{HoG} = \frac{(|\frac{\partial}{\partial x}I| + |\frac{\partial}{\partial y}I|)}{(\max(HoG(I(x,y))))} \quad (2.1)$$

Де I — зображення розміром $n \times m$, HoG. А вертикальний шов визначається так:

$$s^x = \{s_i^x\}_{i=1}^n = (x(i), i)_{i=1}^n, s.t. \forall i, |x(i) - x(i-1)| \leq k \quad (2.2)$$

де $k \in \mathbb{R}$, а x — відображення $x : [1, ..., n] \wedge [1, ..., m]$. Щоб зменшити розмір зображення, видаліть шов з найменшими втратами енергії; щоб збільшити розмір, знайдіть шов з найбільшою енергією та середнім значенням пікселів шва з їхніми сусідами.

Розширено 1D шви з 2D зображень на 2D многовиди швів з 3D просторово-часових об'ємів відео, як показано на рисунку 2.9. На відміну від 1D швів, які обчислюються за допомогою динамічного програмування, воно вирішує 2D шви шляхом знаходження мінімального розрізу в графіку. Крім того, пропонує нову функцію енергії, яка підкреслює енергію, спричинену видаленням швів. Спільно оптимізує енергію назад, пропонує захист ізоповерхні та кодує функцію передачі непрозорості. Використано фільтр Калмана для оптимізації різьблення швів, що є теоретично простим та зручним для апаратного забезпечення. Маніпулюють гістограмою помітності та пропонують алгоритм вирівнювання гістограми помітності - різьблення швів (SHE-SC). Вони перенацілюють перший кадр відео за допомогою SHE-SC та адаптують цей алгоритм для решти кадру відповідно до їхньої різниці від попереднього кадру. Оскільки 3D різьблення швів має високі обчислювальні витрати. Пропонують знайти неоптимальне об'ємне різьблення шва за допомогою багатопрохідного динамічного програмування.

Перенацілювання стереовідео повинно відповідати трьом принципам: збереження часової когерентності, запобігання спотворенню глибини та мінімізація спотворень форми перенаціленого відео. Пропонують сегментувати

					БР.ІІІ - 85.00.00.000 ПЗ	Арк.
						39
Змн.	Арк.	№ докум.	Підпис	Дата		

стереовідео на групи кадрів відповідно до витрат енергії, що включають помітність та стереоскопічну достовірність, а різьблення швів у групі є фіксованим. Спочатку обчислюється різьблення лівого шва.

Відео зведення. Існує кілька рішень для зменшення розміру відео. Одним з них є зменшення частоти кадрів шляхом пропуска кадрів, що називається гіперлапсом. Одним з них є вилучення одного або кількох ключових кадрів та створення нерухомих зображень. А одним є вибір підкадрів з помітністю та виведення скорочених відеокліпів. Загалом, слід дотримуватися двох основних критеріїв: 1) збереження якомога більшої кількості інформації; 2) максимальне зменшення кількості кадрів або розмірів. Таким чином, відеорежим полягає в пошуку балансу між обсягом інформації та розмірами виводу відповідно до потреб різних сцен. Подібно до описаного вище робочого процесу редагування, відеорежим зазвичай складається з сегментації, оцінки важливості кожного кадру або кадру, вибору найважливіших кадрів або кадрів та, як альтернатива, оптимізації виводу та рендерингу [35]. Для певних відео попередня обробка є обов'язковою. Наприклад, створення гіперлапсових відео в режимі реального часу для ручних камер вимагає стабілізації.

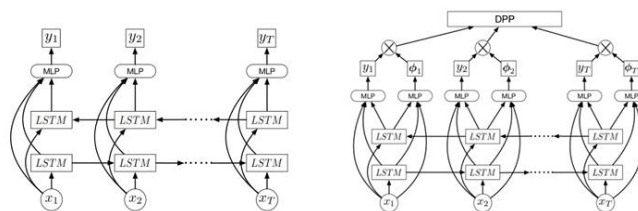


Рисунок 2.9 - vsLSTM та dppLSTM.

Сегментація є ключовим кроком. Рівномірно сегментують вхідні відео на кліпи по 50 кадрів для пришвидшення. Запропоновано використовувати нову криву кумулятивного зміщення для сегментації егоцентричних відео, що включають складний рух, і доводять, що інтегровані вектори руху працюють краще, ніж вектори миттєвого руху. Те, що обчислення гістограми кольору для кадрів та виявлення меж кадру, коли різниці між двома сусідніми кадрами нижчі за

					БР.ІП - 85.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		40

певний поріг, є простим рішенням. Запропоновано алгоритм сегментації ядра за часом (KTS). Сегментують відео відповідно до змін кадру та карт помітності. Виявляють межі кадру, обчислюючи інтенсивність руху за допомогою адаптивного алгоритму пошуку візерунка між двома кадрами протягом усього відео, а потім визначаючи поріг. Сегментують відео на основі руху камери, наприклад, панорамування, масштабування, пауза. Запропоновано дві мережі LSTM для моделювання часової залежності між кадрами та вилучення ключових кадрів або ключових підкадрів. Структури vsLSTM та dppLSTM показано на рисунку 2.9.

Оцінка важливості кадрів і знімків дуже різниться. Запропоновано часове відображення, а саме перетворення відео з високою частотою кадрів на відео з низькою частотою кадрів. У цій роботі пропонується новий метод визначення помітності, техніка пересинхронізації для часової передискретизації на основі важливості кадру, згенерованої цим методом визначення помітності, а також представлені два нові часові фільтри (адаптивний блоковий фільтр та фільтр розмиття руху на основі помітності) для покращення візуалізації помітного руху, також згенерованого цим методом визначення помітності. Новий метод визначення помітності є методом «знизу вгору» та використовує сентиматичну сегментацію та оптичний потік. Результати показують, що фільтр розмиття руху на основі помітності працює найкраще. Запропоновано попарну модель ранжування, яка навчається на онлайн-відео та оцінює найяскравіші моменти відеокліпів без обмежень. LiveLight спочатку сканує та сегментує вхідне відео на основі словника. Він створює словник відеосегментів, додаючи новий сегмент, який не можна було розріджено реконструювати за допомогою поточного словника. Також запропоновано підхід до навчання сегментованих відео на основі словника, який враховує втрати реконструкції, регуляризацію групової розрідженості, регуляризатор збереження структури на рівні патчів та на рівні кадрів.

Вихідними даними можуть бути ключовий кадр, кілька ключових кадрів (статичні розкадровки), статичний розкадровки або короткий відеокліп. Багато класичних робіт з відеореюме були завершені. З 2006 року пропонуються методи, засновані на глибокому навчанні, для резюме відео. Резюмовано відео на

					БР.ІІІ - 85.00.00.000 ПЗ	Арк.
						41
Змн.	Арк.	№ докум.	Підпис	Дата		

основі їхніх рядків. Розглянуто проблему вилучення ключових кадрів та перегляду відео як проблему вибору словника. Запропоновано вилучати ключові кадри з бездротової капсульної ендоскопії. Запропоновано новий метод виявлення та відстеження людини на основі poselet та нового детектора помітності, навченого на даних погляду. Запропоновано послідовний процес детермінантних точок (seqDPP) для резюме відео з контрольованим способом. Він також надає метрики оцінки. Маючи два зведені дані, визначається збіг кадрів, візуальна відстань яких нижча за поріг, та обчислюється їхня точність, повнота та F-оцінка. Також синтезується зведена інформація про базові дані для кожного відео шляхом максимізації F-оцінки між базовими даними та кількома анотаціями, доданими людиною.

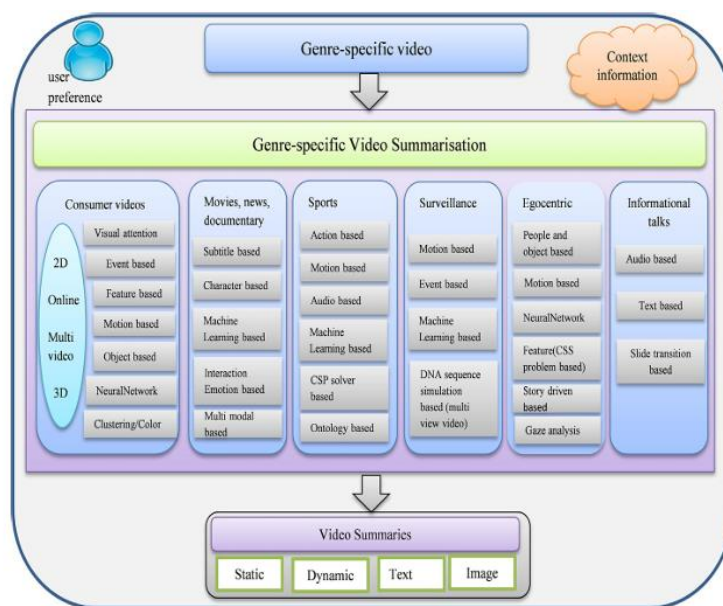


Рисунок 2.10 - Алгоритм роботи

Для відео різних жанрів методи їх узагальнення також різні. Див. малюнок 2.10. Візьмемо для прикладу егоцентричні відео. Запропоновано систему для підсумовування відео, що носяться, за допомогою сигналів мозкових хвиль. Відтоді було проведено багато досліджень щодо підсумовування егоцентричних відео. Використовують метод опорних векторів для підсумовування відео, що носяться [36]. Зосереджено на виявленні високорівневої помітності в егоцентричному відео та формуванні розкадрування. Запропоновано нове рішення для

редагування відео, яке орієнтоване на історію та підкреслює причинно-наслідковий зв'язок між підкадрами. Використано веб-зображення перед вилученням ключових кадрів з відео низької якості, створених користувачем. Вибирає суперкадри відповідно до їхньої цікавості та пропонує бенчмарк SumMe. Додатково вдосконалює алгоритм підсумовування контрольованим способом та спільно оптимізує для кількох цілей. Наголошено на стабілізації егоцентричних відео та використовує їхні зміщення для перетворення 2D-відео на стерео. Запропоновано серію алгоритмів на основі графів для виявлення меж кадру, вибору ключових кадрів, захоплення характеристик кадру, вилучення ознак та кластерування ключових кадрів. Використано веб-зображення перед виявленням точок прив'язки в егоцентричних відео.

Використали метод розпізнавання та відстеження людських облич для вирішення задачі сумарного опису персонажів. У той час як узагальнюють багатокористувацькі ігри з 3D-сценами на основі правил гри.

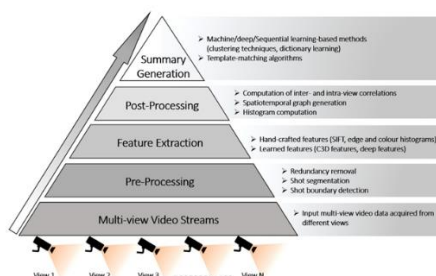


Рисунок 2.11 - Огляд відео з кількома режимами перегляду

Окрім одного вхідного відео, також існує практична потреба у відеозведенні кількох відео. На рис. 2.11 показано загальний робочий процес системи багатовидового відеозведення. Спочатку систематично досліджують проблему багатовидового відеозведення та розглядають проблему багатовидового відеозведення як завдання маркування графів. Вона сегментує відео на кадри та обчислює їхні важливі оцінки на основі низькорівневих ознак (ознака гістограми кольору, ознака гістограми країв та ознака вейвлета) та високорівневих ознак [37]. Шляхом вибору будується просторово-часовий граф кадрів імпортованих

					БР.ІП - 85.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		43

кадрів, ребра якого представляють подібність між вузлами. Кадри кластеризуються за допомогою випадкових блукань та багатоцільової оптимізації. Пропонують метод корегуляризації центроїдів (MSDS-CC) для вибору репрезентативних візуальних елементів. Тоді використовують алгоритм SVM для абстрагування ключових кадрів та подальшого зменшення ключових кадрів за допомогою грубого набору. Прагнуть зменшити включення латок у відео з результатами зведення для кількох камер з фіксованою точкою огляду шляхом оптимальної реконструкції. Також розглянуто проблему зведення багатовидового відео як теоретико-графову. Окрім звичайних кроків, таких як сегментація та вилучення ознак, вона використовує гаусівську ентропію для видалення надлишкових кадрів, використовує двочасткове зіставлення графів для обчислення міжвидових залежностей та застосовує алгоритм лісу оптимального шляху до кластерних ключових кадрів. Вивчають метрики та виводять ключові кадри, а не відео. Щоб зменшити споживання енергії стиснення та передачі бездротовими відеодатчиками, розробили простий, але корисний алгоритм для зведення багатовидових відео. Він повинен бути онлайн, з низькою обчислювальною складністю та вимогами до пам'яті. Крім того, як і для мереж багатовидових датчиків, накладні витрати на зв'язок між вузлами повинні бути якомога меншими. Цей алгоритм складається з двох етапів: внутрішньовидового етапу та міжвидового етапу. На етапі внутрішньовидового перегляду характеристики кадру витягуються за допомогою дескриптора кольорового розташування MPEG-7, кластеруються за допомогою спрощеної моделі гаусової суміші (GMM), і вибираються кадри з меншими вагами GMM та більшими дисперсіями. На етапі міжвидового перегляду зберігається лише один кадр тієї ж події. Також враховано внутрішні та міжкореляційні зв'язки та зосереджується на реалізації розрідженості. Використано метод кластеризації на основі домінантного набору на основі гіперграфа для визначення ключових кадрів та використання веб-зображень для подальшого зменшення надлишковості.

Окрім відео з кількома ракурсами та перекриттям полів зору, абстрагували візуальні кадри спільної появи з безлічі відео на певну тему без обмежень.

					БР.ІП - 85.00.00.000 ПЗ	Арк.
						44
Змн.	Арк.	№ докум.	Підпис	Дата		

Вони сегментують відео за змінами кольору між двома кадрами та витягують особливості кадрів за допомогою CENTRIST, VLFeat та колірних моментів HSV. Кадри побудовані у вигляді двочасткового графу. Також пропонується алгоритм максимального біклічного пошуку для відкидання кадрів, які з'являються лише в одному відео.

Відео-конспект. Оскільки відеореzumе полягає у вилученні послідовних або непослідовних кадрів з оригінального відео, відеосинописис додатково модифікує вилучені кадри. Крім того, зміщення цікавих дій у часовій області є вибором для подальшого ущільнення вхідних даних. Ірані та ін. спочатку запропонували створити синописис для пошуку відео. Відеосинописис переставляє рухомі об'єкти в різний час і місцях на тих самих кадрах, зберігаючи їх розташування незмінним [38]. Пропонований робочий процес виглядає наступним чином: генерація активності, перестановка трубок, генерація фону, зшивання об'єктів, розв'язання функції витрат енергії протягом заданого часового інтервалу за допомогою імітованого відпалу. Ключовий момент відеосинописису полягає у знаходженні оптимальних часових позицій вибраних дій порівняно з вилученням ключових кадрів. Ця робота має багато послідовників.

Багатоканальний відеосинописис також є дослідженою проблемою. Пропонують рішення для створення багатоканального відеосинописису, що складається з 5 кроків: створення спільного фону, відповідність спільній площині, виявлення об'єктів, розпізнавання дій та динамічний відеосинописис. Спільний фон - це зображення виду зверху кількох камер, яке можна отримати за допомогою Google Map для зовнішньої зйомки або шляхом зміни попереднього чернетки ескізу. Відповідність спільній площині досягається шляхом зіставлення всіх камер зі спільною координатою. Розпізнавання дій реалізується за допомогою SimpleMKL. Енергетична функція з точки зору втрати інформації, колізій та довжини результуючого синописису оптимізована за допомогою імітованого відпалу [39]. Створення розкадровки потребує відстеження, семантичної сегментації, вилучення ключових кадрів, розширення макета кадру, макета анотацій, композиції та рендерингу.

					БР.ІІІ - 85.00.00.000 ПЗ	Арк.
						45
Змн.	Арк.	№ докум.	Підпис	Дата		

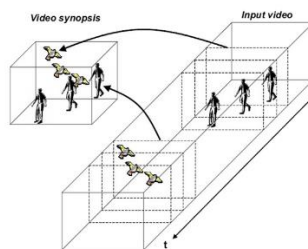


Рисунок 2.12 - Відео-конспект.

Окрім стиснення, відеосинописис також може функціонувати як метод індексації. Створюють монтажі, в яких значки або спрайти представляють події помітності та функціонують як індекси у вхідному відео. Ці спрайти витягуються шляхом побудови гаусової моделі суміші зі спряженими апріорними значеннями для фону поступово, а елементи переднього плану сегментуються за допомогою швидких морфологічних операторів.

Відеомозаїки. Відеомозаїка також перевпорядковує просторово-часові об'єми вхідних відео, але маніпулює послідовністю кадрів і не змінює вміст жодних кадрів. Запропоновано динамосаїку, яка об'єднує просторово-часові об'єми 2D-відео, що рухаються під час запису, для створення 360-градусних відео. Спочатку фільтрує відео за рухами камери, рухомими об'єктами та візуальною якістю. Синтезуйте панорами сцен за допомогою об'єднання, використовуючи алгоритм розтушовування, або додатково додайте рухомі об'єкти до панорам.

Спеціальні ефекти. Кліплет – це щось середнє між відео та зображенням, невелика його частина є динамічною, а ліва – статичною. Пропонується інтерактивну систему, яка підтримує стоп-кадр, відтворення, дзеркальне відображення та циклічне відображення чотирьох ідіом культового часового відображення з вхідного відео на цільовий кліп лет [40]. Після відображення результати уточнюються шляхом автоматичного вирівнювання, оптимізації циклу та розтушовування, одночасного матування та композиції, а також лапласіанського змішування. Cinemagraphs – це окремий випадок цього, оскільки його рух періодичний. Витягують цільовий жорсткий об'єкт з вхідного відео або зображення, а користувачі змінюють його фізичні параметри. Потім нова форма буде змодельована та вписана в оригінальне відео.

					БР.ІІІ - 85.00.00.000 ПЗ	Арк.
						46
Змн.	Арк.	№ докум.	Підпис	Дата		

Представлено новий алгоритм для знаходження візуального ритму та такту у відео, а потім деформації відео аудіодоріжкою для формування танцювального відео. Аудіо обробляється за допомогою STFT для отримання його спектрограми потужності, початкової обвідної, темпограми та тактів. Аналогічно, директограма відео отримується за допомогою оптичного потоку, а потім генерується обвідна удару, візуальна темпограма та візуальні такти. Стратегія інтерполяції використовується для синхронізації між аудіо та відео. Окрім створення танцювальних відео, вона також може змінювати танцювальні такти, щоб вони відповідали іншій пісні. У цьому сенсі можливе відеомонтаж на основі музики.

Запропоновано напіваавтоматизовану техніку вибіркового деанімування відео для видалення великомасштабних рухів одного або кількох об'єктів. Користувачеві потрібно малювати три види штрихів: зелені штрихи вказують, які області відео слід деанімувати; червоні штрихи – які області слід залишати статичними; сині штрихи – які повинні залишатися динамічними. Ця техніка складається з двох етапів: деформації та композиції. Відстеження (KLT) використовується для відстеження характерних точок у вхідному відео, і ті доріжки, тривалість яких становить менше 15% від вхідного сигналу, видаляються. Кожен вхідний кадр розділяється на прямолінійну сітку 64x32. У початковій деформації деформуються лише опорні доріжки, щоб мінімізувати функцію енергії. А в уточненій деформації вихід початкової деформації та плаваючі доріжки використовуються як вхідні дані для розв'язання кінцевої функції енергії за допомогою методу найменших квадратів. На етапі композиції розрізи графів використовуються для виконання оптимізації випадкового поля Маркова.

Запропоновано vid2play, систему, яка вивчає поведінку тенісистів за допомогою величезної анотованої бази даних. Вона може генерувати інтерактивно керовані відеоспрайти, які поведуться та виглядають як професійні гравці. Представлено систему, яка підтримує редагування відео на рівні об'єктів. Вона сегментує об'єкти та генерує альфа-мати, а також оцінює інформацію про 3D-сцену на основі алгоритму структури з руху (SfM). Користувачі можуть застосовувати 3D-перетворення до об'єктів, дублювати об'єкти або навіть переносити об'єкти між

					БР.ІІІ - 85.00.00.000 ПЗ	Арк.
						47
Змн.	Арк.	№ докум.	Підпис	Дата		

відео. Потім система моделюватиме 3D-сцену, візуалізуватиме кадри, використовуючи розріджені структурні точки та композитні шари. Запропоновано редагувати атласи відео. Враховуючи природне відео з грубою маскою об'єктів, що цікавлять, вона оцінює набір атласів для фону та об'єктів, що цікавлять. Користувачі маніпулюють одним або кількома атласами, наприклад, змінюють колір та додають текстуру. Потім цей алгоритм оцінює відображення кожного пікселя відео на двовимірну точку в кожному атласі, його непрозорість та послідовно поширює зміну по всьому відео. Для відображення, атласів та альф використовується представлення на основі координат багатошарового перцептрона (MLP). Цей алгоритм є самокерованим, а його функція втрат складається з втрати жорсткості, втрати узгодженості та втрати розрідженості.

Представлено два методи дубляжу ляльок: один напівавтоматичний на основі зовнішнього вигляду, а інший повністю автоматичний на основі аудіо. Крім того, у роботі також пропонуються три рекомендації щодо виконання лялькового мовлення: 1) кожен склад у мовленні повинен відповідати одному закритому-відкрито-закритому сегменту рухів губ ляльки, який називається візуальними складами; 2) Губи повинні бути нерухомими та закритими, коли лялька не говорить, що називається візуальними складами тиші; 3) у швидких мовних послідовностях кілька розмовних складів можуть відповідати одному візуальному складу. Вхідними даними для двох методів є задане лялькове відео та фрагмент нового аудіо мовлення, тривалість якого коротша за тривалість відео. Два підходи можна розділити на 4 кроки. Сегментуйте нову аудіодоріжку мовлення на послідовність складів, транскрибуючи її в текст із субтитрами з YouTube, вирівнюючи транскрипт з аудіо за допомогою P2FA, поєднуючи фонем на склади та об'єднуючи короткі склади. Сегментуйте лялькове відео на послідовність візуальних складів, використовуючи методи на основі зовнішнього вигляду або аудіо. У методі, що базується на зовнішньому вигляді, кадри класифікуються на три категорії: з відкритим ротом, із закритим ротом та недійсні, за допомогою мережі, що базується на попередньо навченій мережі Google, або вручну, або обома способами. У той же час в методі, що базується на аудіо, візуальні склади

					БР.ІІІ - 85.00.00.000 ПЗ	Арк.
						48
Змн.	Арк.	№ докум.	Підпис	Дата		

відповідають першому кроку. Що стосується вирівнювання аудіоскладів з візуальними складами, три основні принципи: 1) тиша відповідає тиші; 2) нетиша відповідає нетиші; 3) довжини складів є максимально подібними. Потім вирішіть варіант динамічного деформування часу. На останньому кроці пересинхронізуйте аудіосклади за допомогою додавання Waveform Similitiy Overlap, пересинхронізуйте візуальні склади за допомогою семплінгу найближчих сусідів або інтерполяції оптичного потоку та пересинхронізуйте аудіо та відео разом, встановивши нову довжину як їхнє геометричне середнє. У цій роботі декларуються їхні результати в додаткових матеріалах, але я не зміг їх знайти. Зрештою, ляльковий дубляж набагато простіший, ніж дубляж людиною.

Відновлення відео також поділяється на спеціальні ефекти. Використано нелокальні регуляризаційні члени як першого, так і другого порядку для відновлення відео низької якості. Запропоновано багатоплощинну авторегресивну модель для використання кореляції в міжвимірних площинах групи подібних ділянок сусідніх кадрів, а спільний багатоплощинний AR та алгоритм на основі низького рангу реконструює групу.

2.3 Набори даних, метрики оцінювання та методи тестування якості відеомонтажу

Стандартні набори даних значно звільняють дослідників від трудомісткого збору даних та очищення робочих місць. Це економить час і кошти, а також забезпечує орієнтир для алгоритмів цієї області. Тут ми сортуємо 29 наборів даних з двох основних джерел: проблеми та дослідники. Sports-360 складається з 342 відео 360°. Вони належать до п'яти спортивних категорій: баскетбол, паркур, BMX, скейтбординг та танці. Відео Pano2Vid [14] зібрані з Youtube. Створюють LEDOV, що складається з 32 фіксацій суб'єктів на 538 відео з роздільною здатністю щонайменше 720P у 158 підкатегоріях. Створюють корпус зображень з 5 простими підписами. Swedish leaf, KTH IDOL, категорія сцени з 15 класів, спортивна подія з 8 класів та розпізнавання сцени в приміщенні з 67 класів – це

					БР.ІІІ - 85.00.00.000 ПЗ	Арк.
						49
Змн.	Арк.	№ докум.	Підпис	Дата		

набори даних зображень для вилучення ознак. Створено набір даних споживчого відео Kodak з 25 концепцій. Відео надійшли від користувачів та Youtube, а концептуальні позначки доповнені вручну. Концепції включають дії (танці, спів), події (весілля, день народження тощо), сцену, об'єкт, людей та звук. Вибрано 100 відеокліпів з набору даних споживчого відео Kodak та додали до них ключові кадри. Створили набір даних панорамних зображень 360 ° у 26 категоріях місць, щоб навчити метод опорних векторів вивчати категорію місця та ракурс сцени зображень, а також симетрію об'єктів.

Метрики оцінювання. Метрики оцінювання відіграють важливу роль у розробці алгоритмів, оскільки вони надають зворотний зв'язок та виявляють недоліки тестового рішення. Існує кілька метрик, які зазвичай використовуються у сфері відеомонтажу, таких як ступінь його автоматизації, швидкість, вимоги до пам'яті. Точність, повнота та F-міра (наприклад, F1-оцінка), акуратність, CC-оцінка є популярними метриками [7]. Визначення F1-оцінки таке:

$$F1 - score = \frac{(2 \times Precision \times Recall)}{(Precision + Recall)} \quad (2.1)$$

Досі важко оцінити якість згенерованих відео, як обговорювалося вище, красу та логіку важко запрограмувати. Незважаючи на це, дослідники пропонують два рішення. Одне з них – дослідження користувачів. Відгуки користувачів є основним показником. Збирають дані про візуальну помітність людей під час перегляду вихідних відео як показник, що відображає, наскільки добре працює запропонований алгоритм.

Інше рішення полягає у визначенні метрик відповідно до конкретних експериментів. Чжан визначає точність (P) та повноту (R) як рівняння (2) для відеозведення:

$$P = \frac{\text{overlapped duration of A and B}}{\text{duration of A,R}} = \frac{\text{overlapped duration of A and B}}{\text{duration of B}} \quad (2.2)$$

					БР.ІІІ - 85.00.00.000 ПЗ	Арк.
						50
Змн.	Арк.	№ докум.	Підпис	Дата		

де A , B – відеокліпи. Визначають два класи метрик. Одна базується на HumanCam і включає розрізняльність, подібність до HumanCam та переносимість. Інша – це метрики на основі HumanEdit, що складаються з середньої косинусної подібності, об'єднання кадрів та середнього перекриття. Запропоновано *синхронну значущість*, яка вимірює синхронізацію візуальних та звукових подій. Потапов пропонує метод оцінки важливості відеосегмента, враховуючи, чи містить сегмент докази заданої категорії подій. Використовують перетин над об'єднанням (IOU) та середню точність (mAP) для оцінки часової сегментації:

$$IOU = \frac{\frac{1}{N} \sum_{i=1}^N (\tau_i^* \cap \tau_i')}{(\tau_i^* \cup \tau_i')} \quad (2.3)$$

де N – кількість сегментів, m^* – сегмент істинного рівня, а t' – передбачуваний сегмент. Для адаптації до неконтрольованих алгоритмів використовує міру подібності кластерів, яка надає істинний рівень рівня методом перебору. Пропонують метрику для вимірювання якості перенацілювання зображень, яка базується на SIFT.

Хоча ми зосереджуємося на відеомонтажі, редагування зображень все ще є цінною галуззю досліджень, яку варто вивчити, оскільки алгоритми, розроблені для зображень, можна було б перенести у відеосферу. Тут ми представимо останні дослідження в галузі редагування зображень, але не розглядатимемо алгоритми, які генерують зображення з нуля, такі як рендеринг.

Редагування зображень



Рисунок 2.13 - Ліворуч оригінальне чорно-біле зображення; Праворуч кольоровий результат.

					БР.ІІІ - 85.00.00.000 ПЗ	Арк.
						51
Змн.	Арк.	№ докум.	Підпис	Дата		

Покращення якості відео є важливою проблемою для багатьох зображень, а відео мають довгу історію. Надроздільна здатність - це не просто масштабування зображень шляхом інтерполяції, а акцент на покращенні їхньої якості. Це невизначена обернена задача. Використовують глибоку CNN для вивчення відображення зображення з низькою роздільною здатністю на зображення з високою роздільною здатністю. Пропонують NEGAN для попереднього навчання з шумом, який усуває шум, додає фарби та розфарбовує старі фотографії.

Застосовано різьблення швів на карті ознак вхідного зображення, а потім реконструюють отримане зображення за допомогою CNN. Перенацілювання пари стереозображень повинно зберегти їхні пари пікселів та стереоструктуру незмінними шляхом спільної оптимізації. Використовують різьблення швів та деформацію для перенацілювання пари стереоскопічних зображень, використовуючи їхню консистенцію диспаритету. Розфарбовування (як показано на рисунку 2.14) та передача стилю є поширеними галузями. Спочатку досліджують розфарбовування за допомогою CNN. Пропонують автоматичне рішення для розфарбовування, яке розглядає його як завдання класифікації та намагається збільшити різноманітність результуючих кольорів. Деякі роботи потребують вхідних даних від користувачів. Наприклад, розфарбовано зображення у градаціях сірого за допомогою CNN, навченої на великих наборах даних та розріджених підказках від користувачів. На відміну від попередніх методів розфарбовування, які зосереджені на всьому зображенні, пропонують об'єднати ознаки рівня об'єкта та рівня зображення, отримані з двох моделей, для визначення кінцевого кольору кожного об'єкта. Об'єкти виявляються за допомогою Mask R-CNN та обрізаються. Модель екземпляра, моделі рівня зображення та модуль об'єднання навчаються у три кроки.. Перенесення стилю - це видалення текстури з домену вихідного зображення та її перенесення в цільовий домен зображення за допомогою глибокої нейронної мережі. Представлено новий модуль Ghost в архітектурі GANILLA для вивчення та перенесення стилів зображень.

					БР.ІІІ - 85.00.00.000 ПЗ	Арк.
						52
Змн.	Арк.	№ докум.	Підпис	Дата		



Рисунок 2.14 – Інструмент м'які ножиці

Запропоновано інтерактивний інструмент для вилучення альфа-матових елементів об'єктів переднього плану, а саме «М'які ножиці», який поєднує в собі поступову оцінку матових елементів, поступову оцінку кольору переднього плану, інтелектуальний інтерфейс користувача та надійний алгоритм матування, щоб він міг працювати в режимі реального часу та ефективно. У той час захоплюють кілька фотопотоків однієї теми, вирівнюють їх за схожістю та спільно сегментують спільні області вирівняних зображень на основі графа зображень. Ця робота спрямована на пошук спільних закономірностей серед моря веб-зображень.

Щоб уникнути повторюваних зусиль, пропонують нову систему для створення навчальних посібників з маніпуляцій з фотографіями та макросів, залежних від контенту, шляхом запису демонстрації. Вона складається з пристрою для запису демонстрацій, мітки зображень та генератора навчальних посібників. Записувач демонстрацій записує всі зміни в інтерфейсі та результуючі зміни в стані програми, і вони будуть згруповані відповідно до параметрів та операцій пізніше. Мітка зображень використовує існуючі методи розпізнавання на основі комп'ютерного зору для позначення семантично важливих областей на зображеннях. Він генерує текстові описи у фіксованому стилі зі згрупованими змінами, анотаціями знімків екрана та генерацією тексту відповідно до певних рекомендацій, таких як покрокове, лаконічне, комбінація тексту та зображень та макет на основі сітки.

Деякі роботи роблять крок до розуміння зображень. Наприклад, для зображення з кількома простими підписами Ходош намагається визначити кореляцію сутності.

					БР.ІІІ - 85.00.00.000 ПЗ	Арк.
						53
Змн.	Арк.	№ докум.	Підпис	Дата		

Відеомонтаж розпочався в останні 20 років і швидко розвивається завдяки комп'ютерному зору, обробці природної мови та іншим дисциплінам. Як бачимо, системи відеомонтажу розвиваються від простих до складних, від одного вхідного сигналу до кількох, від одного елемента до кількох. Діапазон доменів вхідного відео широкий: від спортивної гри до спостереження, від фільмів до коротких відео для соціальних платформ. Однак, відеомонтаж досі не реалізує це повністю автоматично. Зараз алгоритми монтажу значною мірою залежать від ідеальних вхідних даних і навіть попередньо обробляють дані людьми. Ознаки, отримані з аудіо, зображення та часової залежності послідовних кадрів і субтитрів, нелегко зрозуміти комп'ютерам або недостатньо для алгоритмів. Крім того, розробка логічно узгоджених планів монтажу дуже складна. Створення дивовижного змонтованого відео сьогодні ще більше неможливо. Ми сподіваємося, що це опитування зможе допомогти дослідникам і сприяти розвитку відеомонтажу

2.4 Висновки по розділу

У другому розділі було розглянуто методи та інструменти інтелектуального монтажу відео. Досліджено використання методів машинного навчання та згорткових нейронних мереж для аналізу відеоданих, автоматичного розпізнавання сцен, об'єктів та подій. Проведено аналіз сучасних інструментів автоматичного редагування, генерації та маніпуляції відеоконтентом, що забезпечують оптимізацію процесів монтажу та підвищення якості результатів. Також розглянуто набори даних, метрики оцінювання та методи тестування якості відеомонтажу, які дозволяють визначити ефективність алгоритмів штучного інтелекту. Встановлено, що сучасні інтелектуальні технології здатні значно скоротити витрати часу на обробку відео та забезпечити стабільну якість кінцевого продукту.

					БР.ІІІ - 85.00.00.000 ПЗ	Арк.
						54
Змн.	Арк.	№ докум.	Підпис	Дата		

РОЗДІЛ 3. РЕАЛІЗАЦІЯ СИСТЕМИ АВТОМАТИЧНОГО МОНТАЖУ ТА ОЦІНКА ЇЇ ЕФЕКТИВНОСТІ

3.1 Інтелектуальна автоматизація відеовиробництва та сценарії використання системи

У цьому підрозділі ми розглянемо дослідження, що стосуються керування та контролю камери, візуального аналізу, кінематографічної допомоги та оптимізації робочого процесу.

Кінематографічна мова використовується для передачі стану виробництва та відповідних процесів. Вона дозволяє виробничому персоналу ефективно спілкуватися та узгоджено реагувати на непередбачені ситуації. Кінематографічна теорія часто формалізується для таких завдань, як автоматизоване керування камерою та аналіз сцени.

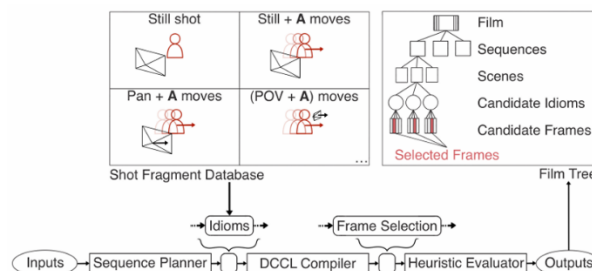


Рисунок 3.1 - Візуалізація DCCL для автоматизованого керування камерою.

Триетапний конвеєр використовує ідіоми, визначені станами камери та актора, як вхідні дані для компілятора DCCL. Вибрані кадри оцінюються відносно «дерева плівки».

Кінцеві автомати використовуються для моделювання кінематографічних кадрів для планування зйомки камерою у віртуальному середовищі. У цьому контексті ми обговорюємо сучасні підходи до автоматизованого керування камерою, які вводять або покладаються на формальну кінематографічну мову для логічного висновку; ми називаємо це *ідіоматичним захопленням*.

Евристичний підхід до керування камерою був заснований на алгоритмі планування причинно-наслідкових зв'язків з декомпозицією та частковим порядком. Це формалізує ідіоматичне захоплення, пов'язуючи наближене представлення сцени із заздалегідь визначеним набором відповідей керування камерою. Алгоритм вирішення конфліктів ранжує відомі дії за допомогою умовних операторів, щоб визначити оптимальну дію та відповідну тривалість. Щоб виразити ідіоми, виокремлюють набір із чотирьох вимог для опосередкування контекстуальних відмінностей між різними постановками: (1) Представлення історії, (2) Представлення реального світу, (3) Риторична узгодженість та (4) Часова узгодженість. Перші два пункти описують фізичні та контекстуальні ландшафти набору, тому необхідно мати як геометричні (або візуальні), так і семантичні представлення постановки. Третій пункт виражає необхідність риторичної структури для забезпечення рішучого виконання дій, наприклад, ієрархічна модель для вибору типів кадрів. Останній пункт визначає узгодженість щодо часового аспекту зйомок.

Аналогічно, декларативна мова керування камерою (DCCL) використовує евристичне дерево рішень для ідіоматичного захоплення рис. 3.2. Цей метод встановлює ієрархічну структуру для автоматизованого композиційного кадру та керування камерою шляхом розбиття сцен на послідовності кадрів, що відповідають певним ідіомам. Тут взаємозв'язок послідовних послідовностей кадрів залежить від часових зв'язків між ідіомами. Потім евристичний оцінювач вибирає дію-кандидат, оцінюючи можливі відповіді та оцінюючи рішення за допомогою дерева рішень.

Ідіоматичні мовні структури також використовувалися для позначення кадрів та оцінювання. Зовсім недавно була запропонована мова розкадровки прози як мова високого рівня для опису кадрів відносно візуальної композиції та ідіоматичного керування камерою. Це продемонстровано у фільмі Альфреда Гічкока «На північ через північний захід», де підказка описує не лише тип кадру та композиції переднього, середнього та заднього планів (окремо), але й тип переходу. Оскільки PSL використовує дерева рішень І/АБО, її можна легко

					БР.ІІІ - 85.00.00.000 ПЗ	Арк.
						56
Змн.	Арк.	№ докум.	Підпис	Дата		

налаштувати для додавання/зміни позначок. Хоча автори не тестували це на автоматизованій схемі позначення, є підстави для такого дослідження, оскільки мови на основі ідіом можуть ігнорувати нещодавно прийняті ідіоми та іноземні практики захоплення.

Вбудовані обмежені шаблони (ЕСР) – це спеціалізована мова запитів, яка враховує фізичні кінематографічні обмеження. На відміну від PSL, ЕСР є більш повною у своєму описі – позначає монтажі, а не окремі кадри. Мітки призначаються за допомогою таких дескрипторів:

1. Обмеження кадрування: розмір акторів та об'єктів, кут зйомки, область зйомки актора та розташування об'єктів і рух камери
2. Зв'язок між кадрами: розмір, кут та область відносно послідовності кадрів
3. Підпослідовності: локальне групування пов'язаних послідовностей знімків

Запропоновано ЕСР разом з автоматизованим алгоритмом пошуку для оптимального опису послідовності кадрів. Метод пошуку розділено на два етапи: (1) створення кешу дійсних рішень та (2) застосування *подвійного рекурсивного пошуку в глибину* для вибору оптимального опису з набору дійсних рішень. На етапі (1) дійсні рішення поділяються на три набори: (i) кадри, що задовольняють обмеження, попередньо визначені ЕСР, (ii) пари кадрів, що задовольняють реляційне обмеження ЕСР, та (iii) послідовності кадрів. На другому етапі процес пошуку повторюється по всіх кадрах у кожній послідовності, де наступні кадри перевіряються як частина локальної або глобальної послідовності. Поряд з технічною реалізацією ЕСР, запропонували плани інтеграції з Unity 3-D для застосування ЕСР до своїх монтажів; що ми розглядаємо як позитивний внесок у виробничий конвеєр.

Крім того, мови, створені для автоматизованого керування камерою, не легко порівнювати. Вибір структури мови може суттєво відрізнитися залежно від застосування. Хоча наразі не існує загального стандарту для кінематографії, ми спостерігаємо інтерес з боку Movie Labs, які прагнуть визначити набір обчислювальних лінгвістичних структур як стандарт для хмарних лабораторій

					БР.ІІІ - 85.00.00.000 ПЗ	Арк.
						57
Змн.	Арк.	№ докум.	Підпис	Дата		

оптимізації робочих процесів. Перехід до хмарних обчислень свідчить про іншу парадигму виробництва. Примітно, що оскільки ресурси та інструменти зазвичай зберігаються або виконуються офлайн, зміни у спільних ресурсах може бути важко контролювати на одній платформі. Цьому можуть сприяти підходи до управління файлами, які імітують, наприклад, практику спільного використання коду. Однак це може бути важко масштабувати для виробництв, що потребують великої кількості ресурсів, або виробництв, які залучають зовнішніх міжнародних співробітників, які можуть не знати стандартних практик.

Помічники з питань директив

Перетворення більшості кінематографічних концепцій на контролери є складним завданням. Поняття композиції кадрів, типів кадрів та переходів між кадрами пов'язані з реальними проблемами, такими як вартість виробництва, фізика та топологія знімального майданчика. Таким чином, для полегшення цих виробничих проблем були введені помічники-директори штучного інтелекту (DA). Сучасні дослідження в галузі ІК спрямовані на такі завдання, як складання списків кадрів, планів зйомок, оптимізація розміщення камери та керування роботизованими установками камер.

Існує тенденція використовувати семантичні представлення робочого середовища для провокування DA. Наприклад, за допомогою DCCL, демонструють евристичний підхід, який спирається на ідіоматичну практику для прийняття рішень щодо складу пострілу. Аналогічно, обговорюють інший евристичний підхід, використовуючи набір скінченних автоматів для обробки ідіоматичних дій камери. Обидва ці методи використовують віртуальні симуляції з подіями, керованими користувачем, щоб продемонструвати здатність своїх інструментів отримувати ідіоматичні відповіді за короткий проміжок часу.

Існують більш нові підходи до проблеми отримання списку кадрів, які використовують архітектуру, що імітує чотири критичні ролі зйомки: (1) сценариста, який спостерігає за контекстом поточної сцени та надсилає інформацію до (2), (2) організатора екрану, який налаштовує розміщення акторів та об'єктів для драматичного ефекту та передає цю інформацію до (3) та (4), (3) режисера, який

					БР.ІІІ - 85.00.00.000 ПЗ	Арк.
						58
Змн.	Арк.	№ докум.	Підпис	Дата		

витагує важливу інформацію та використовує кілька машин опорних векторів (SVM) для прийняття рішень щодо вибраних кадрів, та (4) оператора, який дотримується ідіом під час зйомки. Цей метод об'єднує три двовимірні матриці ознак у багатовимірний простір ознак, вибираючи найбільш релевантну SVM з попередньо навченого набору для класифікації оптимального положення кадру та кута огляду. SVM працюють, використовуючи карти ядра для перерозподілу всіх ознак у багатовимірну матрицю. Це робиться за припущення, що класи ознак не є лінійно роздільними, доки вони не об'єднані у багатовимірну матрицю. На практиці, нормалізовані позиційні значення представляють особливості середовища, тоді як сприйнятий емоційний стан кожного актора представлений рисами акторів, а актор, на якому зосереджена основна увага, представлений в межах рис сцени. Вони використовуються для вибору правильної SVM для оцінки оптимального розташування та кута камери, враховуючи контекст сцени. Оскільки SVM можна вважати застарілими відповідно до сучасної літератури, ймовірно, існують більш підходящі підходи до моделювання середовища, акторів та характеристик сцени для класифікації оптимального кута та розташування камери з попередньо визначеного набору. Наприклад, з достатньо великим набором даних глибокі нейронні мережі будуть краще узагальнювати.

Як альтернатива, CamDroid - це добре відома архітектура керування камерою на основі станів, як показано на рисунку. 3.2.

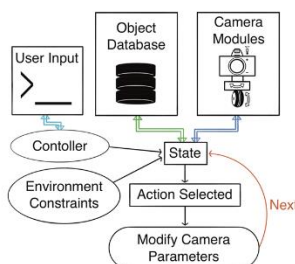


Рисунок 3.2 - Візуалізація CamDroid для керування камерою

Цей метод використовує інформацію про локальне реальне середовище для визначення послідовності дій, які необхідно виконати для оптимального

захоплення. Завдяки цьому різні візуальні черги або підказки користувача будуть запускати послідовність дій камери. Умовні функціональні фреймворки, такі як CamDroid, є класичним способом автоматизованого керування камерою та є потужними для захоплення дій у режимі реального часу, що залежать від часу.

Архітектура на основі станів, яка спирається на вхідні дані користувача та обмеження середовища для зміни параметрів камери для послідовності дій. Незважаючи на нестачу досліджень цифрової анімації (DA), нещодавно з'явилися такі роботи, які досліджують вибір кадрів за допомогою багаторакурсної камери для зйомки театральних вистав. Автори записують вистави з кількох ракурсів у форматі 4K, і шляхом обрізання кадрів високої роздільної здатності на кадри з нижчою роздільною здатністю виводиться набір дій камери. Згодом для кожного актора в наборі обрізаних відео оцінюються пози скелета та обличчя. Вони використовуються в автоматизованому сценарії монтажу, де, наприклад, рухомі губи можуть бути використані для визначення актора як головного фокуса сцени, тому відповідний кліп (тобто вибраний ракурс та дія камери) вибираються для заданого періоду часу.

Основні напрямки дослідження DA: ідіоматична мова, класифікація та підходи на основі станів, не є новими за сучасними стандартами досліджень ШІ, але тим не менш здатні досягати надійних ідіоматичних дій камери у відповідь на введення користувача. Однак, оскільки обговорення паралельних методів обмежене нестачею нещодавніх робіт, ми ще не бачили перспективних робіт, які відходять від офлайн-автоматизована DA та в напрямку прямого ефіру, наприклад, у спортивних трансляціях. Незважаючи на це, численні суміжні галузі підходять до подібних парадигм. Наприклад, обговорюють найсучасніші підходи до взаємодії з 3D-візуалізаціями. Автори окреслюють численні області для вдосконалення, зокрема, щодо інструментів та метрик помилок, що використовуються для розуміння та оцінки взаємодії людини з комп'ютером. Оцінка скелета та розпізнавання облич, недостатні для отримання контексту сцени. Аналогічно, обговорюється застосування, пов'язані з керуванням камерою у віртуальних середовищах.

					БР.ІІІ - 85.00.00.000 ПЗ	Арк.
						60
Змн.	Арк.	№ докум.	Підпис	Дата		

Оптимізація робочого процесу та автоматизована композиція кадрів

Робочі процеси передпродакшну та виробництва суттєво відрізняються через бюджетні обмеження, терміни виконання та творчі цілі. Незважаючи на це, є аспекти кіновиробництва, які залишаються незмінними, включаючи дизайн декорацій, процеси постановки об'єктів та фону, а також зйомку та монтаж фільмів. У галузі помітними зусиллями Movie Labs ілюструє потенціал хмарної платформи для оптимізації виробництва, що передбачає нові структури обчислювальної мови, методи хмарної безпеки та робочі процеси для співпраці. Movie Labs підготували набір офіційних документів Labs, у яких детально описані їхні амбіції на наступне десятиліття.

Machinima Production (MP) для ІС використовує інтелектуальне написання сценаріїв для створення поз камери для кадрів (тестування на віртуальній сцені). Замість автоматизації всього процесу виробництва, ми бачимо потенціал в адаптації інструментів MP для задоволення загального обсягу робочих процесів. Наприклад, система GLAMOUR використовує генерацію природної мови на основі кінематографічних ідіом, для створення кінематографічної композиції кадрів з вичерпними описами сцени зі статичних зображень. Anantrasirichai та Bull, таким чином, коли анімовані актори керуються ІІІ під час сцени, визначається оптимальний вибір камери та переходу. Як обговорювалося, рівень потенційних стимулів, які можна відчувати від фільму, набагато перевершує складність уваги аудиторії. GLAMOUR можна розширити, включивши інші процеси, що беруть участь в автоматизованій композиції кадру, такі як керування кернелізованим фільтром (KCF), використовуване для виявлення акторів та кадрування в автоматизованому отриманні аерозйомки. Додатково корисним є те, що кінематографічна передача руху використовує 3D-зображення, згенеровані з реальних зображень та зіставлені з послідовністю кадрів із сцени в заданому фільмі. Це можна поєднати з оптимізацією уваги для покращення вибраного ідіоматичного кадру для передачі, з чого ми могли б отримати більш змістовні пози та рухи камери.

Напівавтоматизований метод, був розроблений для обробки постановки об'єктів, автоматичного позиціонування камери та композиції кадру за

					БР.ІІІ - 85.00.00.000 ПЗ	Арк.
						61
Змн.	Арк.	№ докум.	Підпис	Дата		

допомогою анотованого сценарію. Фреймворк, показаний на рисунку 3.3 оптимізує параметри камери залежно на основі згенерованої послідовності дій для кожного присутнього персонажа. Список дій перетворюється на сценічну виставу, де кожна дія відповідає руху у сцені. Це автоматизує процес планування сцени. Під час кроку оптимізації камери до вистави спільно застосовуються моделі «естетики» та «точності», щоб визначити оптимальну камеру для використання на кожному кроці часу. Естетична модель аналізує шість факторів для планування камери: видимість персонажа, дії персонажа, конфігурація камери, безперервність екрану, безперервність руху та тривалість кадру. Модель точності спочатку припускає, що математична модель може наближено визначити зв'язок між сценарієм та згенерованим відео. Для цього модель використовує модель вбудовування глобального вектора для слів (GloVe) для генерації тексту зі згенерованого відео, а потім аналізує подібність між згенерованим сценарієм та цільовим сценарієм. Вбудовування GloVe + це векторне представлення слів, які навчаються з використанням глобальної спільної появи пар/груп слів. Таким чином, підструктури векторного простору можуть визначати синоніми та канонічні структури.

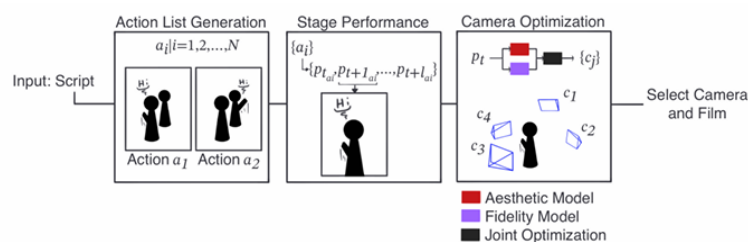


Рисунок 3.3 - Text2animation для вибору кадру

Загалом, обидва ці методи створюють наївні підходи до оптимізації робочого процесу, обмежені як кінематографічним розумінням, так і сучасними технологічними можливостями. Наприклад, суть механізму оцінювання залежить не лише від припущення, що можна побудувати модель точності, але й від можливості перевести відео в текст у спосіб, придатний для оцінювання. Крім того,

потрібен удосконалений метод оцінки таких моделей, оскільки ми маємо сумніви щодо достовірності текстового тестування для кінематографічних моделей.

Враховуючи, що робочі процеси часто фрагментовані через залежність від зовнішніх співавторів, дослідження, зосереджене на підтримці співпраці, видається доцільним. Наприклад, моделі можна доручити виконувати роль адміністратора спільного ресурсу – вирішувати, чи слід приймати зафіксовані зміни, враховуючи деякі обмеження, пов'язані з камерою. Спираючись на ICSVFX, камери, такі як AxiBo PT 4. Слайдери з керування рухом використовують хмарний робочий процес Unreal Engine та безпосередньо відповідають за парадигму спільних ресурсів. Використання таких інструментів можна поширити на дослідження, що проектують динамічні ресурси та перевіряють внутрішньокамерну композицію з урахуванням цільової області на основі функцій .

Виклики та майбутня робота

Одна з проблем впровадження нового III у виробничий процес полягає в тому, як його тестувати та перевіряти. На жаль, реальні виробничі середовища не тільки складні для доступу, але й важко контролювати. Тому більшість пропозицій надають перевагу тестуванню за допомогою симуляції через віртуальні середовища. Очевидно, що існує невідповідність між напівінформованою симуляцією та реальним випадком використання. У майбутньому нам потрібно визнати значну користь експериментальних постановок. Експериментальні постановки можуть стати платформою для тестування інструментів ІС, чи то для виконання рутинних завдань, щоб митці могли зосередитися на творчих експериментах, чи то для явної підтримки власного творчого самовираження. Приклади цього вже існують, такі як Люкс Аетерна, який продюсував науково-фантастичний фільм «*RENO*» в рамках програми «MyWorld Strength in Places» як засіб дослідження наборів інструментів III у візуальних ефектах. Ця постановка RENO була використана для дослідження використання III в постпродакшні, проте таку роботу можна було б підняти до завдань, орієнтованих на виробництво. Як варіант, ми можемо уявити собі використання 3D-реконструкцій реальних сцен, щоб досягти більш надійних тестів у реальних умовах. Це забезпечить більш

					БР.ІІІ - 85.00.00.000 ПЗ	Арк.
						63
Змн.	Арк.	№ докум.	Підпис	Дата		

зрозумілі результати для кінематографістів, які планують використовувати ІС для керування камерами в реальних умовах, а також для тих, хто бажає знімати кінематографічні кадри за допомогою 3D-реконструкцій.

Щодо підтримки кінематографічної мови, існує багато міркувань. Наприклад, застосування мови може бути корисним для семантичного аналізу/генерації сценаріїв, але вона також може виступати як форма комунікації між виробничим персоналом та інтерфейсами штучного інтелекту. Вважається PSL особливо цікавою, оскільки вона представлена як мова середнього рівня. Це корисно, оскільки зменшує залежність від вирішення проблем обробки природної мови (NLP), що може бути проблематичним при зіткненні зі складними описами кадрів або при описі суб'єктивних кінематографічних спостережень. Однак, ми повинні визнати прогрес, досягнутий у дослідженнях NLP та семантичного аналізу за останнє десятиліття, що може призвести до кращих пропозицій щодо вивчення лінгвістичних структур для кінематографічного виробництва. Наприклад, для семантичного аналізу кадрів можна запропонувати комплексний набір даних та метод позначення кадрів за допомогою PSL, щоб зменшити складність структур природної мови для вбудовування слів. В іншому випадку, дослідження DA забезпечують платформу для використання низки контролерів штучного інтелекту. Їх можна адаптувати для конкретних цілей, а архітектури часто реалізуються як модульні конвеєри. Однак оцінка продуктивності в реальних умовах обмежена доступом до реальних наборів, виробничого персоналу та обмеженою кількістю наборів даних та моделей для еталонів. Це вказує на те, що стан досліджень DA ще має значний шлях розвитку, перш ніж рішення можна буде застосувати в комерційних масштабах.

Щодо стратегій оптимізації робочого процесу, ми повинні дослідити соціальну користь від досліджень у сфері маніпуляційного моделювання. Як обговорювалося, метою ІС є краще сприяння художньому кіновиробництву. Однак автоматизація всього процесу може залишити мало місця для творчого внеску. МР можна використовувати для підтримки розкадровки та інших завдань передвиробництва. Розвиваючи це далі, можна сформулювати розкадровку як набір

					БР.ІІІ - 85.00.00.000 ПЗ	Арк.
						64
Змн.	Арк.	№ докум.	Підпис	Дата		

анімованих кліпів для прийняття рішень щодо композиції кадру. Таким чином, майбутня робота може бути спрямована на оптимізацію композиції кадру, наприклад, для передачі стилю або оптимізації уваги.

3.2 Реалізація модулів віртуального та автоматизованого відеомонтажу

У цьому підрозділі ми зосередимося на віртуальному виробництві з використанням реальних камер. Це важливо, оскільки відрізняє його від класичного віртуального виробництва, що використовується в анімації чи іграх. Існують дві помітні відмінності: по-перше, композиція віртуальних сцен для анімації є менш складною, ніж композиція віртуальних ресурсів у реальній сцені для виробництва інтегрованих схем, оскільки нюанси освітлення, кольору та перспективи між реальними та віртуальними ресурсами потребують уваги. По-друге, реальні камери обмежені фізичними та фінансовими обмеженнями, такими як топологія набору, додаткове обладнання та технічні навички для досягнення певних рухів камери. Ці питання легко вирішити за допомогою віртуальної камери, яка має 6 ступенів свободи (6DoF).

Ці відмінності лежать в основі проблем досліджень, пов'язаних з віртуальним виробництвом для ІС. Отже, в цьому розділі ми розглядаємо дослідження, що стосуються ICVFX та світлодіодних томів, торкаючись робіт, що досліджують перефарбовування та освітлення на основі зображень (IBL).

ICVFX та гучність світлодіодів

Численні наукові роботи присвячені аналізу візуальних ефектів (VFX) та їх застосування в камері в галузі кінематографії. Однак відповідні варіанти використання ІС чітко визначені. З суто кінематографічної точки зору, ICVFX зазвичай включає CGI та/або методи композитингу, які виконуються в режимі реального часу, надаючи кінематографістам пряму трансляцію того, як віртуальні ефекти виглядатимуть відносно реальної сцени. Це може підтримувати виробництво різними способами, наприклад, забезпечуючи індикатор поганого освітлення або

					БР.ІІІ - 85.00.00.000 ПЗ	Арк.
						65
Змн.	Арк.	№ докум.	Підпис	Дата		

як спосіб попередньої візуалізації композитингу та VFX для полегшення проблем постпродакшну. Тут технічна практика в основному включає розрізнення елементів переднього та заднього плану, а також модифікацію освітлення та кольору. Коротко представлено *хромакей* та *ротоскопінг* як парадигми, зосереджуючись на першому; як показано на рисунку. 3.5(a). У той же час, представляють, як IBL, добре зрозумілий метод у VFX на рисунку 3.4, перетворився на робочий процес освітлення, що оточує світлодіодні відеоекрани для керування переосвітленням та розфарбовуванням у камері, рисунок 3.5(b). Це натякає на два різні випадки використання досліджень ІС:

1. Автоматична сегментація переднього плану: (i) з використанням та (ii) без кольорових екранів (тобто синіх/зелених екранів).
2. Автоматичне IBL та перефарбування. IBL відображає зображення середовища на 3D-примітив (наприклад, сферу). Середовище IBL розглядається як джерело світла, і за допомогою трасування променів модель композицується.

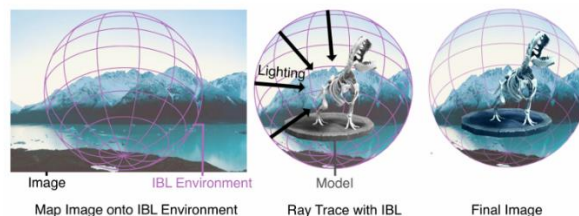


Рисунок 3.4 - Процес освітлення на основі зображення для композиції

Загальновідомо, що варіант використання (1) легко реалізується за допомогою кольорових екранів (1.i) та виділення кольору, що відповідає екрану. Однак це створює проблеми для сегментації зображень на основі кольору. Застосування відповідних методів сегментації зображень на основі кольору, орієнтованих на ширшу аудиторію штучного інтелекту. Автори обговорюють такі рішення, як порогове значення гістограми, використовуючи бінарні дерева для зберігання обсягів даних у 3D колірних просторах; Висвітлюють проблеми із затіненням та текстурованням. Найбільш очевидним випадком цього є інтерферуюче відбиття світла від кольорового екрана. Це призводить до специфічних підходів,

					БР.ІІІ - 85.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		66

наприклад, підходять до сегментації шкіри в реальному часі для відео, на відміну від змінного в часі освітлення. Оптимізовано марковську модель другого порядку для прогнозування гістограми кольору шкіри з часом. Крім того, хромакейінг обмежує можливість відображення кольорів, подібних до кольорів на хрома-екрані.

Зовсім недавно ми спостерігали підходи, що стосуються випадку (1.ii). Це популярна парадигма поза межами кінематографії, в рамках якої ми бачимо нові дослідження, такі як Segment Anything Model (SAM) – рішення для загальної сегментації зображень одним клацанням миші. Щодо кінематографії, ми знаходимо такі роботи, як Roto++ – інструмент ротоскопування, спрямований на повагу до вимог художників.

Roto++ покращує традиційні методи інтерполяції, поєднуючи відстеження зовнішнього вигляду в реальному часі та новий процес регуляризації многовидів форми (побудований на моделі латентних змінних гаусового процесу (GP-LVM)). Згодом, у послідовності кадрів метод (1) передбачає зміну многовиду форми та (2) визначає, який наступний ключовий кадр потрібно вручну обмежити за допомогою ротації.

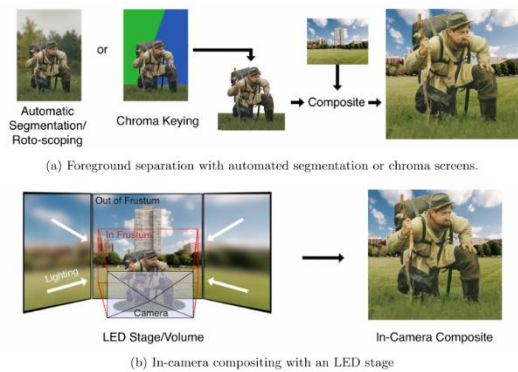


Рисунок 3.5 - Підходи до композиції в камері: використання автоматичної сегментації, хромакейінгу та світлодіодної сцени/об'єму

Щодо випадку (2); хоча IBL є класичною технікою 3D-рендерингу, її можна застосовувати до реальних сцен, де налаштування освітлення можна легко змінити. Обговорюють підходи, що реконструюють матрицю переносу світла

(LTM). Класифікують ці підходи на три категорії: (i) метод грубої сили, безпосереднє моделювання LTM, (ii) на основі розрідженості, моделюючи набір базисних функцій, припускаючи, що кожен рядок LTM можна лінійно апроксимувати, та (iii) на основі когерентності, аналізуючи поле когерентного відбиття для отримання LTM. Обмеженням цих методів є те, що вони вимагають отримання кількох зображень за різних умов освітлення, а це означає, що для відео цю проблему складніше вирішити.

Цікаво, що у випадках (1) та (2) акцент змістився в бік використання світлодіодних панелей для проектування віртуальних фонів як практичного рішення. У випадку (1) розділення переднього плану та фону спрощується шляхом заміни кольорових екранів віртуальними фонами, що відображаються в реальному часі на взаємопов'язаних світлодіодних панелях. Хоча це й дорожче та енергоємніше, це зменшує навантаження на постпродакшн та дозволяє уникнути необхідності ротоскопії. У випадку (2) результат не такий серйозний. Натомість, впровадження світлодіодних панелей пропонує нові можливості для автоматизованого калібрування освітлення, тепер включаючи світлодіодні панелі як джерело світла. Наприклад, розглянуто освітлення панелі як навколишнє світло, що дає приклад результату на рисунку 3.6.



Рисунок 3.6 - Результати зображень, зроблених за природного освітлення (ліворуч), та налаштування навколишньої світлодіодної панелі сцени/об'єму для освітлення в камері (праворуч)

Досягнуто цього, досліджуючи набір матриць попередньої корекції (ядер), застосованих до внутрішніх та зовнішніх усічених стінок світлодіодів. Потім, за допомогою окремої матриці пост-корекції, застосованої до зображення

					БР.ІІІ - 85.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		68

в камері, пікселі переналаштовуються на бажану/очікувану колірну схему. Матриці попередньої корекції розв'язуються шляхом порівняння та мінімізації різниці між очікуваними функціями спектральної чутливості випромінювання світлодіодів та результуючою функцією чутливості, захопленою камерою. Ця оптимізація виконується для заданої колірної діаграми. Тестування на різних колірних діаграмах показало майже оптимальні результати, хоча результуючі помилки обмежені розміром ядер до та після корекції. Висвітлюють проблеми, що призводять до зниження насиченості кольору шкіри і тканини; що пояснюється обмеженою кольоровою схемою світлодіодного панельного освітлення.

Аналогічно, розглядають корекцію динамічних змін відтінку в камері за допомогою усіченого коліщатка, хоча й використовуючи простіші колірні перетворення. Крім того, досліджують метод відтворення освітлення HDR-зображення (HDRI), переходячи від віртуального налаштування HDRI до налаштування гучності світлодіода. Це по суті інвертує класичну проблему IBL і вирішується шляхом розширення пікселів вище заданого порогу, щоб відповідати обмеженням на локальні середні значення пікселів, що відображаються на віртуальній світлодіодній стіні. Загалом, використання світлодіодних панелей у виробництві все ще є відносно новим. Окрім обговорення освітлення та колоризації, ландшафт ІС поза цією парадигмою не визначений.

Калібрування та локалізація камери. Метою калібрування та локалізації камери (або оцінки пози) є відображення 3D-світу на площині 2D-зображення. Це включає моделювання внутрішніх/інерційних параметрів камери та зовнішніх/зовнішніх параметрів з використанням реальних даних зображення. Ілюстрацію відповідних параметрів, що підлягають моделюванню, показано на рисунку 3.7. Внутрішні параметри визначають модель камери, фокусну відстань та спотворення об'єктива.

Зовнішні параметри моделюють матрицю перетворення камери для кожної пози, а також шлях для рухомих кадрів.

Це популярна проблема, яка стосується загального випадку використання однієї камери, а також конкретних випадків використання, таких як хірургічний

					БР.ІІІ - 85.00.00.000 ПЗ	Арк.
						69
Змн.	Арк.	№ докум.	Підпис	Дата		

моніторинг. Щодо ІК, опубліковані роботи стосуються кількох застосувань, включаючи аерофотозйомку, фотограмметрію, 3D-реконструкцію на основі зображень та підводна зйомка. Два практичні способи використання цього в кінематографії – автоматизоване керування камерою та справжня 3D-реконструкція.

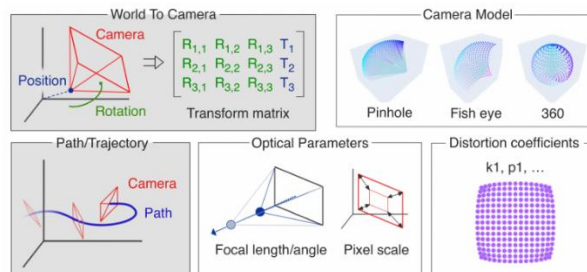


Рисунок 3.7 - Візуалізація параметрів, що беруть участь у калібруванні камери: Зовнішні параметри показані сірим кольором. Внутрішні параметри показані білим кольором.

Для автоматизованого керування камерою дослідницький ландшафт схиляється до контролю зовнішніх параметрів. Однак існують роботи, що поєднують зовнішнє та внутрішнє параметричне керування для задоволення потреб кінооператора. Наприклад, представлено CineMPC, який шукає оптимальну траєкторію та кут камери, використовуючи структуру прогнозованого моделювання керування (MPC). MPC досягає такого результату процесу, як дія камери, враховуючи майбутні моменти /горизонти часу та мінімізуючи витрати на вибір різних дій. CineMPC моделює скінченний горизонт, який постійно зміщується, доки всі дії не припиняться, та обмежує цільову функцію гнучкістю до різних конфігурацій камери та візуальної естетики. Автори розробляють математичні вирази для врахування композиції, глибини різкості та канонічних кадрів. Таким чином, наступні функції вартості можна вважати наївними наближеннями канонічного кінематографічного стилю.

Хоча існує багато робіт з калібрування камери для автоматизованого керування, можливості для 3D-реконструкції є більш плідними. Загальні підходи до калібрування описані, які моделюють внутрішні параметри спотворення для

камери-обскури. Тим не менш, не існує універсальної або гнучкої схеми автоматизації, яка б досягала цього. Ось чому ми досі бачимо розрізнені дослідження парадигм підмножин, які зосереджені на калібруванні функцій масштабування та фокусування для сценаріїв, де моделювання об'єктів з фіксованим фокусом та змінною фокусною відстанню є неможливим. Розглянуто покращення геометричної стабільності відео за допомогою SfM багаторакурсних стереозображень (MVS) для фотограмметричної реконструкції.

Цікаво, що існують роботи, пов'язані з моделюванням INR та NeRF, які розглядають всю проблему. Класична модель наївно використовує COLMAP, тоді як нещодавні роботи зосереджені на оптимізації суглобів.

Нейронні 3D-зображення динамічних сцен. Реконструкція реальних сцен як 3D-зображень пропонує безліч нових рішень існуючих кінематографічних проблем, таких як синтез нових зображень для отримання контенту, та підносить попередні парадигми, засновані на 2D, до 3D.

Існує два загальних формулювання цієї проблеми:

1. Монокулярні сцени, зняті однією камерою, що рухається навколо рухомого об'єкта
2. MVS-сцени, що містять одну дію, зняту за допомогою кількох камер.

Існує третє формулювання, сцени, спрямовані вперед, де одна камера обмежена однією площиною руху. Однак це слабо прийнято як універсальна парадигма. Наприклад, K-площини імітує загальний консенсус щодо використання нормалізованих координат пристрою (NDC) та стиснення сцени, що використовуються для «взлому» цієї проблеми. Однак це не дає можливості рендерингу динамічного відео з 6 ступенями свободи (6DoF). Для цього знадобиться галюцинація геометрії перешкод («позаду» сцени), що вводить абсолютно нову парадигму.

Методи, що відповідають загальним формулюванням, наведено в таблиці 3.1. Ці методи не є єдиними одночасними рішеннями, проте вони представляють широкий спектр доступних рішень .

					БР.ІП - 85.00.00.000 ПЗ	Арк.
						71
Змн.	Арк.	№ докум.	Підпис	Дата		

Огляд динамічних NeRF

Метод (Method)	Швидкість навчання (Training Speed)	Компактність (Compact)	Явність (Explicit)	Попередники (Ancestors)
D-NeRF	D	✓	✗	NeRF
Dy-NeRF	D	✓	✗	NeRF
V4D	h	✗	✓	-
NeRFPlayer	h	✗	✗	Instant-NGP, TensoRF
K-Planes	m/h	✓	✓	NeRF-W, Instant-NGP, DVGO
HyperReel	m/h	✓	✗	TensoRF
NeRF-DS	h	✗	✗	HyperNeRF
HexPlane	m/h	✗	✗	TensoRF
Tensor4D	h	✗	✗	NeRF-T, D-NeRF
DynIBar	D	✗	✗	NSFF, IBRNet

Існує дві риси, які наразі відрізняють дослідницькі підходи. Перша - це пропозиція нових просторово-часових представлень. Друга - це маніпулювання представленнями для покращення вивчення часових елементів.

Швидкість навчання вимірює час навчання, де D – дні, h – години, а m – хвилини. Це може залежати від сцени, тому m/h вказує, що швидкість навчання варіюється від хвилин до годин. Компактний параметр вказує, чи можна навчати моделі на одному графічному процесорі. Явний параметр вказує, де було використано явне представлення для підвищення швидкості виведення.

Однією з найперших і найлегших для розуміння пропозицій щодо нового просторово-часового представлення є D-NeRF. Цей метод моделює деформацію (зміну положення) об'ємів відносно часу. Хоча моделі деформації інтуїтивно зрозумілі, вони створюють проблеми, коли сцена/об'єкти не перебувають безперервно в кадрі. Натомість, такі методи, використовують низькорангові розклади для послідовного моделювання топології сцени, навіть коли об'єкти залишаються невидимими протягом кількох кадрів. Щоб подолати проблеми з великими, повільними для навчання динамічними представленнями, кращими є альтернативи

гауссового сплетінгу.

Ще одним цікавим рішенням проблеми об'ємної узгодженості є інтерполяція статичних полів ключовими кадрами. HyperReel ілюструє це, вивчаючи вектори зміщення та геометричні примітиви спільно вивченого статичного поля ключових кадрів. Подібно до принципів D-NeRF, це покращує часову якість представлення поля випромінювання.

Загалом, методи лише помірно здатні моделювати динамічні сцени з 6 ступенями свободи та зазвичай потребують додаткової модифікації для обробки нестандартних сцен, таких як сцени, спрямовані вперед. Незважаючи на це, ми все ще знаходимо роботи, що відповідають специфічним потребам виробництва, такі як редагувані NeRF та представлення GS. Для кінематографістів це означає бути терплячими, оскільки ми спостерігаємо зростання інтересу та внеску в цю галузь.

Нейронні 3D- зображення людей

Моделювання нежорсткої або деформованої геометрії, наприклад, людини, є класичною проблемою для графічних досліджень. Ця тема добре узгоджується з дослідженнями динамічних NeRF та розширює дослідження комп'ютерного зору, орієнтованого на людину, до тривимірного випадку.

Подібно до попереднього підрозділу, два загальні формулювання проблеми – це багаторакурсні та монокулярні сцени, які ми використовуємо для розмежування поточних дослідницьких цілей. Ми також можемо розрізнити роботу за її опорою на узагальнення людей та їхніх поз. HumanNeRF Zhao ілюструє обидва ці пункти як метод, що фокусується на парадигмі багаторакурсності, одночасно генеруючи узагальнювані людські пози. Автори пропонують використовувати NeRF для вивчення багатоособової лінійної моделі Скіннера (SMPL), тобто геометрії та зовнішнього вигляду актора. Ці ознаки вивчаються та використовуються для навчання нового нейронного поля змішування зовнішнього вигляду. Поле змішування зовнішнього вигляду уточнює деталі текстури, враховуючи колір вирівняних ознак із сусідніх ракурсів. Ця модель працює добре, проте вона вимагає ретельно розміщених камер та має труднощі з новими позами.

					БР.ІІІ - 85.00.00.000 ПЗ	Арк.
						73
Змн.	Арк.	№ докум.	Підпис	Дата		

MonoHuman долає це, використовуючи спільні двонаправлені модулі деформації, які розділяють деформацію вперед і назад на жорсткий скелетний рух і нежорсткий рух. Деформація вперед стосується перетворення з канонічного простору в унікальний простір спостереження, тоді як деформація назад досягає протилежного. Для керівництва навчанням новим позам, ознаки прямої відповідності у відомих ключових кадрах вибираються з банку спостережень, а візуальні ознаки оцінюються відносно ознак у новому просторі спостереження. Це покращує об'ємну узгодженість між різними просторами спостереження, а це означає, що послідовності дій можна відновлювати з більшою впевненістю. Крім того, метод менш вразливий до проблем із використанням монокулярного відео.

У кінематографічній практиці методи нейронного моделювання людини переживають обмеження доступності захоплення руху, які наразі вважаються золотим стандартом для систем MoCap. Автори моделі нейронного руху (NeMo), тестуючи запропоновану ними структуру на спортивних діях, взятих з набору даних Penn Action. Як і MonoHuman та HumanNeRF, NeMo узагальнює рухи за допомогою багатовидового відео. Хоча, примітно, це досягається шляхом введення відео однієї й тієї ж дії з різними умовами сцени, такими як різні актори та освітлення. Для обробки несинхронізованих дій вводиться фазова мережа як функція деформації на основі часу для вирівнювання поз у послідовності дій. Крім того, вбудовування специфічних для сцени функцій обробляє візуальні відмінності між сценами. Отже, фазові та екземплярні ознаки використовуються як вхідні дані для мережі NeMo, де виходи визначають кути з'єднання та трансляцію руху – що використовується для візуалізації моделі SMPL.

У випадку з ICVFX дослідження обмежені доступом до дорогих технологій, таких як камери та світлодіодні об'ємні моделі. Враховуючи новизну формату виробництва світлодіодних моделей, ми обов'язково побачимо більше досліджень, оскільки буде побудовано більше сцен. Представляючи коментарі численних керівників виробництва та фахівців, які висловлюють великі очікування щодо цієї технології.

Окрім поточних досліджень щодо покращення якості сценічного освітле-

					БР.ІІІ - 85.00.00.000 ПЗ	Арк.
						74
Змн.	Арк.	№ докум.	Підпис	Дата		

ння, існують інші перспективні напрямки. Світлодіодне об'ємне виробництво (або віртуальне сценічне виробництво) стало популярним, значною мірою спираючись на планування перед виробництвом; отже, існує потенціал для допоміжних завдань, таких як візуалізація кадрів перед виробництвом або 3D-дизайн сцени. Це включатиме вирішення проблем рендерингу високої роздільної здатності в реальному часі, пов'язаних з проектуванням великих живих фонів.

З нещодавньою появою конвеєрів з відкритим кодом для NeRF, модифікація та тестування різних NeRF стали тривіальними. Це продемонстрували популярні онлайн-творці, такі як Рен Вайхман (учасник YouTube-каналу Corridor Crew), який демонструє використання NeRF для численних завдань CGI та VFX, таких як створення ресурсів та проекція фону, у відео під назвою «Я випробовую технологію, яка одного дня замінить CG». Це можна покращити для використання в конвеєрі ICVFX, поєднавши методи NeRF/GS, що моделюють фони, з автоматизованими методами сегментації переднього плану в реальному часі. Перевага цього порівняно з існуючими рішеннями ICVFX полягає в тому, що нам більше не потрібно вручну моделювати реалістичний віртуальний фон, а натомість можна моделювати цифрових двійників реальних середовищ, що значно менше трудомістко та дорого. Як поточний приклад, BBC досліджує NeRF для синтезу нових кадрів, які важко отримати вручну. Вони здатні досягти цього на практиці завдяки добре продуманому, специфічному для сцени методу захоплення.. Крім того, сучасні методи оцінки NeRF з використанням метрик на основі 2D зображень, таких як PSNR, SSIM та LPIPS, можливо, не повністю відображають точність 3D полів, які, як ми очікуємо, вивчить мережа NeRF. Це особливо стосується динамічного сценарію використання. Отже, ця галузь потребує подальшої роботи, перш ніж ми зможемо *впевнено* вибрати моделі для великомасштабного комерційного виробництва. Варто також зазначити, що автоматизовані технології 3D-зйомки впливають на поточні проблеми щодо власності візуальної сутності акторів.

					БР.ІІІ - 85.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		75

3.3 Аналіз ефективності роботи системи та перспективи розвитку технологій ІІІ у відеопродакшені

Дослідження ІК для живого виробництва залежить від типу середовища, типів дій, які очікуються відображення, розташування глядачів та центральної точки зору для цифрової аудиторії. Крім того, складні середовища та сценарії можуть призвести до погіршення візуальної якості, тому потрібні рішення в реальному часі для корекції або покращення зображення. У цьому підрозділі ми пов'язуємо роботи, що стосуються додаткових досліджень відстеження об'єктів та оцінки пози людини.

Оцінка пози людини (HPE) є популярною парадигмою в дослідженнях комп'ютерного зору. Ми бачимо випадки використання HPE в аналізі кінематографічних сцен, а також у трансляціях прямих подій. Для спорту основним фактором, який розділяє дослідження, є середовище. Ми робимо це розмежування перед оглядом методів, щоб підкреслити, що рішення є неоднорідними.

Розглянуто оцінку постави футболіста в режимі реального часу для аналізу презентацій у реальному часі. Для досягнення цієї мети використовується зважене злиття з впевненістю та візуальна увага для вирішення проблем кольорового камуфляжу та статичних ознак переднього плану, щоб спочатку ідентифікувати цільові ознаки переднього плану. На рисунку 3.8 ілюструється піксельна перевірка суглобів для ідентифікації ключових цільових ознак з використанням локальних бінарних шаблонів подібності (Ibsp). Потім генерується теплова карта за допомогою ResNet, багатошарової мережі пісочного годинника та агрегації глибоких шарів (DLA). DLA збирає ознаки з шарів CNN для визначення спільних ознак, тобто вона агрегує ознаки на кожному шарі, що стосуються перевіреного суглоба, і намагається класифікувати, що це за суглоб. Для оптимізації модель оновлюється адаптивними стратегіями, що стосуються ваги впевненості кожного пікселя та їх відповідної зваженої суми злиття - на основі класифікації суглобів. Цей метод демонструє значне покращення у виявленні суглобів та постави порівняно з іншими сучасними методами.

					БР.ІІІ - 85.00.00.000 ПЗ	Арк.
						76
Змн.	Арк.	№ докум.	Підпис	Дата		

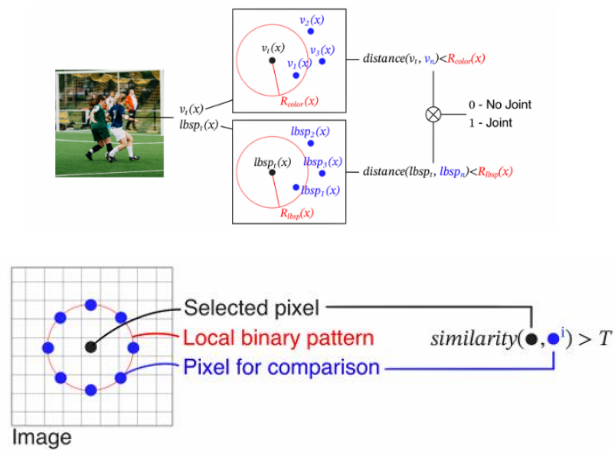


Рисунок 3.8 - Перевірка бінарних суглобів для кожного пікселя, що використовується для ідентифікації пікселів для класифікації суглобів

Мета полягає в тому, щоб побудувати 3D модель пози, використовуючи кілька 2D-видів, одночасно вирішуючи попередні проблеми тривалого часу обробки та залежності від черг на основі зовнішнього вигляду для ініціювання виявлення ознак. Це досягається шляхом представлення жадібного підходу до багатовидового виявлення. Підхід має три кроки: (1) Корекція помилок (2) Покадрова асоціація та (3) 3D-відстеження скелета Новизною цієї роботи є використання багатовидового HPE для перевірки розташування суглобів у 3D.

З аналогічними амбіціями, явно вдосконалюють методи визначення пози, використовуючи різні модулі з мережі VGG11 для різних методів об'єднання ознак. Для досягнення цієї мети оцінки ознак вибірково вибираються з мережі VGG11, а локальні ознаки точок спочатку агрегуються. Після того, як зображення проходить через мережу семантичної сегментації, тобто частина тіла, передається в мережу об'єднання ознак. Це сплайсує та об'єднує RGB-ознаки сегментованого зображення, локальні ознаки хмари точок та глобальні ознаки вигляду для формування кінцевого вектора ознак, який потім зводиться до скалярного значення для класифікації частини тіла.

Розглянуто високопродуктивне функціонування за допомогою стаціонарних аерофотокамер. Як і в попередніх моделях, для виявлення цілі (гравця) використовується CNN, а для виявлення пози – модель YOLOv2, яка навчається на

					БР.ІІІ - 85.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		77

загальнодоступних наборах аерофотознімків. Оброблені зображення проходять подальшу класифікацію нормальних та аномальних поз. Отже, модель забезпечує проникливий спосіб виявлення пози під загальним кутом огляду для подій, хоча й обмежена своєю здатністю виявляти аномальні пози внаслідок упередженого набору даних – важко знайти загальнодоступні набори даних, що стосуються аномальних поз.

Відстеження об'єктів. Як і у випадку з НРЕ, виявлення та відстеження об'єктів має значну присутність у спортивному трансляції з аналогічних причин. На відміну від НРЕ, рішення не прописуються в кожному окремому випадку. Швидше, вони вирішують неявні проблеми, враховуючи набір ідіоматичних характеристик середовища. Наприклад, для футболу ми могли б уявити метод виявлення м'яча в русі в шумному середовищі..

Представлено комплексний огляд підходів ODT, включаючи еволюційні алгоритми, звичайні нейронні мережі та нейронні мережі прямого зв'язку (FNN). В рамках ІК ми знаходимо дві різні парадигми: (1) ODT для однієї цілі та (2) ODT для кількох цілей. (1) набагато легше вирішити, і рішення, як правило, використовують детектування YOLO. Наприклад, розглядають два види спорту, зокрема, плавання та настільний теніс, для трансляції на великому екрані та на стадіоні. Для настільного тенісу використовується модель YOLOv4 для оцінки відскоку м'яча від відеопотоку та використовує це для визначення 3D-траєкторії. Для плавання рішенням є скорочена модель нейронної мережі ResNet-50, яка діє як базова мережа, разом з модифікованою моделлю SiamRPN++ для відстеження плавця. Для підтримки розпізнавання ракурс камери маскується для придушення фонового шуму. Зроблено висновок щодо труднощів зі зв'язком різнорідних інтерфейсів між співавторами, а також забезпеченням практичної доцільності цієї моделі. Методи YOLO найшвидше тестуються та налаштовуються і таким чином є привабливим кандидатом для практичної розробки. Вони також довели свою ефективність за наявності перекриттів. Менш успішна альтернатива, Mask R-CNN використовує дві мережі: одну для виявлення областей інтересу (тобто регіональну мережу пропозицій (RPN)) та глибоку CNN для визначення

					БР.ІІІ - 85.00.00.000 ПЗ	Арк.
						78
Змн.	Арк.	№ докум.	Підпис	Дата		

ймовірності того, що група пікселів у області інтересу пов'язана з об'єктом. Розглянуто кілька популярних методів трансляції спортивних подій на полі. Традиційні підходи виявляють візуальні зміни за допомогою класичних методів обробки зображень. Однак ці рішення є наївними щодо кількох обмежень, таких як (ii). Як альтернатива, методи на основі DL зазвичай будуються як Fast R-CNN та YOLO. Знову ж таки, є також аргументи на користь Masked R-CNN. Дивно, але виявили, що традиційні методи перевершують підходи DL, хоча й не набагато. Аргументація полягає в тому, що за більшої кількості цілей для відстеження глибини CNN не здатні розрізняти об'єкти з нижчою роздільною здатністю, тоді як традиційні методи повністю залежать від зображень з високою роздільною здатністю. У випадку 2 підходи DL наразі нездійсненні, оскільки виявлення малих об'єктів створює додаткові проблеми, пов'язані з виявленням об'єктів з низькою роздільною здатністю.

Порівняно з іншими формами виробництва, прямі трансляції мають давній зв'язок з дослідженнями ШІ. Наприклад, ВВС брала участь у підтримці досліджень штучного інтелекту в умовах прямого виробництва, збираючи та безкоштовно надаючи набір даних для підтримки телевізійних виробничих завдань, таких як покращене кадрування. Зв'язок з галуззю можна побачити у значних зусиллях щодо підвищення ефективності детекторів поз та об'єктів для людських цілей. Однак, у нас є сумніви щодо поточної практичності цих моделей, враховуючи їхню недостатню точність порівняно з людськими можливостями, хоча це добре розуміється дослідницькою спільнотою. Враховуючи кількість задіяних процесів, існує кілька інших можливих напрямків для досліджень. Наприклад, ми могли б розширити допоміжну інфраструктуру, дослідивши різні стратегії маскування для придушення загального набору шумного оточення. Крім того, значна кількість відеоматеріалів легкодоступна в Інтернеті. Однак дослідники повинні враховувати упередження щодо популярності, гендерної та расової різноманітності спортивних подій. Це вказує на необхідність пошуку більш різноманітних наборів даних для загального використання.

Тим не менш, використання штучного інтелекту в сучасному суспільстві

					БР.ІП - 85.00.00.000 ПЗ	Арк.
						79
Змн.	Арк.	№ докум.	Підпис	Дата		

створює низку проблем, які виходять за рамки технічних питань та потребують широкого міждисциплінарного обговорення. Здатність штучного інтелекту впливати на різні аспекти життя відкриває нові перспективи, але також викликає занепокоєння, пов'язане з порушеннями етичних норм, загостренням соціальної нерівності та загрозами конфіденційності.

Однією з фундаментальних проблем використання ШІ у створенні аудіо-візуального контенту є визначення авторства. У традиційному розумінні авторство базується на унікальному внеску людини, але в контексті ШІ це поняття стає складнішим. Алгоритми, що генерують контент на основі великих наборів даних, не діють автономно у творчому сенсі, але їхні результати часто викликають питання щодо розподілу прав інтелектуальної власності. Зокрема, незрозуміло, чи має автором бути програміст, який розробив алгоритм, компанія, яка його використовує, чи кінцевий користувач, який визначив параметри для генерації контенту. Водночас процес навчання ШІ, який базується на аналізі існуючих творів, створює ризик відтворення контенту, подібного до оригіналів, що може призвести до юридичних конфліктів та звинувачень у порушенні авторських прав. Залежність творчих процесів від технічних можливостей алгоритмів призводить до розмивання меж між людською творчістю та автоматизованим виробництвом. ШІ генерує контент, який не має прямого зв'язку з культурним контекстом чи суб'єктивним досвідом, характерним для людської творчості, що викликає сумніви щодо достовірності таких результатів. Ця ситуація підкреслює необхідність адаптації правових норм та створення нових правових механізмів, що враховують специфіку творчої діяльності ШІ. Зокрема, важливо забезпечити справедливий розподіл прав та обов'язків між усіма учасниками процесу створення контенту, включаючи розробників, компанії та користувачів, щоб сприяти регуляторній узгодженості та мінімізації конфліктів.

Ще серйознішою проблемою є порушення конфіденційності. ШІ часто використовується для створення порнографічних матеріалів із залученням відомих осіб без їхньої згоди, що є серйозним порушенням їхніх прав. Технології Deepfake дозволяють змінювати відео чи аудіо таким чином, що вони виглядають

					БР.ІІІ - 85.00.00.000 ПЗ	Арк.
						80
Змн.	Арк.	№ докум.	Підпис	Дата		

цілком реалістичними, підриваючи репутацію окремих осіб та установ. Окрім репутаційних ризиків, такі матеріали становлять загрозу для психологічного благополуччя постраждалих осіб, викликаючи стрес, громадський осуд або дискримінацію. Це вимагає впровадження суворих правових та технічних механізмів для запобігання зловживанню такими інструментами. Превентивні заходи повинні включати як кримінальну відповідальність за створення таких матеріалів, так і технічні рішення для їх ідентифікації.

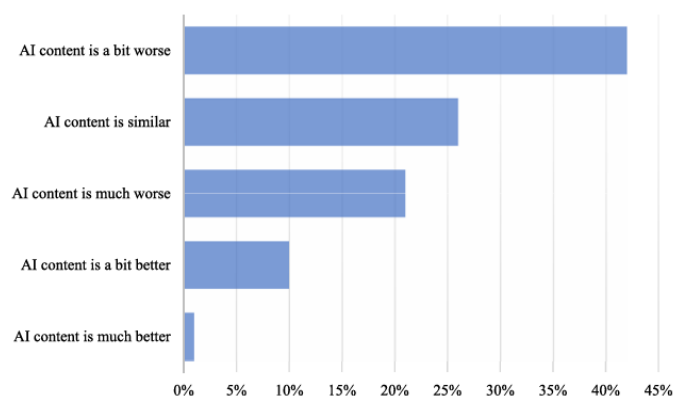


Рисунок 3.9 - Якість контенту, створеного штучним інтелектом, порівняно з контентом, створеним людиною.

Автоматизація творчих процесів також має значний соціальний вплив, особливо на зайнятість. Хоча ШІ допомагає оптимізувати рутинні завдання, такі як редагування або створення субтитрів, він також витісняє людей з цих сфер, що сприяє зменшенню можливостей працевлаштування. Працівники аудіовізуальної галузі змушені адаптуватися до нових умов, часто змінюючи кваліфікацію або переходячи на ролі, пов'язані з контролем та адмініструванням ШІ. Це призводить до трансформації ринку праці, де компетенції в галузі технологій цінуються все більше, але традиційним творчим навичкам приділяється менше уваги. Такі зміни призводять до структурної нерівності в суспільстві: працівники, які не мають доступу до технічної освіти або ресурсів для перепідготовки, можуть опинитися під загрозою безробіття, тоді як добре оплачувані посади в технологічному секторі стають доступними лише обмеженому колу фахівців.

Чинні закони про інтелектуальну власність недостатньо здатні вирішити проблеми, пов'язані зі складнощами, що виникають внаслідок використання ШІ під час розробки контенту. Основною проблемою є встановлення авторства на твори, створені за допомогою штучного інтелекту, оскільки чинне законодавство, включаючи законодавство ЄС та США, передбачає авторство людини. Тим не менш, незалежно від того, чи використовується штучний інтелект для створення музики, сценаріїв чи візуальних ефектів, право власності на авторські права залишається неоднозначним, незалежно від того, чи належить воно розробнику штучного інтелекту, користувачеві, який надає дані, чи самому штучному інтелекту. Підкреслено цей недолік у правовій базі, припускаючи, що чинне законодавство може потребувати змін, щоб охопити твори, створені за допомогою штучного інтелекту.

Більше того, технології дипфейків, що використовують штучний інтелект для створення дуже переконливих підроблених фільмів чи звуків, становлять значні ризики для цілісності медіа. Ці технології сприяють створенню зміненого матеріалу, який може легко поширювати дезінформацію. Чинне законодавство неадекватно регулює використання або поширення дипфейків, роблячи людей та організації вразливими до репутаційної шкоди або політичної експлуатації. Існує явний регуляторний дефіцит у пом'якшенні ризиків, пов'язаних з цими технологіями на базі ші.

Більше того, правила конфіденційності даних, такі як GDPR, часто не пом'якшують небезпеки, пов'язані з технологією штучного інтелекту, особливо коли системи штучного інтелекту обробляють персональні дані для створення персоналізованого контенту. Залежність штучного інтелекту від великих наборів даних для створення контенту створює суттєві проблеми з підзвітністю та прозорістю, особливо коли включаються персональні дані. Відсутність відкритості в процесах прийняття рішень щодо штучного інтелекту стає суттєвою етичною проблемою. Роз'яснення моделей ШІ є важливим для гарантування того, що результати, створені штучним інтелектом, є морально стійкими та неупередженими. За відсутності чітких правових рамок для прозорості штучного інтелекту

					БР.ІП - 85.00.00.000 ПЗ	Арк.
						82
Змн.	Арк.	№ докум.	Підпис	Дата		

існує небезпека того, що матеріали, створені штучним інтелектом, можуть вводити громадськість в оману або посилювати упередження, особливо в чутливих ЗМІ, таких як новини чи політичні дискурси.

Більше того, автоматизація загострює проблему зниження соціального статусу традиційних професій у креативному секторі. Використання ШІ може призвести до стигматизації людської праці як менш ефективної або застарілої, що негативно впливає на мотивацію та професійну самореалізацію митців та працівників креативної індустрії. Це також створює нові виклики для соціального сприйняття креативності, оскільки автоматизовані системи знижують важливість індивідуального внеску, замінюючи його стандартизованими підходами. Розширення автоматизації сприяє поглибленню цифрової нерівності, коли країни чи соціальні групи, які мають менший доступ до передових технологій, втрачають свою конкурентоспроможність у світовій економіці.

Очікується, що широка інтеграція штучного інтелекту в аудіовізуальну творчість вплине на робочі місця в креативній індустрії. Технології штучного інтелекту, що автоматизують такі операції, як відеомонтаж, кольорокорекція та анімація, можуть призвести до витіснення робочих місць для творчих людей, включаючи редакторів та аніматорів. Хоча можуть виникнути нові професії в адмініструванні та контролі за допомогою штучного інтелекту, традиційні творчі ролі можуть зіткнутися зі скороченням. Крім того, розробка контенту на основі штучного інтелекту становить загрозу для стандартизації творчості, тим самим зменшуючи різноманітність художнього самовираження. Культурна гомогенізація може виникнути, якщо системи штучного інтелекту переважно відображатимуть західні медіа-тенденції, тим самим обмежуючи вираження незахідних перспектив у глобальних медіа.

Брак прозорості в моделях штучного інтелекту, зокрема в глобальних мережах (GAN) та глибокому навчанні, викликає занепокоєння, особливо коли ці інструменти створюють або модифікують медіаматеріали.

Використання штучного інтелекту в аудіовізуальному секторі вимагає балансу між технологічним прогресом, етичними стандартами та соціальними

					БР.ІІІ - 85.00.00.000 ПЗ	Арк.
						83
Змн.	Арк.	№ докум.	Підпис	Дата		

викликами. Вкрай важливо розробити механізми, що забезпечують прозорість процесів створення контенту, захищають права на конфіденційність та запобігають маніпуляціям. Тільки дотримуючись етичних стандартів, штучний інтелект може стати не лише технологічним досягненням, а й інструментом, який приносить користь суспільству в цілому, підтримуючи соціальну справедливість та безпеку.

Розвиток ШІ створює нові можливості та виклики в різних галузях, зокрема в медіа, де ці технології сприяють змінам у створенні та споживанні контенту. Порівняння результатів цього дослідження з роботами інших авторів дозволяє оцінити значення застосування штучного інтелекту в контексті сучасних наукових розробок. Оцінка впливу технологій штучного інтелекту на медіавиробництво, а також врахування етичних та безпекових аспектів таких змін є важливою частиною академічної дискусії, оскільки поглиблює розуміння як позитивних, так і негативних наслідків їх впровадження.

Застосування ШІ у створенні контенту все ще перебуває на початковій стадії, він вже справив суттєвий вплив, підвищивши ефективність виробництва та трансформувавши творчий процес. Наголошено на необхідності подальших досліджень довгострокових наслідків інтеграції ШІ, особливо для творчих робочих процесів та виробничих процесів. Це узгоджується з висновками цього дослідження, які вказують на те, що хоча ШІ забезпечує очевидні короткострокові переваги в ефективності, його довгостроковий вплив на автономію творчих працівників та загальну якість матеріалу залишається неоднозначним. Це вимагає збалансованої стратегії інтеграції ШІ в медіа, що забезпечує збереження як інновацій, так і художньої цілісності.

Проаналізовано проблеми, які твори, створені штучним інтелектом, створюють для законодавства про авторське право, особливо щодо авторства. Їхні дослідження підкреслюють, що, хоча деякі аспекти можуть вимагати додаткових роз'яснень, особливо щодо авторства та прав творців, які використовують технології штучного інтелекту, чинні правові рамки переважно оснащені для

управління контентом, створеним штучним інтелектом, відповідно до чинних законів про авторське право.

Результати відображають занепокоєння щодо інтеграції ШІ у виробництво контенту, що створює суттєві етичні проблеми, особливо щодо гарантування того, що матеріал, створений ШІ, позбавлений упереджень, а процеси прийняття рішень є зрозумілими та обґрунтованими. ШІ революціонізує створення контенту, недостатня відкритість у прийнятті рішень є значною перешкодою, особливо коли він генерує матеріал, який впливає на громадську думку або зображує реальні події.

Досліджено загрози безпеці, що виникають від матеріалів, створених штучним інтелектом, зокрема діпфейків, які викликали занепокоєння щодо їхньої здатності підривати віру в легітимність медіа. ШІ може покращити створення контенту, його негативні аспекти, представлені такими технологіями, як діпфейки, становлять значні загрози для довіри до медіа. Підкреслено нагальну необхідність інструментів на базі ШІ для виявлення діпфейків, у зв'язку з цим важливі інструменти для автоматичного виявлення діпфейків повинні бути інтегровані в медіаплатформи, щоб запобігти їх поширенню. Розробка таких інструментів повинна постійно оновлюватися, щоб забезпечити їхню ефективність у боротьбі з новими методами фальсифікації контенту.

Проаналізовано етичні наслідки використання ШІ, відбувається розширення можливостей персоналізації контенту та сприяння поширенню дезінформації, та підкреслює необхідність чіткої політики, що контролює використання ШІ у створенні медіа, щоб запобігти його зловживанню у створенні оманливих або шкідливих матеріалів. Заклик до вивчення бізнес-моделей платформ є особливо актуальним, враховуючи зростання частоти дезінформації, що генерується ШІ, яка може посилюватися алгоритмами соціальних мереж.

Персоналізація контенту є однією з ключових сфер застосування ШІ в медіасекторі, оскільки вона значно покращує взаємодію з аудиторією та підвищує якість аудіовізуального контенту. Підкреслено здатність штучного інтелекту

					БР.ІІІ - 85.00.00.000 ПЗ	Арк.
Змн.	Арк.	№ докум.	Підпис	Дата		85

сприяти створенню персоналізованого медіаконтенту, адаптуючи матеріал відповідно до вподобань користувача. Поняття персоналізації на основі штучного інтелекту є центральним питанням у сучасних дослідженнях, підкреслюючи, що штучний інтелект може забезпечити автоматизацію виробництва контенту, що призведе до створення більш персоналізованих та захопливих аудіовізуальних товарів.

Досліджено функцію глибокого навчання в автоматизації класифікації та адміністрування інформації, зокрема, з великими цифровими архівами. Цей аспект використання глибокого навчання особливо актуальний для отриманих вище результатів, де здатність штучного інтелекту ефективно обробляти великі дані була визнана ключовим елементом у медіавиробництві. Важливість штучного інтелекту на етапах постпродакшну, сприяючи організації, зберіганню та пошуку медіаактивів. Штучний інтелект може забезпечити для управління медіа, особливо в контекстах, що вимагають адміністрування великих аудіовізуальних архівів. Глибоке навчання дозволяє створювати точніші та швидші системи пошуку, здатні аналізувати та структурувати величезну кількість різноманітних медіаресурсів, таких як відео, аудіо, зображення та текст, на основі їхнього вмісту, а також виділяти та зберігати найцінніші фрагменти інформації. Крім того, важливо зазначити, що ці технології не лише покращують етапи постпродакшну, де контент потрібно організовувати та зберігати, але й можуть використовуватися на всіх етапах медіавиробництва. Методи глибокого навчання дозволяють автоматизувати процеси моніторингу та модерації контенту в режимі реального часу, що критично важливо в ситуаціях, коли обробка великих обсягів матеріалу вимагає значного часу та людських ресурсів. Таким чином, ці технології значно підвищують ефективність та швидкість виробництва, а також забезпечують вищу якість кінцевого продукту.

Наголошено на труднощах, з якими стикається аудиторія у виявленні матеріалів, створених штучним інтелектом. Дослідження підкреслює необхідність удосконалених методів виявлення, оскільки більшість матеріалів, створених

					БР.ІІІ - 85.00.00.000 ПЗ	Арк.
						86
Змн.	Арк.	№ докум.	Підпис	Дата		

штучним інтелектом, залишаються непоміченими громадськістю. Поточне дослідження підтверджує це, підкреслюючи зростаючу проблему диференціації медіа, створених штучним інтелектом, від контенту, створеного людиною. Це створює значний виклик для легітимності медіа та викликає занепокоєння щодо можливості маніпуляцій. З розвитком технологій виявлення існує необхідність підвищення обізнаності та освіти громадськості щодо можливостей використання штучного інтелекту у медіавиробництві для зменшення небезпеки дезінформації. Цей розвиток подій має серйозні наслідки для довіри до медіа. У ситуаціях, коли громадськість може бути необізнаною або не в змозі відрізнити штучно створений контент від реальних джерел, існує ризик поширення дезінформації та маніпуляцій, які можуть вплинути на громадську думку та політичні процеси. Рішення цієї проблеми включають створення міток або водяних знаків для контенту, що вказують на його штучне походження, а також розробку методів, що дозволяють аудиторії самостійно перевіряти справжність джерел через спеціалізовані платформи або додатки.

Таким чином, розвиток технологій штучного інтелекту в медіасекторі не лише надає нові можливості для створення та редагування контенту, але й створює значні етичні, правові та безпекові проблеми для суспільства. Порівняння з іншими дослідженнями показує важливість багатогранного підходу до використання штучного інтелекту, який включає не лише технологічні інновації, але й розробку стандартів для запобігання потенційним ризикам.

Майбутні дослідження впливу штучного інтелекту на аудіовізуальний контент повинні пріоритезувати розробку прозорих алгоритмів для забезпечення підзвітності та пом'якшення маніпуляцій, особливо за допомогою технології дипфейків, що вимагає швидкого створення регуляторних баз та механізмів виявлення для боротьби з дезінформацією. Етичні принципи щодо контенту, створеного штучним інтелектом, мають вирішальне значення для запобігання увічненню упереджень та сприяння рівності, різноманітності та культурній обізнаності. Крім того, оцінка впливу штучного інтелекту на зайнятість у креативному секторі має вирішальне значення, оскільки автоматизація ШІ може замінити

					БР.ІІІ - 85.00.00.000 ПЗ	Арк.
						87
Змн.	Арк.	№ докум.	Підпис	Дата		

традиційні творчі ролі, що порушує питання адаптації та підвищення кваліфікації. Крім того, майбутні дослідження повинні дослідити екологічні наслідки навчання великих моделей ШІ, які потребують значних обчислювальних ресурсів, та визначити стратегії для підвищення енергоефективності ШІ. Крім того, забезпечення конфіденційності та безпеки даних у матеріалах, створених ШІ, особливо щодо особистої інформації, буде першочерговим обов'язком. Зрештою, дослідження повинні вивчити вплив ШІ на сприяння культурному різноманіттю, щоб запобігти гомогенізації контенту, гарантуючи, що системи ШІ поважають місцеві культури та збагачують глобальний медіа ландшафт.

Вплив штучного інтелекту в цій галузі проявився у значному скороченні часу виробництва, оптимізації ресурсів та впровадженні інноваційних технологій. Завдяки автоматизації рутинних завдань штучний інтелект спрощує такі процеси, як монтаж, корекція відео та створення складних візуальних ефектів. Генеративні моделі, такі як GAN, дозволяють створювати реалістичні об'єкти та сцени з мінімальними витратами. Крім того, алгоритми глибокого навчання успішно застосовуються для відновлення старих аудіовізуальних матеріалів, покращуючи якість звуку та відео, автоматизуючи створення трейлерів та персоналізуючи контент. Ці досягнення відкривають нові перспективи для галузі, дозволяючи продуктам швидше адаптуватися до вимог ринку.

Однак, дослідження також виявило низку проблем, включаючи стандартизацію контенту, ризики, пов'язані з поширенням фейкових матеріалів (діпфейків), порушення авторських прав та неадекватний регуляторний нагляд. Етичні проблеми, пов'язані з відсутністю прозорості алгоритмів та загрозами конфіденційності, вимагають розробки нових регуляторних механізмів та нагляду за використанням технологій штучного інтелекту. Тим не менш, дослідження має певні обмеження. Воно в першу чергу зосереджувалося на якісному аналізі сучасних технологій штучного інтелекту та їхнього впливу, залишаючи кількісні оцінки ефективності впроваджених рішень поза його сферою. Більше того, через швидкий розвиток штучного інтелекту деякі з представлених технологій можуть застаріти, що вимагатиме постійного оновлення дослідницької бази. Тому

					БР.ІІІ - 85.00.00.000 ПЗ	Арк.
						88
Змн.	Арк.	№ докум.	Підпис	Дата		

подальші дослідження можуть бути зосереджені на розробці прозорих та контекстно-адаптованих алгоритмів, створенні систем для запобігання маніпуляціям інформацією та оцінці довгострокових соціальних наслідків автоматизації творчих процесів.

3.4 Висновки по розділу

У третьому розділі було реалізовано систему автоматичного монтажу відео та проведено оцінювання її ефективності. Проаналізовано сценарії використання інтелектуальної автоматизації відеовиробництва та розроблено модулі автоматизованого й віртуального відеомонтажу. Проведене тестування системи підтвердило можливість ефективного використання алгоритмів штучного інтелекту для автоматичного формування відеоконтенту, оптимізації монтажу та скорочення часу обробки матеріалів. Отримані результати показали, що використання ШІ дозволяє підвищити продуктивність процесу створення відео та мінімізувати вплив людського фактора. Водночас визначено перспективи подальшого розвитку технологій штучного інтелекту у сфері відеопродакшену, пов'язані з удосконаленням генеративних моделей, персоналізацією контенту та підвищенням рівня автоматизації творчих процесів.

					БР.ІП - 85.00.00.000 ПЗ	Арк.
						89
Змн.	Арк.	№ докум.	Підпис	Дата		

ВИСНОВКИ

У ході виконання дипломної роботи було досліджено та проаналізовано методи застосування штучного інтелекту для автоматичного монтажу та редагування відео-контенту, а також розроблено концепцію програмної системи, здатної автоматизувати процеси обробки відеоматеріалів. У роботі розглянуто сучасні підходи до використання нейронних мереж, алгоритмів комп'ютерного зору та машинного навчання для оптимізації створення мультимедійного контенту. Особливу увагу приділено автоматизації монтажу, інтелектуальному аналізу відео, синхронізації аудіо- та відеодоріжок, а також підвищенню ефективності виробничих процесів у сфері цифрового медіа.

На початковому етапі було виконано аналіз теоретичних основ застосування штучного інтелекту у сфері створення відео-контенту. Було досліджено сучасний стан розвитку технологій автоматичного монтажу, визначено ключові принципи функціонування систем на основі штучного інтелекту та проаналізовано переваги їх використання у порівнянні з традиційними методами редагування відео. Окрему увагу приділено впливу штучного інтелекту на контент-індустрію, включаючи як позитивні аспекти — підвищення швидкості виробництва, зниження витрат та автоматизацію рутинних операцій, так і потенційні ризики, пов'язані з авторським правом, дезінформацією та генерацією синтетичного контенту.

У другому розділі було досліджено методи та інструменти, які застосовуються для автоматичного редагування відео. Зокрема, розглянуто принципи роботи згорткових нейронних мереж, алгоритмів розпізнавання об'єктів, сцен та руху у відеопотоці. Проведено аналіз технологій редагування та маніпуляції відео-контентом, включаючи автоматичне кадрування, стабілізацію, генерацію переходів та визначення ключових сцен. Також було досліджено набори даних та метрики оцінювання якості моделей штучного інтелекту, що дозволило визначити найбільш ефективні підходи до побудови інтелектуальних систем монтажу.

					БР.ІІІ - 85.00.00.000 ПЗ	Арк.
						90
Змн.	Арк.	№ докум.	Підпис	Дата		

У процесі реалізації програмного рішення було сформовано структуру системи автоматичного монтажу відео-контенту, яка забезпечує аналіз відеоматеріалу, виявлення важливих фрагментів та автоматичне створення змонтованого відео відповідно до заданих критеріїв. Реалізація передбачала інтеграцію алгоритмів машинного навчання, модулів обробки відео та інструментів для роботи з мультимедійними файлами. Значна увага приділялася продуктивності системи, оптимізації обробки великих обсягів відеоданих та забезпеченню високої якості кінцевого результату.

Окремим етапом роботи стало тестування та оцінка ефективності запропонованої системи. Було проведено серію експериментів для перевірки точності автоматичного визначення сцен, швидкості обробки відео та якості автоматично сформованого монтажу. Результати тестування показали, що використання алгоритмів штучного інтелекту дозволяє значно скоротити час редагування відео та підвищити ефективність виробничих процесів у сфері цифрового контенту. Також було визначено ключові обмеження існуючих моделей, зокрема потребу у великих обсягах навчальних даних та високих обчислювальних ресурсах.

Загалом, поставлені у роботі цілі були досягнуті. Розглянуті методології та інструменти підтвердили ефективність використання штучного інтелекту для автоматичного монтажу та редагування відео-контенту. Запропонований підхід дозволяє автоматизувати значну частину процесів створення мультимедійного контенту, підвищити швидкість роботи та забезпечити стабільну якість результату. Перевагами розробленої концепції є можливість масштабування, інтеграція сучасних AI-технологій та підтримка автоматизованого аналізу відео.

У перспективі система може бути вдосконалена шляхом використання генеративних моделей штучного інтелекту, інтеграції технологій синтезу мовлення та автоматичного створення сценаріїв, а також підтримки хмарної обробки відео у реальному часі.

					БР.ІІІ - 85.00.00.000 ПЗ	Арк.
						91
Змн.	Арк.	№ докум.	Підпис	Дата		

СПИСОК ПОСИЛАНЬ НА ДЖЕРЕЛА

1. Adobe Premiere Pro Documentation. (2025). adobe.com.
<https://helpx.adobe.com/premiere-pro/user-guide.html>

2. AI Video Editing Tools Overview. (2024). techradar.com.
<https://www.techradar.com/best/best-ai-video-editing-software>

3. Amazon Web Services. (2025). Machine Learning for Media and Entertainment. aws.amazon.com. <https://aws.amazon.com/media/machine-learning/>

4. Andrej Karpathy Blog. (2023). Deep Learning for Visual Recognition. karpathy.ai. <https://karpathy.ai/>

5. Artificial Intelligence in Video Production. (2024). ibm.com.
<https://www.ibm.com/topics/artificial-intelligence>

6. Blender Documentation. (2025). docs.blender.org.
<https://docs.blender.org/manual/en/latest/>

7. Brown, T., & Davis, M. (2024). AI-assisted cinematic video generation and editing systems. ACM Transactions on Multimedia Computing, 20(4), 55–72.
<https://doi.org/10.1145/3678901>

8. Computer Vision with OpenCV. (2025). opencv.org. <https://opencv.org/>

9. Convolutional Neural Networks for Visual Recognition. (2023). cs231n.stanford.edu. <https://cs231n.stanford.edu/>

10. Cutler, R., & Turk, M. (2022). Smart video editing using machine learning techniques. IEEE Multimedia, 29(3), 44–58.
<https://doi.org/10.1109/MMUL.2022.3156789>

11. DaVinci Resolve Documentation. (2025). blackmagicdesign.com.
https://documents.blackmagicdesign.com/UserManuals/DaVinci_Resolve_Manual.pdf

12. Deep Learning. (2024). deeplearningbook.org.
<https://www.deeplearningbook.org/>

13. DeepMind Technologies. (2025). Video understanding with AI systems. deepmind.google. <https://deepmind.google/research/>

14. FFmpeg Documentation. (2025). ffmpeg.org.

<https://ffmpeg.org/documentation.html>

15. Fowler, M. (2002). Patterns of Enterprise Application Architecture. martinowler.com. <https://martinfowler.com/eaCatalog/>

16. Freeman, E., & Robson, E. (2021). Head First Design Patterns. oreilly.com. <https://www.oreilly.com/library/view/head-first-design/9781492077992/>

17. Gamma, E., Helm, R., Johnson, R., & Vlissides, J. (1994). Design Patterns: Elements of Reusable Object-Oriented Software. pearson.com. <https://www.pearson.com/store/p/design-patterns-elements-of-reusable-object-oriented/P100000252058>

18. Google AI Blog. (2024). Advances in generative AI for multimedia content. ai.googleblog.com. <https://ai.googleblog.com/>

19. Google Cloud Video Intelligence API. (2025). cloud.google.com. <https://cloud.google.com/video-intelligence>

20. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. mitpress.mit.edu. <https://www.deeplearningbook.org/>

21. Hassan, M., & Clark, P. (2024). Automated scene detection in AI-driven video editing systems. IEEE Access, 12, 55678–55695. <https://doi.org/10.1109/ACCESS.2024.3378901>

22. Keras Documentation. (2025). keras.io. <https://keras.io/>

23. Kim, S., & Lee, J. (2023). Neural networks for automatic cinematic editing. Multimedia Tools and Applications, 82(14), 21455–21473. <https://doi.org/10.1007/s11042-023-14567-2>

24. Martin, R. C. (2017). Clean Architecture: A Craftsman's Guide to Software Structure and Design. pearson.com. <https://www.pearson.com/store/p/clean-architecture-a-craftmans-guide-to-software-structure-and-design/P100001465201>

25. Microsoft Azure AI Services Documentation. (2025). learn.microsoft.com. <https://learn.microsoft.com/en-us/azure/ai-services/>

26. MongoDB Documentation. (2025). mongodb.com. <https://www.mongodb.com/docs/>

27. Neural Networks and Deep Learning. (2024). coursera.org.

					БР.ІІІ - 85.00.00.000 ІІЗ	Арк.
						93
Змн.	Арк.	№ докум.	Підпис	Дата		

<https://www.coursera.org/specializations/deep-learning>

28. NVIDIA Developer Documentation. (2025). developer.nvidia.com.
<https://developer.nvidia.com/documentation>

29. OpenAI API Documentation. (2025). platform.openai.com.
<https://platform.openai.com/docs>

30. OpenCV AI Kit Documentation. (2025). docs.luxonis.com.
<https://docs.luxonis.com/>

31. OpenShot Video Editor Documentation. (2025). [openshot.org](https://www.openshot.org).
<https://www.openshot.org/user-guide/>

32. OWASP Top Ten Security Risks. (2024). owasp.org.
<https://owasp.org/www-project-top-ten/>

33. PyTorch Documentation. (2025). pytorch.org.
<https://pytorch.org/docs/stable/index.html>

34. Radford, A., Kim, J. W., Hallacy, C., et al. (2021). Learning transferable visual models from natural language supervision. [arxiv.org](https://arxiv.org/abs/2103.00020).
<https://arxiv.org/abs/2103.00020>

35. Shotstack API Documentation. (2025). shotstack.io.
<https://shotstack.io/docs/>

36. Szeliski, R. (2022). Computer Vision: Algorithms and Applications. szeliski.org.
<https://szeliski.org/Book/>

37. TensorFlow Documentation. (2025). [tensorflow.org](https://www.tensorflow.org).
<https://www.tensorflow.org/learn>

38. Wang, Y., & Zhang, H. (2025). Intelligent video summarization using deep neural networks. ACM Transactions on Multimedia Computing, 21(1), 1–18.
<https://doi.org/10.1145/3690123>

39. YouTube Creator Academy. (2025). Video editing best practices. [creators.youtube.com](https://support.google.com/youtube/answer/11221953).
<https://support.google.com/youtube/answer/11221953>

40. Zhang, X., & Liu, Y. (2024). AI-based real-time video enhancement and editing systems. Journal of Visual Communication and Image Representation, 98, 104012. <https://doi.org/10.1016/j.jvcir.2024.104012>

					БР.ІІІ - 85.00.00.000 ІІЗ	Арк.
						94
Змн.	Арк.	№ докум.	Підпис	Дата		

БІБЛІОГРАФІЧНА ДОВІДКА

Тема бакалаврської роботи: " Застосування штучного інтелекту для автоматичного монтажу та редагування відео-контенту"

Обсяг пояснювальної записки: 91 аркуш

Дата закінчення бакалаврської роботи 10 червня 2026р.

Підпис студента _____