

МАГІСТЕРСЬКА РОБОТА

МР. Шм - 62.00.00.000 ПЗ

Група Шм-24-2

Кордяк Ростислав

2025

Івано-Франківський національний технічний університет нафти і газу

Факультет інформаційних технологій

Кафедра інженерії програмного забезпечення

Кордяк Ростислав Михайлович

(прізвище, ім'я, по батькові)

УДК 004.9
(індекс)

МАГІСТЕРСЬКА РОБОТА

Моделі, методи та алгоритми генерації тривимірних моделей із

зображень

(назва роботи)

Інженерія програмного забезпечення

(назва освітньої програми)

121 - Інженерія програмного забезпечення

(шифр і назва спеціальності)

Кордяк Р.М.

(підпис, ініціали та прізвище здобувача освітнього ступеня)

Науковий керівник Григорчук Любомир Іванович, к.т.н., доцент

(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

Допущено до захисту

Завідувач кафедри

доц. Бандура В.В.

(посада) (підпис) (дата) (ініціали та прізвище)

Нормоконтроль

доц. Вовк Р.Б.

(посада) (підпис) (дата) (ініціали та прізвище)

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

Івано-Франківськ – 2025

Івано-Франківський національний технічний університет нафти і газу

Факультет інформаційних технологій

Кафедра інженерії програмного забезпечення

Освітній рівень магістр

Спеціальність 121 – Інженерія програмного забезпечення

ЗАТВЕРДЖУЮ:

Зав. кафедрою

ІПЗ

доц.

В.В. Бандура

“ 04 ” вересня 2025 р.

ЗАВДАННЯ

НА МАГІСТЕРСЬКУ РОБОТУ СТУДЕНТУ

Кордяку Ростиславу Михайловичу

(прізвище, ім'я, по-батькові)

1. Тема магістерської роботи “ **Моделі, методи та алгоритми генерації тривимірних моделей із зображень** ”

керівник проекту (роботи) Григорук Л.І., к.т.н., доцент

затверджені наказом закладу вищої освіти від “ 05 ” листопада 2025 р. № 695/7

2. Строк подання студентом проекту (роботи) 15 грудня 2025 р.

3. Вихідні дані до проекту (роботи) Формальні моделі і методи побудови інформаційних та програмних технологій певного класу

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

1. Аналіз предметної області генерації 3D моделей та структур із зображень

2. Етапи 3D-реконструкції з послідовності зображень

3. Методологія та оцінка критеріїв для генерації тривимірної об'ємної моделі об'єктів

4. Імплементация моделей та методів генерації тривимірних моделей

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

1. Огляд підходу до 3D-запису на основі зображень (рис. 1.1)

2. 3D-запис культурної спадщини (рис. 1.2)

3. Графічна інтерпретація методу епіполярної геометрії (рис. 1.3)

4. Приклади реалізацій алгоритму RANSAC (рис. 1.4)

5. Методологія 3D реконструкцій (рис. 1.5)

6. Консультанти розділів проекту (роботи)

| Розділ | Консультант | Підпис, дата |
|----------------------|------------------------|--------------|
| Перевірка на плагіат | доц., к.т.н. Вовк Р.Б. | |
| | | |

7. Дата видачі завдання 04 вересня 2025 р.

Керівник _____

(підпис)

Завдання прийняв до виконання _____

(підпис)

КАЛЕНДАРНИЙ ПЛАН

| № п/п | Назви етапів магістерської роботи | Строк виконання етапів роботи | Примітка |
|-------|--|-------------------------------|----------|
| 1 | Підбір і вивчення літератури по темі магістерської роботи | 15.09.2025 | виконано |
| 2 | Аналіз предметної області генерації 3d моделей та структур із зображень | 01.10.2025 | виконано |
| 3 | Етапи 3D-реконструкції з послідовності зображень | 17.10.2025 | виконано |
| 4 | Методологія та оцінка критеріїв для генерації тривимірної об'ємної моделі об'єктів | 02.11.2025 | виконано |
| 5 | Імплементация моделей та методів генерації тривимірних моделей | 19.11.2025 | виконано |
| 6 | Реалізація методології генеративних моделей у задачах синтезу 3D-моделей та сцен | 02.12.2025 | виконано |
| 7 | Затвердження пояснювальної записки роботи завідувачем кафедри | 15.12.2025 | виконано |

Студент – магістр _____

(підпис)

Керівник роботи _____

(підпис)

АНОТАЦІЯ

Магістерська робота: 80 с., 25 рис., 5 табл., 35 джерел.

Тема: Моделі, методи та алгоритми генерації тривимірних моделей із зображень

Метою магістерської роботи є дослідження, аналіз і узагальнення моделей, методів та алгоритмів генерації тривимірних моделей із зображень, а також обґрунтування ефективних підходів до 3D-реконструкції та генеративного синтезу тривимірних об'єктів.

Об'єктом дослідження є процес генерації та реконструкції тривимірних моделей на основі двовимірних зображень.

Предметом дослідження є моделі, методи та алгоритми 3D-реконструкції і генеративного синтезу тривимірних моделей. **Результати дослідження**

В роботі виконано комплексний аналіз сучасних підходів до генерації тривимірних моделей із використанням генеративного моделювання та розглянуто еволюцію генеративних моделей від двовимірного до тривимірного простору та визначено ключові етапи цього переходу.

Висновок

Досліджено математичні та алгоритмічні засади генерації тривимірних моделей із двовимірних зображень, а також сучасні підходи до синтезу 3D-контенту з використанням класичних методів комп'ютерного зору та генеративних моделей штучного інтелекту

3D-РЕКОНСТРУКЦІЯ, ГЕНЕРАЦІЯ ТРИВИМІРНИХ МОДЕЛЕЙ, КОМП'ЮТЕРНИЙ ЗІР, ФОТОКОНСИСТЕНТНІСТЬ, ВОКСЕЛЬНІ МЕТОДИ, ГЕНЕРАТИВНІ МОДЕЛІ, ТРИВИМІРНА ГРАФІКА, РЕНДЕРИНГ, ПРОСТОРОВЕ МОДЕЛЮВАННЯ.

ABSTRACT

Master Thesis: 80 pp., 25 fig., 5 tab., 35 sources.

Topic: Models, methods and algorithms for generating three-dimensional models from images

The purpose of the master's thesis is to study, analyze and generalize models, methods and algorithms for generating three-dimensional models from images, as well as to substantiate effective approaches to 3D reconstruction and generative synthesis of three-dimensional objects.

The object of the study is the process of generating and reconstructing three-dimensional models based on two-dimensional images.

The subject of the study is models, methods and algorithms for 3D reconstruction and generative synthesis of three-dimensional models. Research results

The work provides a comprehensive analysis of modern approaches to generating three-dimensional models using generative modeling, examines the evolution of generative models from two-dimensional to three-dimensional space, and identifies key stages of this transition.

Conclusion

The mathematical and algorithmic principles of generating three-dimensional models from two-dimensional images have been investigated, as well as modern approaches to synthesizing 3D content using classical computer vision methods and generative artificial intelligence models

3D RECONSTRUCTION, GENERATION OF THREE-DIMENSIONAL MODELS, COMPUTER VISION, PHOTOCONSISTENCY, VOXEL METHODS, GENERATIVE MODELS, THREE-DIMENSIONAL GRAPHICS, RENDERING, SPATIAL MODELING.

ЗМІСТ

| | |
|--|-----------|
| ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ | 9 |
| ВСТУП..... | 10 |
| РОЗДІЛ 1. АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ГЕНЕРАЦІЇ 3D МОДЕЛЕЙ ТА СТРУКТУР ІЗ ЗОБРАЖЕНЬ | 14 |
| 1.1. Теоретичні основи та сучасні підходи в комп'ютерному зорі щодо вилучення 3D-структури зі зображень | 14 |
| 1.2. Алгоритмічні основи 3D-реконструкції на основі послідовності зображень | 15 |
| 1.2.1. Етапи 3D-реконструкції з послідовності зображень..... | 16 |
| 1.2.2. Уникнення процесів необхідності калібрування | 20 |
| 1.3. Математичний апарат у 3D-реконструкції | 22 |
| 1.3.1. Епіполярна геометрія (Epipolar Geometry) | 22 |
| 1.3.2. Алгоритм RANSAC (RANDOM SAMPLE CONSENSUS) | 24 |
| 1.3.3. Опис методології | 26 |
| Висновки до розділу | 28 |
| РОЗДІЛ 2. МЕТОДОЛОГІЯ ТА ОЦІНКА ФОТОКОНСИСТЕНТНИХ КРИТЕРІЇВ ДЛЯ ГЕНЕРАЦІЇ ТРИВИМІРНОЇ ОБ'ЄМНОЇ МОДЕЛІ ОБ'ЄКТІВ..... | 30 |
| 2.1. Вступ до проблеми 3D-Реконструкції | 30 |
| 2.1.1. Конфігурації збору даних..... | 31 |
| 2.1.2. Постановка завдання реконструкції | 31 |
| 2.2. Огляд об'ємних методів | 32 |
| 2.2.1. Реконструкція візуальної оболонки | 32 |
| 2.2.2. Заповнення вокселів | 33 |
| 2.2.3. Метод фарбування вокселів | 35 |
| 2.3. Методологія та перевірка консистентності | 38 |
| 2.3.1. Основні етапи реалізації фарбування вокселів | 38 |

| | |
|--|-----------|
| 2.3.2. Метрики консистентності..... | 39 |
| 2.4. Представлення та опис методології фарбування вокселів..... | 41 |
| Висновки до розділу | 43 |
| РОЗДІЛ 3. ІМПЛЕМЕНТАЦІЯ МОДЕЛЕЙ ТА МЕТОДІВ ГЕНЕРАЦІЇ | |
| ТРИВИМІРНИХ МОДЕЛЕЙ..... | 45 |
| 3.1. Систематичний огляд генерації 3D-контенту | 45 |
| 3.1.1. Еволюція генеративних моделей у 2D-просторі | 45 |
| 3.1.2. Перехід до 3D-генерації | 46 |
| 3.2. Деталізація методологій 3D-генерації | 49 |
| 3.2.1. Метод прямої 3D генерації..... | 49 |
| 3.2.3. Процедурна 3D генерація..... | 52 |
| 3.2.4. Метод генеративного синтезу нових видів | 53 |
| 3.3. Представлення сцени та рендеринг у генерації 3d-моделей на основі штучного інтелекту..... | 54 |
| 3.3.1. Експліцитні представлення | 56 |
| 3.3.2. Застосування сіток | 57 |
| 3.3.3. Багатошарові представлення..... | 58 |
| 3.4. Методологія та еволюція генеративних моделей у задачах синтезу 3D-моделей та сцен..... | 59 |
| 3.5. Математичні парадигми імовірнісного моделювання та послідовного синтезу в 3D-генерації..... | 65 |
| 3.6. Генерація на основі оптимізації | 67 |
| 3.6.1. Текстово-орієнтована генерація..... | 68 |
| 3.6.2. Візуально-орієнтована генерація | 69 |
| 3.7. Методи синтезу тривимірних моделей з використанням генеративного моделювання антропоцентричних об'єктів | 69 |
| Висновки до розділу | 74 |
| ВИСНОВКИ | 75 |
| ПЕРЕЛІК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ..... | 78 |

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

3DGS — 3D Gaussian Splatting
3DMM — 3D Morphable Model
AE — Autoencoder
ELBO — Evidence Lower Bound
GAN — Generative Adversarial Network
GNVS — Generative Novel View Synthesis
ImAM — Improved Autoregressive Model
INR — Implicit Neural Representation
IoU — Intersection over Union
LPIPS — Learned Perceptual Image Patch Similarity
Mesh — Polygon Mesh
NeRF — Neural Radiance Field
PBR — Physically Based Rendering
PSNR — Peak Signal-to-Noise Ratio
SDF — Signed Distance Function
SDS — Score Distillation Sampling
SJC — Score Jacobian Chaining
SMPL — Skinned Multi-Person Linear model
SSIM — Structural Similarity Index Measure

ВСТУП

Актуальність теми.

Стрімкий розвиток цифрових технологій, комп'ютерної графіки та штучного інтелекту зумовлює зростаючий інтерес до задач автоматичної генерації тривимірних моделей на основі двовимірних зображень. Такі задачі є ключовими для широкого спектра прикладних галузей, зокрема комп'ютерного зору, віртуальної та доповненої реальності, цифрового дизайну, робототехніки, медицини, архітектури та ігрової індустрії. Перетворення візуальної інформації у повноцінні тривимірні представлення дозволяє значно підвищити рівень автоматизації процесів моделювання та аналізу об'єктів реального світу.

Генерація 3D-моделей із зображень є складною міждисциплінарною проблемою, що поєднує методи геометричного та імовірнісного моделювання, алгоритми оптимізації, теорію проєктивної геометрії та сучасні підходи машинного навчання. Класичні методи 3D-реконструкції, засновані на багатовидовому аналізі та фотоконсистентності, забезпечують геометричну коректність результатів, однак часто обмежені високою обчислювальною складністю та чутливістю до якості вхідних даних. Водночас генеративні моделі штучного інтелекту відкривають нові можливості для синтезу складних тривимірних форм, але потребують теоретичного узагальнення та методологічного обґрунтування.

У зв'язку з цим актуальним є комплексне дослідження моделей, методів та алгоритмів генерації тривимірних моделей із зображень, що охоплює як класичні підходи комп'ютерного зору, так і сучасні генеративні методології. Такий підхід дозволяє не лише систематизувати існуючі рішення, а й визначити перспективні напрями інтеграції геометричних і нейромережевих методів у єдину обчислювальну парадигму.

Актуальність даної магістерської роботи зумовлена зростаючою потребою в автоматизованих засобах створення тривимірного контенту з

мінімальним залученням людських ресурсів. Традиційні методи ручного 3D-моделювання є трудомісткими, потребують високої кваліфікації та значних часових витрат, що обмежує їх масштабованість у промислових і наукових застосуваннях. У цьому контексті методи генерації 3D-моделей із зображень виступають ефективною альтернативою, здатною суттєво оптимізувати процес створення цифрових об'єктів.

Сучасні задачі комп'ютерного зору дедалі частіше орієнтовані не лише на аналіз зображень, а й на відновлення повної просторової структури сцен. Це особливо важливо для систем автономної навігації, цифрових двійників, реконструкції культурної спадщини та медичної візуалізації. Водночас зростання обсягів візуальних даних потребує розробки методів, здатних ефективно працювати в умовах шуму, неповноти та неоднорідності вхідної інформації.

Додаткову актуальність темі надає активний розвиток генеративних моделей, які демонструють високий потенціал у задачах синтезу 3D-об'єктів і сцен. Однак їх інтеграція з класичними методами реконструкції потребує глибокого теоретичного аналізу та методологічного узгодження. Саме тому дослідження моделей, методів і алгоритмів генерації тривимірних моделей із зображень є своєчасним і науково обґрунтованим.

Метою магістерської роботи є дослідження, аналіз і узагальнення моделей, методів та алгоритмів генерації тривимірних моделей із зображень, а також обґрунтування ефективних підходів до 3D-реконструкції та генеративного синтезу тривимірних об'єктів.

Об'єктом дослідження є процес генерації та реконструкції тривимірних моделей на основі двовимірних зображень.

Предметом дослідження є моделі, методи та алгоритми 3D-реконструкції і генеративного синтезу тривимірних моделей.

Для досягнення поставленої мети у роботі необхідно **розв'язати такі завдання:**

1. Проаналізувати основи та сучасні підходи 3D-реконструкції із зображень.
2. Дослідити алгоритмічні етапи відновлення тривимірної структури за послідовністю зображень.
3. Розглянути математичний апарат, що використовується в задачах 3D-реконструкції.
4. Проаналізувати об'ємні методи генерації тривимірних моделей та фотоконсистентні критерії.
5. Дослідити методологію фарбування вокселів і метрики оцінювання якості реконструкції.
6. Проаналізувати способи представлення сцен і методи рендерингу у 3D-генерації.
7. Узагальнити перспективи інтеграції класичних і генеративних підходів.

Методи дослідження

У роботі використано такі методи дослідження:

- методи теоретичного аналізу та узагальнення наукових джерел;
- математичні методи проєктивної та епіполярної геометрії;
- алгоритмічні методи комп'ютерного зору;
- об'ємні методи 3D-реконструкції;
- імовірнісні та оптимізаційні методи генеративного моделювання;
- методи порівняльного аналізу та оцінювання якості 3D-моделей.

Наукова новизна роботи полягає у комплексному узагальненні класичних і сучасних генеративних підходів до побудови тривимірних моделей із зображень. У роботі систематизовано фотоконсистентні критерії в об'ємній 3D-реконструкції та проаналізовано їх роль у підвищенні якості моделей. Подальшого розвитку набули уявлення про інтеграцію геометричних методів із генеративними моделями штучного інтелекту в єдиному обчислювальному підході.

Практичне значення отриманих результатів полягає у можливості використання матеріалів роботи при розробці систем автоматизованої 3D-реконструкції та генерації контенту. Результати можуть бути застосовані у сфері комп'ютерної графіки, віртуальної та доповненої реальності, цифрового моделювання, навчальних і науково-дослідних проєктів.

Структура магістерської роботи. Представлена робота складається зі вступу, трьох розділів та висновків. Загальний обсяг роботи становить 80 сторінок, і містить 25 рисунків, 5 таблиць, 35 використаних джерел.

РОЗДІЛ 1. АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ГЕНЕРАЦІЇ 3D МОДЕЛЕЙ ТА СТРУКТУР ІЗ ЗОБРАЖЕНЬ

1.1. Теоретичні основи та сучасні підходи в комп'ютерному зорі щодо вилучення 3D-структури зі зображень

Сучасна комп'ютерна графіка забезпечує відтворення високореалістичних зображень тривимірних (3D) сцен. Проте генерація таких зображень вимагає наявності адекватних графічних моделей світу. Традиційне отримання цих моделей здійснювалося за допомогою пакетів 3D-моделювання, що є процесом, який характеризується високою трудомісткістю та обмеженим рівнем досяжної деталізації та фотореалізму.

Внаслідок цих обмежень, набуває поширення підхід, заснований на оцифруванні реального світу для генерації 3D-моделей. Серед методів оцифрування відомі лазерні далекоміри та аналогічні пристрої, які функціонують шляхом проєктування структурованого світла на об'єкт та подальшого отримання 3D-інформації через аналіз відбитого сигналу.

Альтернативний підхід полягає у використанні наборів зображень сцени або об'єктів для їхньої реконструкції. Зображення можуть бути використані не лише для вилучення текстурної інформації для 3D-моделі (текстурне картування), але й для вилучення її 3D-геометрії (форми). Саме останній аспект є центральним предметом цього дослідження.

Концептуальне розуміння цього процесу можна отримати, розглядаючи механізми людського сприйняття 3D-світу. Стереоскопічний зір забезпечує сприйняття глибини завдяки наявності двох точок спостереження (очей), які фіксують сцену з дещо відмінних ракурсів. Ця різниця (диспаратність) між зображеннями прямо пропорційна відстані до об'єктів.

Інший механізм, що дозволяє оцінювати 3D-структуру, ґрунтується на русі спостерігача (голови) та подальшому аналізі рухового паралаксу. Крім

того, переміщення у просторі забезпечує оцінку напрямку руху та структури середовища як одночасно виконуваних завдань.

У галузі комп'ютерного зору (Computer Vision) протягом десятиліть розроблялися алгоритми, які імітують подібні можливості для машин. Початковою метою була робототехніка та автоматизація (наприклад, навігація в невідомому середовищі). Останнім часом акцент змістився на застосування у візуалізації та комунікації, що сприяло тіснішій інтеграції зі спільнотою комп'ютерної графіки.

Ключовим напрямком є розробка алгоритмів, здатних автоматично вилучати необхідну 3D-інформацію з множини зображень. Важливі теоретичні здобутки, особливо в геометрії багатьох зображень (multi-view geometry), забезпечили створення більш гнучких та стійких підходів.

Слід зазначити, що подібні методи використовувалися значно раніше в контексті фотограмметрії. З другої половини XIX століття фотографічні знімки застосовувалися для картографування та вимірювання будівель, при цьому основний акцент робився на досягненні високої точності шляхом ретельного моделювання та калібрування процесу формування зображення і статистично коректного усунення похибок.

Спільнота комп'ютерної графіки також активно працювала над отриманням 3D-моделей на основі зображень, що є необхідною передумовою для візуалізації. Робота [2] є прикладом успішної розробки алгоритмів, які дозволяють отримувати високопереконливі 3D-моделі зі знімків з мінімальною інтерацією з користувачем.

1.2. Алгоритмічні основи 3D-реконструкції на основі послідовності зображень

Представлена робота інтегрує знання, методи та алгоритми з різних галузей для вирішення проблеми автоматичної генерації реалістичної тривимірної (3D) моделі на основі вхідної послідовності зображень об'єкта чи

сцени. Ця комплексна задача може бути декомпозиційна на низку менших, послідовних підзадач.

Розглянемо найпростіший випадок: реконструкція 3D-точки з її проєкцій на двох зображеннях. За умови відомого відносного руху камери та наявності моделі, що пов'язує пікселі зображення з променями проєкції, 3D-точка відновлюється шляхом обчислення перетину двох просторових променів, що їй відповідають. Цей процес називається триангуляцією.

Ключові передумови для триангуляції:

1. Відносний рух та калібрування камери.
2. Встановлення відповідних точок на зображеннях (кореспонденцій).

1.2.1. Етапи 3D-реконструкції з послідовності зображень

Першим необхідним кроком є відновлення відносного руху між послідовними зображеннями, яке здійснюється спільно з пошуком відповідних елементів зображення (тобто пікселів, що походять від одного 3D-елемента).

Надійне зіставлення (Matching):

- Використовується детектор елементів для виділення опорних точок, які мають високу відмінність від сусідніх пікселів (не підходять однорідні області або прямі краї).

- Початкові потенційні збіги, отримані порівнянням інтенсивностей, містять велику кількість помилкових кореспонденцій (outliers).

- Для ідентифікації правильного набору кореспонденцій використовується епіполярна геометрія — обмеження, яке накладається на пару відповідних точок просторовим положенням камер (обчислюється з мінімум 7 кореспонденцій).

- Набір узгоджених кореспонденцій визначається за допомогою ітеративної процедури RANSAC, яка оцінює найбільшу підмножину точок, що підтримують гіпотезу про епіполярну геометрію.

Цей процес дозволяє отримати початкову проєктивну реконструкцію.

Початкова реконструкція містить проєктивні спотворення (непаралельність ліній, некоректні кути/відстані) через відсутність попереднього калібрування.

Підхід базується на стратифікації геометрії, спочатку відновлюючи лише проєктивну структуру. Використовуючи додаткові обмеження, властиві реальним камерам, застосовується алгоритм самокалібрування. Усунення спотворення призводить до евклідової реконструкції, еквівалентної оригіналу з точністю до глобального масштабного коефіцієнта.

Перевагою є гнучкість процесу отримання даних, оскільки внутрішні параметри камери (наприклад, фокусна відстань) не вимагають попереднього вимірювання і можуть змінюватися під час зйомки.

Реконструкція, отримана, як описано в попередньому абзаці, містить лише розріджений набір 3D-точок (спочатку розглядається лише обмежена кількість елементів). Хоча інтерполяція може бути рішенням, це зазвичай призводить до моделей із низькою візуальною якістю. Тому наступний крок полягає у спробі зіставити всі пікселі зображення з пікселями на сусідніх зображеннях, щоб ці точки також можна було реконструювати. Це завдання значно полегшується знанням усіх параметрів камери, які ми отримали на попередньому етапі. Оскільки піксель на зображенні відповідає променю в просторі, а проєкція цього променя на інших зображеннях може бути передбачена з відновленої пози та калібрування, пошук відповідного пікселя на інших зображеннях може бути обмежений однією лінією. Додаткові обмеження, такі як припущення про неперервну 3D-поверхню (поділену на частини), також застосовуються для подальшого обмеження пошуку. Можна викривити зображення так, щоб діапазон пошуку збігався з горизонтальними рядками сканування. Алгоритм, який може досягти цього для довільного руху камери. Це дозволяє нам використовувати ефективний стереоалгоритм, який обчислює оптимальне зіставлення для всього рядка сканування одночасно. Таким чином, ми можемо отримати оцінку глибини (тобто відстані від камери до поверхні об'єкта) майже для кожного пікселя

зображення. Шляхом поєднання результатів усіх зображень разом отримується повна щільна модель 3D-поверхні. Зображення, використані для реконструкції, також можуть бути використані для текстурування, щоб досягти остаточного фотореалістичного результату. Різні етапи процесу проілюстровані на рисунку 1.1. У наступних абзацах деякі з критичних кроків описані більш детально.

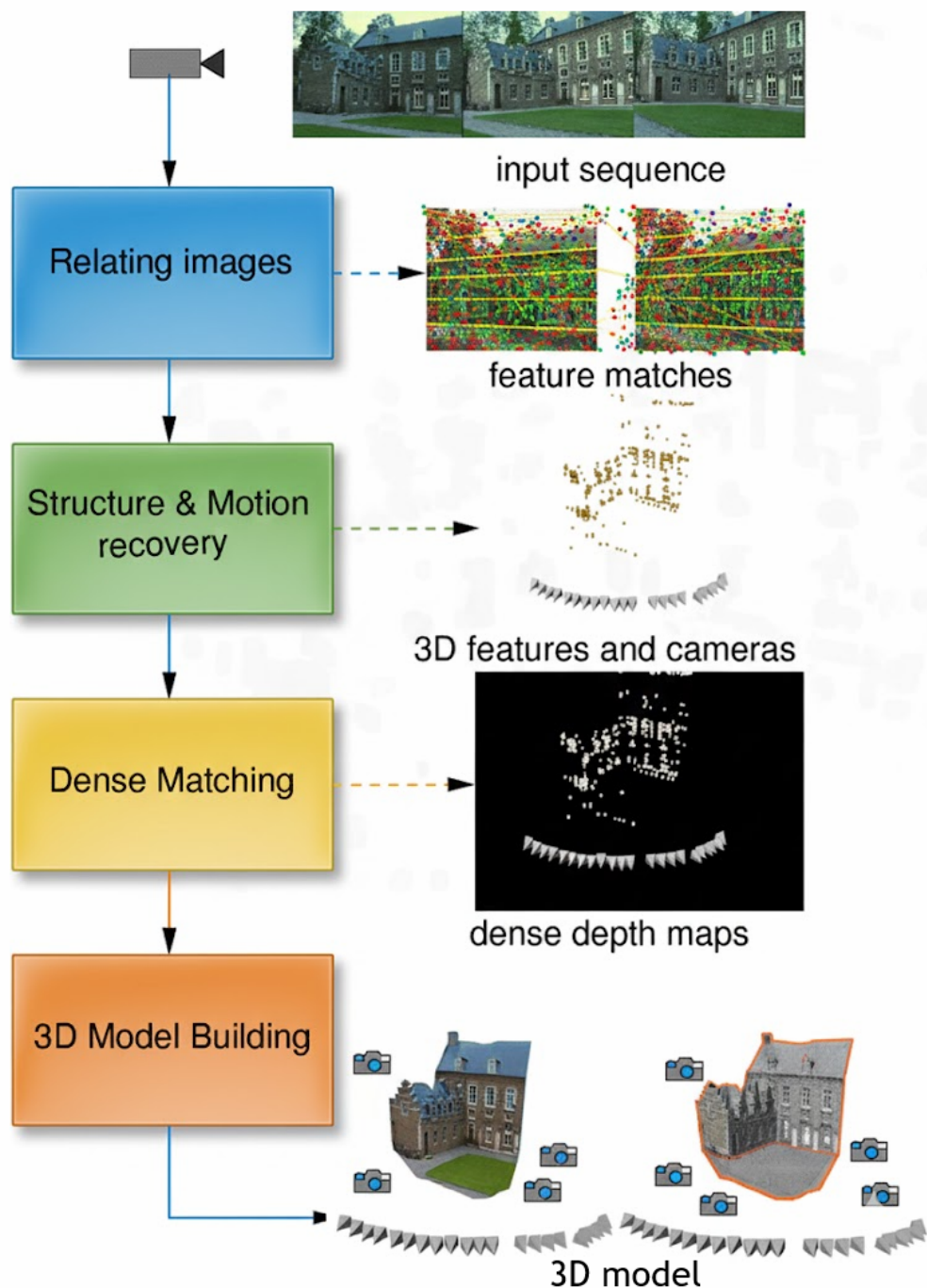


Рис. 1.1. Огляд підходу до 3D-запису на основі зображень

Перша важлива складність полягає в автоматичному отриманні початкових збігів із реальних зображень. Для комп'ютера зображення — це лише велика сукупність значень інтенсивності пікселів. Щоб знайти відповідні точки на різних зображеннях, можна порівняти значення інтенсивності на невеликій області навколо точки. Однак не всі точки придатні для такого типу порівняння. Коли точка не може бути відрізнена від власних сусідів, також не можна очікувати знайти унікальний збіг із точкою на іншому зображенні. Тому точки в однорідних областях або розташовані на прямих краях не підходять для зіставлення на цьому етапі. Метод, який ми використовуємо, витягує набір точок за допомогою детектора елементів, який максимізує відмінність від сусідніх пікселів. Як правило, ми прагнемо виділити 1000 опорних точок, добре розподілених по зображенню, для кожного зображення. Потенційні збіги отримуються шляхом порівняння опорних точок із подібними координатами. Це обмеження згодом послаблюється на другому етапі. Звичайно, багато з цих початкових збігів можуть бути неправильними, і нам потрібна надійна процедура для роботи з ними. Це можна зробити, оскільки набір правильних збігів має задовольняти деяким обмеженням. Оскільки 3D-точка має лише три ступені свободи, а пара точок зображення має чотири, має бути обмеження для кожної пари збіжних точок (з точністю до деяких ступенів свободи, пов'язаних із відносною позою камери). Ця внутрішня структура набору збігів, що називається епіполлярною геометрією, може бути обчислена з 7 або більше збігів точок. Використовуючи процедуру RANSAC, обчислюється найбільший набір узгоджених збігів. Алгоритм генерує гіпотезу з набору 7 довільних збіжних точок і визначає, яка частка інших збігів підтримує гіпотезу. Цей крок повторюється доти, доки ймовірність того, що була згенерована правильна гіпотеза, не перевищить певний поріг. Це дозволяє нам надійно працювати з більш ніж половиною початкових збігів, які є неправильними. Як тільки знайдено правильне рішення, його можна уточнити та використовувати для керування пошуком додаткових збігів.

Водночас відновлена епіполярна геометрія дозволяє нам генерувати початкову проєктивну реконструкцію. Зауважимо, що цей підхід передбачає, що спостережувана сцена є жорсткою.

1.2.2. Уникнення процесів необхідності калібрування

Традиційна робота з 3D-реконструкції із зображень вимагає калібрування камер. Одним із ключових усвідомлень нашого підходу було послаблення цього обмеження. Окрім значного виграшу в гнучкості, одна з переваг полягає в тому, що 3D-реконструкцію також можна отримати з архівних матеріалів або з аматорських відеозображень, записаних із змінним масштабуванням. Уникнення калібрування базується на стратифікації геометрії. Спочатку відновлюється лише проєктивна структура сцени та геометрія камери. Це можна зробити незалежно від калібрування. Перспективна модель камери, що використовується на цьому рівні, є занадто загальною, і доступна низка додаткових обмежень. Вони дозволяють відновити евклідову структуру (з точністю до масштабного коефіцієнта).

Після отримання 3D-моделі одним із основних застосувань, окрім вимірювання, є відтворення нових ракурсів записаної сцени. Однак, якщо це мета, доступні альтернативи явному 3D-моделюванню. Нещодавно були запропоновані суто методи на основі зображень, такі як рендеринг світлового поля. У цьому випадку записані зображення розглядаються як велика сукупність 3D-світлових променів, з яких вибираються найбільш підходящі для заповнення пікселів нового зображення. Як це можна ефективно досягти — використовуючи обчислені нами дані — описано в [4]. Перевага такого типу підходу полягає в тому, що ефекти, залежні від ракурсу, легко відтворюються. Однак, з іншими ефектами легше працювати, маючи явні 3D-моделі.

Описаний вище підхід до 3D-моделювання має два типи застосування:

- 1) 3D-модель може бути використана для вимірювань;
- 2) 3D-модель може бути використана для візуалізації.

Більшість реальних застосувань використовують обидва аспекти. Завдяки вартості та гнучкості цей тип підходу має багато цікавих застосувань у різних сферах. Тут ми лише виділимо два застосування.

Перше цікаве застосування — це запис археологічних пам'яток та об'єктів, щоб їх можна було доповнити віртуальними реконструкціями, заснованими на археологічних гіпотезах. Однак археологія має набагато більше потреб. Запис достатньої кількості вимірювань є дуже важливим у галузі археології, оскільки докази постійно руйнуються під час розкопок. У цьому контексті дешева та гнучка техніка 3D-запису є дуже вигідною. Розкопані блоки можуть бути змодельовані, щоб можна було віртуально перевірити будівельні гіпотези. Стає можливим навіть записати весь процес розкопок у 3D. Кілька прикладів показано на рисунку 1.2.



Рис. 1.2. 3D-запис культурної спадщини

Одне зображення (рис. 1.2) з відеопослідовності та отримана 3D-модель (вгорі-ліворуч), вигляд зверху 3D-запису розкопок (вгорі-праворуч), статуя Діоніса віртуально переміщена з музею назад на своє початкове місце

розташування (внизу-ліворуч) та один кадр відео, де архітектор представляє віртуальну реконструкцію стародавньої пам'ятки.

Для деяких застосувань відновлені параметри камери також дуже важливі. Хорошим прикладом є вставлення віртуальних об'єктів у реальну відеопослідовність. У цьому випадку важливо, щоб вставлений об'єкт зазнав того самого відносного руху відносно камери. Сьогодні описані в цьому розділі методи використовуються для створення спецефектів (внизу ліворуч на рис. 1.2). Обробка виконується в автономному режимі. У майбутньому цей тип обчислень також стане можливим у реальному часі, що дозволить гнучку доповнену реальність на основі зору, і інші аспекти, такі як тіні, оклюзія та освітлення, будуть автоматично враховані.

Отримання 3D-моделей із зображень можливе. Виявляється, що це завдання може бути навіть виконане автоматично комп'ютером. Застосування для цього типу технологій можна знайти в багатьох різних сферах, серед яких лише кілька: археологія, архітектура, судова експертиза, геологія, планетарні дослідження, електронна комерція, кіноспецефекти, віртуальна та доповнена реальність тощо. Важливими перевагами цього типу підходу порівняно з іншими є гнучкість і нижча вартість.

1.3. Математичний апарат у 3D-реконструкції

Розглянемо детальніше дві ключові концепції: епіполярну геометрію та алгоритм RANSAC.

1.3.1. Епіполярна геометрія (Epipolar Geometry)

Епіполярна геометрія описує внутрішні геометричні співвідношення між двома зображеннями однієї й тієї ж 3D-сцени, отриманими з різних позицій. Вона є фундаментальним обмеженням, яке дозволяє визначити, чи відповідають дві точки на різних зображеннях одній 3D-точці.

Ключові елементи:

1. Центри проєкції (Camera Centers, C_1, C_2) - позиції об'єктивів камер, звідки були зроблені знімки.

2. Епіполус (Epipole, e_1, e_2) - точка на одному зображенні, де проєціюється центр проєкції іншої камери. Це точка, в якій лінія, що з'єднує центри камер, перетинає площину зображення.

3. Епіполярна площина (Epipolar Plane) - площина в 3D-просторі, що проходить через 3D-точку P та обидва центри проєкції (C_1 і C_2).

4. Епіполярна лінія (Epipolar Line) - лінія перетину епіполярної площини з площиною зображення. Всі можливі проєкції 3D-точки P на другому зображенні лежать на його епіполярній лінії.

Принцип обмеження: якщо точка p_1 на першому зображенні відповідає 3D-точці P , то відповідна точка p_2 на другому зображенні обов'язково повинна лежати на епіполярній лінії l_2 , яка генерується p_1 та e_2 . Це значно скорочує простір пошуку кореспонденцій.

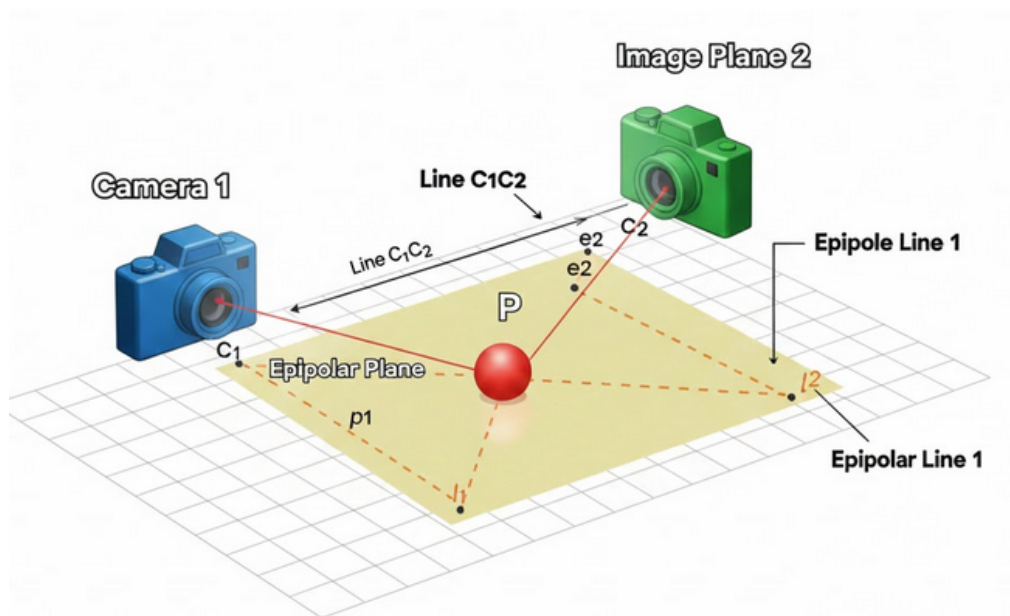


Рис. 1.3. Графічна інтерпретація методу епіполярної геометрії

Геометричне обмеження виражається математично за допомогою фундаментальної матриці (F). F — це несингулярна матриця розміром 3×3 ,

яка інкапсулює епіполярне співвідношення між двома некаліброваними зображеннями.

Якщо p_1 та p_2 — однорідні координати відповідних точок на першому та другому зображеннях, то їхнє співвідношення задається рівнянням:

$$p_2^T F p_1 = 0$$

Це рівняння є епіполярним обмеженням. Воно стверджує, що вектор p_2 лежить на епіполярній лінії, визначеній F і p_1 .

Якщо внутрішні параметри камер (калібрування) відомі, використовується основна матриця (E). E пов'язана з F через матриці внутрішнього калібрування K_1 та K_2 :

$$F = K_2^{-T} E K_1^{-1}$$

У підході, описаному вище, використовується самокалібрування, тому F оцінюється першою (з некаліброваних даних), і лише потім, з додатковими геометричними обмеженнями, відновлюється E та параметри K_1, K_2 .

1.3.2. Алгоритм RANSAC (RANdom SAmple Consensus)

RANSAC — це не детермінований ітераційний алгоритм для оцінки математичної моделі за даними, які містять велику кількість викидів (outliers).

У контексті 3D-реконструкції, він використовується для надійної оцінки фундаментальної матриці (F), оскільки початковий пошук кореспонденцій завжди генерує багато неправильних збігів.

Алгоритм RANSAC працює на припущенні, що дані складаються з:

1. Внутрішніх елементів (Inliers) - дані, які можна пояснити деякою моделлю з деяким допуском шуму.

2. Викидів (Outliers) - дані, які не вписуються в модель (у нашому випадку, неправильні кореспонденції).

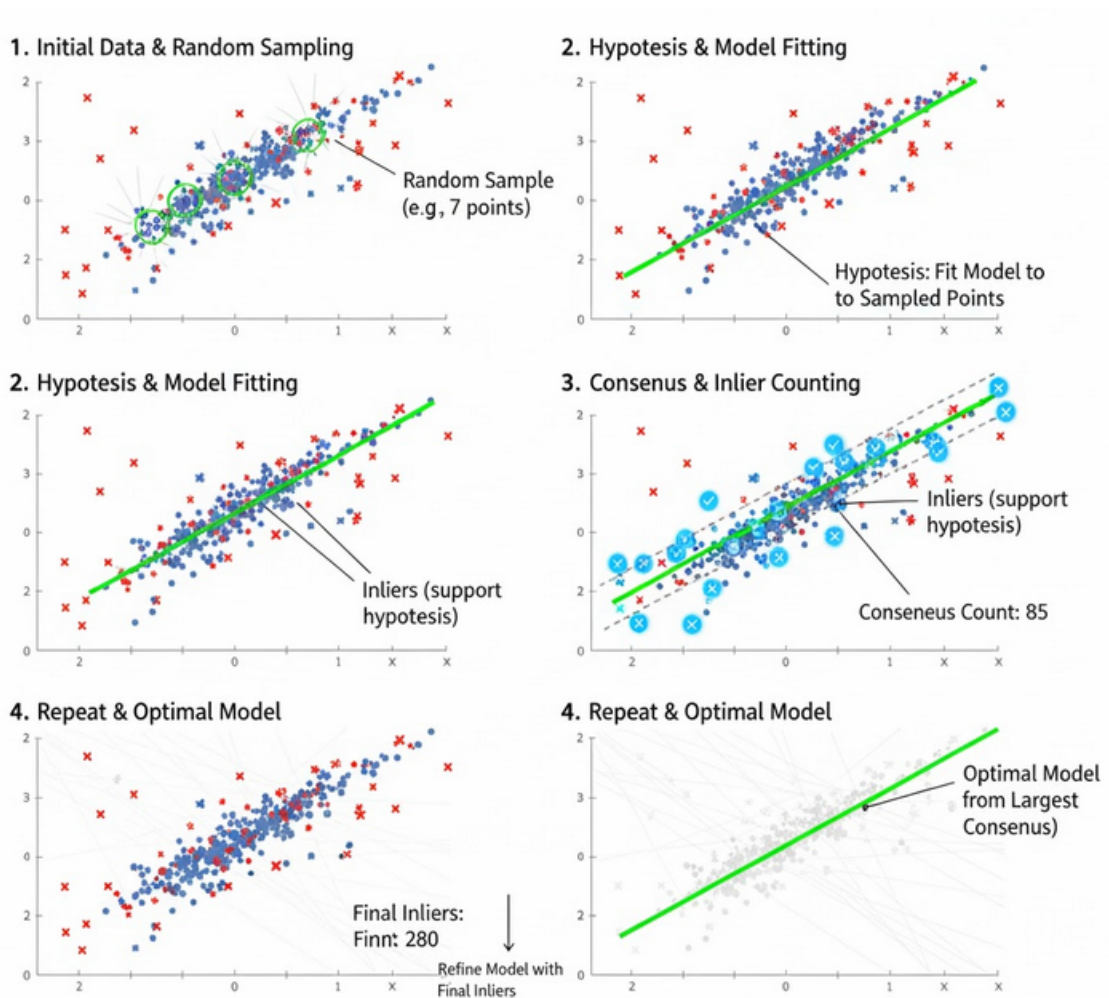


Рис. 1.4. Приклади реалізацій алгоритму RANSAC

Етапи алгоритму наступні.

Вхідні дані: набір можливих кореспонденцій (збігів), мінімальна кількість точок для оцінки моделі (m), та поріг узгодженості (τ).

1. Вибірка (Sampling) - випадковим чином обирається мінімальна підмножина даних. Для оцінки матриці F потрібно $m=7$ пар відповідних точок (7-точковий алгоритм).

2. Генерація гіпотези (Hypothesis Generation) - на основі цієї мінімальної підмножини обчислюється модель (тобто, оцінюється матриця F).

3. Перевірка (Testing) - вся множина вхідних даних (всі кореспонденції) тестується на відповідність обчисленій моделі. Точка p_2 вважається такою, що підтримує модель, якщо вона задовольняє епіполярному обмеженню: $|p_2^T F p_1| < \tau$, де τ — порогове значення.

4. Підрахунок узгодженості (Consensus Count) - підраховується кількість внутрішніх елементів (кореспонденцій), які підтримують згенеровану модель.

5. Повторення - кроки 1-4 повторюються N разів. Кількість ітерацій N обчислюється виходячи з бажаної ймовірності успіху та очікуваної частки викидів.

Після завершення ітерацій, обирається модель (F), яка має найбільшу кількість узгоджених збігів (найбільшу підмножину внутрішніх елементів). Ця найбільша підмножина використовується для фінального уточнення моделі (зазвичай, методом найменших квадратів), щоб отримати найбільш точну оцінку F .

Отже, RANSAC забезпечує надійність, дозволяючи комп'ютеру ігнорувати більшість неправильних збігів, зосереджуючись на пошуку максимальної внутрішньо узгодженої геометрії.

1.3.3. Опис методології

Представлений рисунок 1.5 зображення ілюструє фундаментальний концепт епіполярної геометрії (Epipolar Geometry), який є наріжним каменем у більшості методів багатовидового (мультивидового) комп'ютерного зору, включаючи тривимірну реконструкцію та стереовідповідність. Ця геометрія описує внутрішні геометричні співвідношення між двома зображеннями однієї й тієї ж сцени, отриманими з різних позицій (камер).

1. Просторова конфігурація та основні елементи

Система складається з двох камер (Camera 1 та Camera 2), кожна з яких має свій оптичний центр (C_1 та C_2) та площину зображення (Image Plane 1 та Image Plane 2).

- тривимірна точка: нехай P — це довільна тривимірна точка у просторі, яку ми прагнемо реконструювати.
- проєкції точки: точка P проєктується на площину першої камери у вигляді пікселя p_1 та на площину другої камери у вигляді пікселя p_2 .
- базова лінія (Line C_1C_2): це пряма, що з'єднує оптичні центри обох камер.
- епіполярна площина (Epipolar Plane): Це площина, що визначається оптичними центрами обох камер (C_1, C_2) та тривимірною точкою P . Ця площина містить усі три елементи.

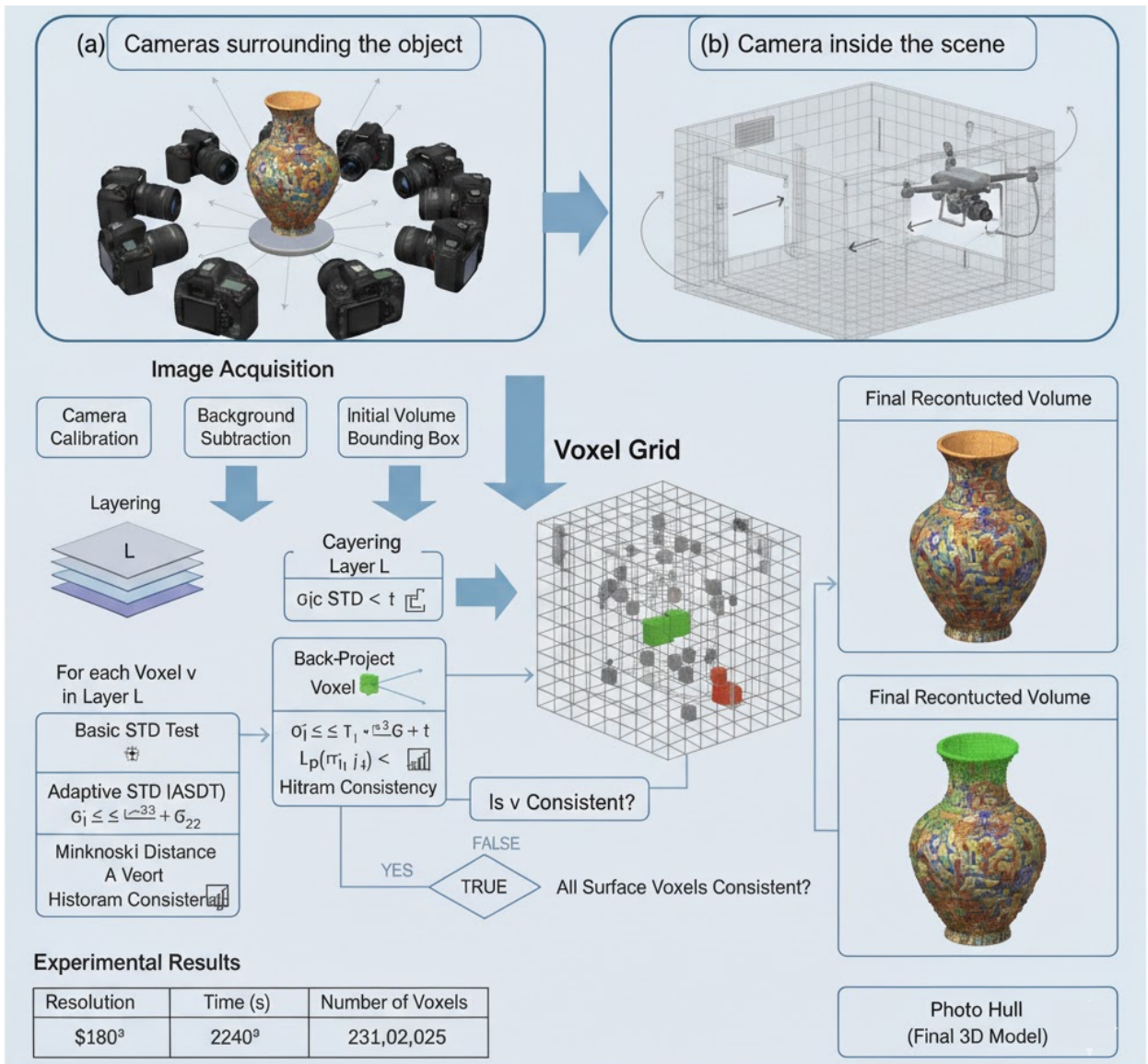


Рис. 1.5. Методологія 3D реконструкцій

2. Епіполлярні обмеження

Ключова ідея епіполлярної геометрії полягає у спрощенні процесу пошуку відповідних точок (matching problem).

- Епіполь e_2 на другій площині зображення — це проєкція оптичного центра першої камери C_1 на площину другої камери. Аналогічно, e_1 — проєкція C_2 на площину першої камери. Епіполі є точками перетину базової лінії C_1C_2 з відповідними площинами зображення.

- Епіполярна Лінія (Epipolar Line l_1, l_2) - це лінії перетину епіполлярної площини з площинами зображення. Лінія l_1 — це проєкція променя C_2P на Image Plane 1, а l_2 — це проєкція променя C_1P на Image Plane 2.

Головне обмеження. Якщо відома проєкція p_1 тривимірної точки P на першому зображенні, то відповідна точка p_2 на другому зображенні мусить лежати на епіполярній лінії l_2 . Ця лінія l_2 однозначно визначається точкою p_1 .

Це обмеження кардинально знижує обчислювальну складність, перетворюючи двовимірний пошук відповідності (2D search) на одновимірний (1D search) вздовж епіполярної лінії.

3. Математичний апарат: фундаментальна матриця

Епіполярне обмеження математично виражається за допомогою фундаментальної матриці. Таким чином, епіполярна геометрія забезпечує строгі геометричні обмеження, необхідні для точного відновлення тривимірних координат точки P (процес, відомий як триангуляція) на основі відповідних пікселів.

Висновки до розділу

У першому розділі здійснено системний аналіз теоретичних та алгоритмічних основ генерації тривимірних моделей із зображень у контексті сучасних підходів комп'ютерного зору. Розглянуто базові принципи

вилучення просторової інформації зі сцен, що формують концептуальний фундамент задач 3D-реконструкції. Детально проаналізовано процес реконструкції за послідовністю зображень, включно з етапами виявлення ознак, їх зіставлення та відновлення геометричної структури сцени. Показано, що коректність побудови тривимірної моделі значною мірою залежить від точності оцінювання параметрів камер. Обґрунтовано актуальність підходів, спрямованих на уникнення явної калібрування, які підвищують гнучкість і практичну застосовність методів. Значну увагу приділено математичному апарату 3D-реконструкції, зокрема епіполлярній геометрії як основі взаємозв'язку між зображеннями. Розкрито роль алгоритму RANSAC у забезпеченні стійкості розрахунків за наявності шуму та помилкових відповідностей. Показано, що поєднання геометричних і статистичних методів є необхідною умовою надійної реконструкції. Узагальнення теоретичних положень дозволило сформулювати цілісне уявлення про сучасний стан предметної області. Отримані результати створюють основу для подальшого аналізу та реалізації об'ємних методів генерації 3D-моделей.

РОЗДІЛ 2. МЕТОДОЛОГІЯ ТА ОЦІНКА ФОТОКОНСИСТЕНТНИХ КРИТЕРІЇВ ДЛЯ ГЕНЕРАЦІЇ ТРИВИМІРНОЇ ОБ'ЄМНОЇ МОДЕЛІ ОБ'ЄКТІВ

2.1. Вступ до проблеми 3D-реконструкції

Тривимірна (3D) реконструкція є фундаментальним процесом у галузі комп'ютерного зору, спрямованим на відновлення просторової інформації про об'єкт виключно на основі його двовимірних (2D) зображень. Комп'ютерний зір займається автоматизованим вилученням різноманітних видів інформації із зображень, а його кінцева мета полягає у забезпеченні здатності машин візуально сприймати світ та ефективно взаємодіяти з ним. Одним із конкретних завдань є створення точних комп'ютерних моделей 3D-об'єктів на основі їхніх цифрових проєкцій.

Традиційні підходи в обробці зображень, оперуючи виключно 2D-даними, неминуче втрачають критичний параметр — глибину об'єкта. Інформація про глибину є вкрай важливою і відіграє вирішальну роль у багатьох прикладних рішеннях. Таким чином, основна мета 3D-реконструкції полягає у відновленні цієї глибинної інформації із простих 2D-зображень об'єкта. Результатом процесу є 3D-модель, яку можна зберігати, модифікувати та використовувати у подальшому аналізі. Завдання реконструкції форми 3D-об'єктів із серії цифрових зображень залишається складним і багатограним.

При використанні систем автоматизованого проєктування (CAD) для 3D-моделювання моделювальнику доводиться надавати велику кількість описів через графічний інтерфейс (GUI), включаючи текстові описи та визначення полігональних патчів. Такий підхід характеризується низьким ступенем реалізму та вимагає значних людських ресурсів. Тому розвиток автоматизованого моделювання, що використовує зображення, отримані за допомогою ПЗЗ-камер, є критично важливим.

2.1.1. Конфігурації збору даних

Було запропоновано низку конфігурацій для 3D-реконструкції:

1. Однокамерні системи: методи класифікуються залежно від того, чи є рухомим об'єкт, чи камера. У випадках, коли об'єкт рухомий, його зазвичай розміщують на поворотному столі, і здійснюється серія знімків при зміні кута повороту столу. Для успішної реконструкції необхідно мати інформацію про калібрування камери та її положення.

2. Багатокамерні системи: використовується кілька камер для збору різних ракурсів об'єкта. У цьому сценарії для кожної камери необхідна інформація про внутрішнє калібрування та зовнішні параметри.

Методи також поділяються на:

- Активні методи: Використовують зовнішнє джерело для виділення форми об'єкта (наприклад, структуроване світло або масив лазерних променів) у поєднанні з одним пристроєм захоплення, як правило, ПЗЗ-камерою.

- Пасивні методи: Базуються на звичайному освітленні та використанні камери для захоплення ракурсів.

2.1.2. Постановка завдання реконструкції

Дане дослідження зосереджено на реконструкції текстурованих об'єктів. Об'єкт дискретизується та реконструюється в кілька етапів, кожен з яких супроводжується специфічними викликами.

Першим кроком є збір зображень. Ми використовували однокамерну систему з об'єктом на обертовому столі. Це вимагає калібрування камери для визначення внутрішніх параметрів та параметрів спотворення, які потім використовуються для визначення зовнішніх параметрів (положення) камери.

Основний підхід об'ємної реконструкції використовує концепцію вокселів (елементів об'єму). При роботі з вокселями вимога до пам'яті є значною, оскільки воксельна сітка потребує величезного обсягу сховища.

Отже, необхідно визначити воксельний куб, який би оптимально апроксимував область уваги.

Використання кольорових зображень робить доцільним застосування інформації про колір для реконструкції форми об'єкта. Традиційні методи перетину об'ємів можуть включати деякі зайві вокселі. Видалення цих зайвих елементів для отримання кращої реконструкції є важливим завданням. Обчислення видимості вокселя з певного ракурсу є критичним, оскільки лише тоді зображення з цього ракурсу може бути використане як додаткова інформація для перевірки консистентності. Видимими є лише вокселі, розташовані на поверхні. Проблемою при обчисленні видимості є точне визначення напрямку, куди спрямований розглянутий воксель (його нормалі).

2.2. Огляд об'ємних методів

Зі зростанням обчислювальної потужності та ємності вторинного сховища, об'ємні підходи (Volume-based approaches) стали більш реалістичними для практичної реалізації.

2.2.1. Реконструкція візуальної оболонки

Раніше більшість методів вирішували завдання об'ємної реконструкції шляхом апроксимації візуальної оболонки (Visual Hull). Візуальна оболонка визначається як максимально можлива форма, яка генерує той самий силует, що й оригінальний об'єкт. Збільшення кількості 2D-вхідних даних зменшує розмір візуальної оболонки, покращуючи апроксимацію. В роботі [4] детально обговорювали цю концепцію.

Процедура:

1. Сегментація зображення (відокремлення переднього плану від фону).
2. Зворотне проєктування сегментованого силуету з кожного зображення.

3. Перетин цих зворотних проєкцій для генерації кінцевого об'єму.
4. Подальша обробка для відновлення кольорової інформації.

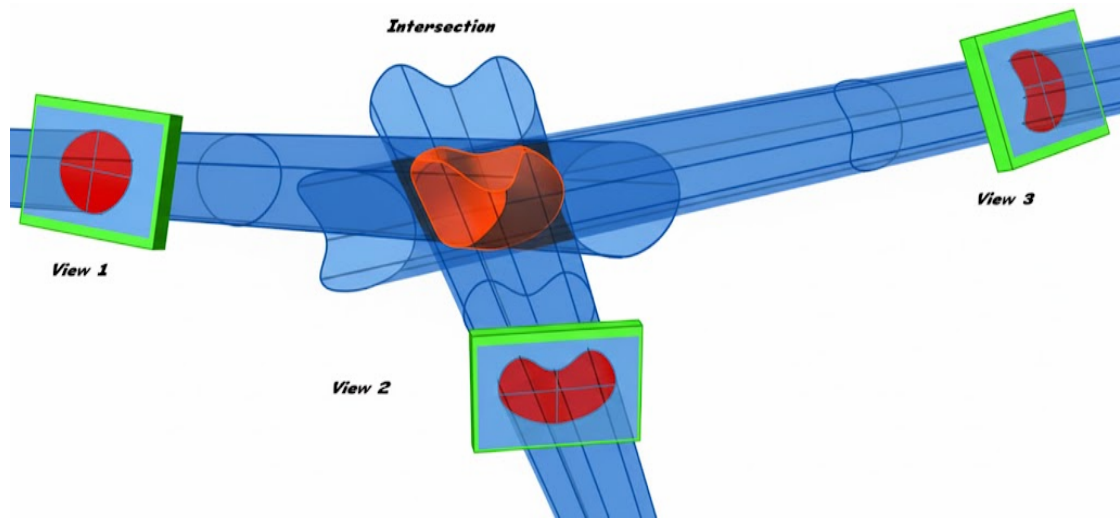


Рис. 2.1. Приклад реконструкції візуальної оболонки

Алгоритми на основі візуальної оболонки є швидкими з точки зору обчислень. Однак, наявність увігнутостей в об'єкті та недостатня кількість ракурсів призводить до створення надлишкового об'єму. Ці алгоритми є детермінованими і не враховують невизначеності у вхідних даних, а також чутливі до помилок сегментації. Помилка сегментації в одному ракурсі може призвести до значних спотворень у реконструйованій 3D-моделі [9].

2.2.2. Заповнення вокселів

Заповнення вокселів є добре відомим методом для визначення 3D-форми [6]. У цьому підході об'єм об'єкта визначається набором вокселів, які разом формують його форму.

Послідовність дій:

1. Ініціалізація обмежувального прямокутника.
2. Поділ прямокутника на маленькі вокселі.
3. Вокселізація - перевірка кожного вокселя за допомогою кількох ракурсів, щоб визначити, чи знаходиться він всередині, на межі, чи ззовні об'єкта.

Методи, засновані на заповненні вокселів, переважно використовують силуети (форма з силуету). Силует обчислюється для кожної камери (шляхом віднімання фонового зображення). Воксель вважається частиною об'єму, що реконструюється, якщо він потрапляє всередину перетину зворотних проєкцій усіх силуетів. Для представлення цих вокселів можуть використовуватися октодеревя.

Силуети виступають як жорсткі обмеження. Проте, помилки в обчисленні силуету (наприклад, збіг значень пікселів об'єкта та фону) можуть спричинити різкі зміни у реконструйованому об'єкті та утворення отворів. Для мінімізації цих артефактів можуть застосовуватися морфологічні оператори.

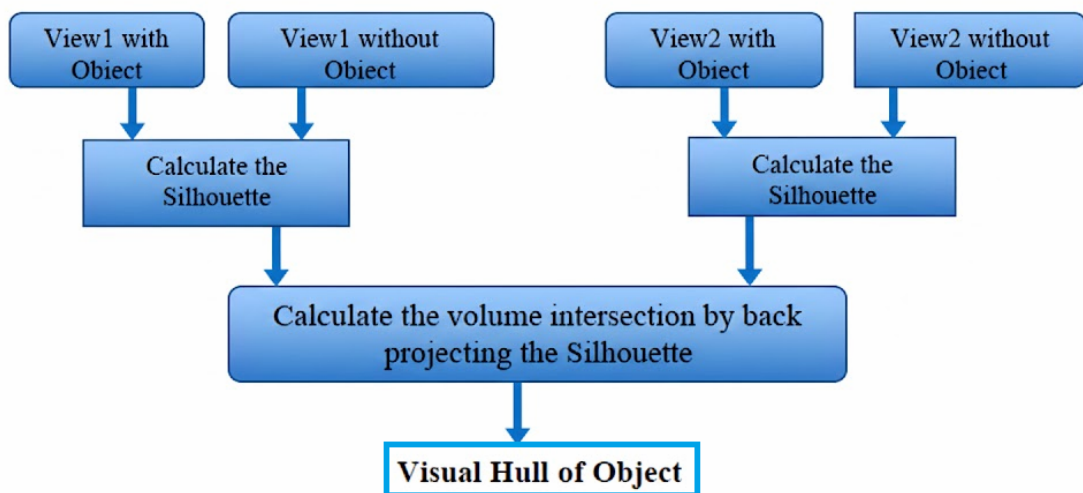


Рис. 2.2. Огляд форми з силуету

Вхідними даними є зображення з каліброваних камер у поєднанні з відповідними фоновими зображеннями. Вихідним результатом є бінарно маркований об'єм, де воксель позначається як 1, якщо він зайнятий, і 0, якщо він порожній. В роботі [6] запропонували метод, що використовує функцію мінімізації енергії:

$$E \sum_{(l)=0,0} 2_i(f_\nu) + i \sum_{y \in W' \in W} [(1 - \delta(f_\nu) - f_\nu)]$$

Тут f_v — мітка вокселя v , C_2 — штраф за присвоєння мітки $\delta(x)$ — одинична імпульсна функція.

2.2.3. Метод фарбування вокселів

Фарбування вокселів — це метод, де кожен воксель маркується своїм кольором та прозорістю. В основі методу лежить принцип консистентності кольору (фотоконсистентності).

Основний механізм:

- Якщо воксель v має ідентичний колір на всіх зображеннях, з яких він є видимим, він маркується цим кольором.

- Якщо колір вокселя різний у різних ракурсах, він оголошується прозорим (тобто, видаляється).

Алгоритм фарбування вокселів вимагає накладення обмеження на розташування камер: усі камери повинні бути розташовані таким чином, щоб вокселі можна було топологічно відсортувати та обробити за один прохід. Це гарантує, що прозорість усіх потенційно закриваючих вокселів була обчислена до обробки поточного [14]. Існує два типи установок для сканування поверхні:

а) камери оточують об'єкт;

б) камера знаходиться всередині сцени.

Алгоритм наголошує на виявленні того, чи відповідає проєкція точки на зображенні пікселю, що лежить на поверхні об'єкта. Для цього обчислюється фотоконсистентність точки шляхом порівняння світла, що надходить від цієї точки на різних зображеннях. За припущенням про ламбертівську поверхню, точка на поверхні повинна мати однаковий колір у різних ракурсах, якщо вона не закрита. Якщо світло, що надходить від будь-якої незакритої частини, не є фотоконсистентним, точка повинна бути позначена як порожня.

Обмежувальний прямокутник ділиться на шари (на основі відстані від камери). Вокселі обробляються шар за шаром:

1. Обчислення видимості - спочатку обчислюється видимість вокселя, і вокселі маркуються.

2. Перевірка консистентності - для кожного видимого вокселя перевіряється фотоконсистентність. Якщо він консистентний, він позначається; інакше він видаляється.

Оптимізація. Через зовнішнє та самозакльовування не всі вокселі видно на всіх зображеннях. Вокселі, які не знаходяться на поверхні, не будуть видимі на жодному зображенні. Видалення не поверхневих вокселів може значно зменшити обчислювальну складність. Поверхневі вокселі можуть бути ідентифіковані як ті, у яких принаймні один із шести сусідніх вокселів не належить об'єму об'єкта [16].

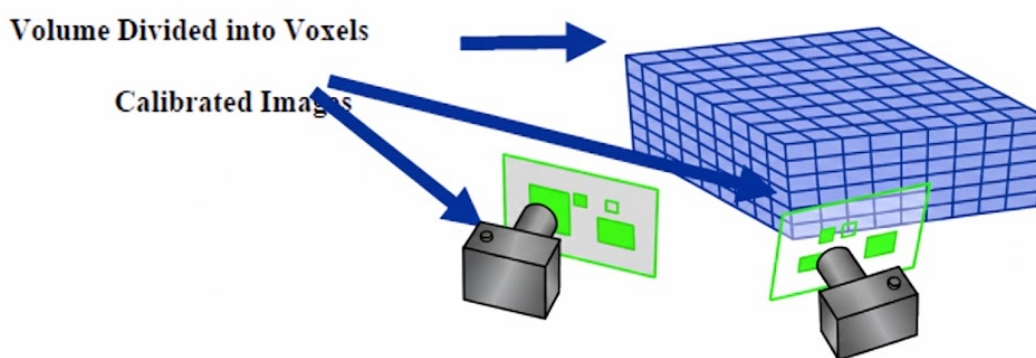
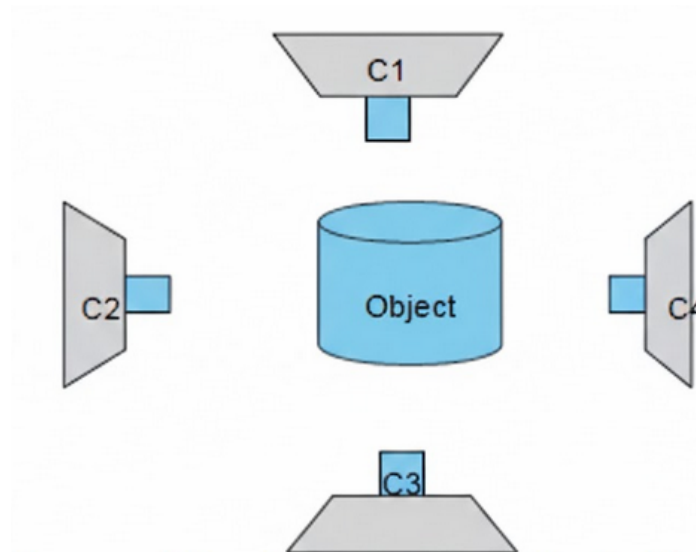


Рис. 2.3. Огляд процесу фарбування вокселів

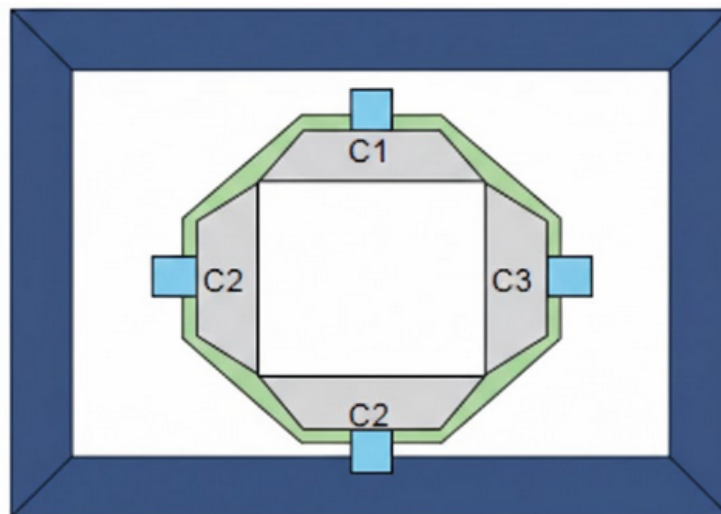
Оскільки більше одного вокселя можуть проектувати одну й ту саму область на зображенні, найближчий до камери воксель приписується кольору пікселя. Карта глибини використовується для позначення того, який воксель знаходиться ближче до камери; глибина вокселя v від зображення j позначається як $Depth_j(v)$.

Результатом є модель сцени, яка є консистентною з усіма пікселями на всіх вхідних ракурсах. Процес видаляє всі непрозорі вокселі, доки кожен граничний воксель не стане фотоконсистентним. Найближчий фотоконсистентний воксель уздовж кожного візуального променя гарантовано знаходиться на поверхні кінцевого об'єму. Кінцевий об'єм можна

розглядати як об'єднання всіх фотоконсистентних реконструкцій сцени, і з цієї причини він називається фотооболонкою (photo hull) [17].



а) Збір зображень, коли камери оточують об'єкт



б) Збір зображень, коли камера знаходиться всередині сцени

Рис. 2.3. Схеми розміщення камер для збору зображень

Алгоритм висікання вокселів (Voxel Carving) можна узагальнити наступним чином [17]:

Крок 1: Ініціалізувати V (об'єм, що містить кінцеву реконструкцію).

Крок 2: Поки $\text{consistent}(v) = \text{true}$; де $v \in \text{Surface}(V)$

- а. Обчислити всі фотографії, з яких v видимий. Нехай це множина P_v .
- б. Спроектувати v на всі фотографії в P_v .
- в. Зберегти піксельні кольори $C[1], \dots, C[j]$, на які проектується v , та обчислити оптичні промені $R[1], \dots, R[j]$, які з'єднують v з відповідними оптичними центрами.
- г. Визначити фотоконсистентність v за допомогою `calculate_consist(C, R)`.

Крок 3:

Якщо `consistent(v) = true`; v , де $v \in \text{Surface}$ встановити $V^* = V$ і завершити.

Інакше встановити $V = V - \{v\}$ і повторити крок 2.

Основна проблема з висіканням вокселів полягає у складності визначення оптимального критерію зупинки обчислення.

2.3. Методологія та перевірка консистентності

2.3.1. Основні етапи реалізації фарбування вокселів

1. Калібрування камери для визначення внутрішніх параметрів.
2. Збір зображень об'єкта.
3. Калібрування зовнішніх параметрів для різних зображень.
4. Видалення фону із зображень.
5. Ініціалізація початкового об'єму для висікання.
6. Розділення об'єму на вокселі відповідно до заданої роздільної здатності реконструкції.
7. Розділення масиву вокселів на шари L .
8. Для кожного вокселя в шарі L повторити наступні кроки.
9. Зворотне проектування вокселів на всі зображення. Якщо воксель видимий, обчислити його консистентність за допомогою різних перевірок:
 - а) Оригінальна перевірка консистентності.
 - б) Адаптивна перевірка консистентності.
 - в) Перевірка на основі відстані Мінковського.

г) Перевірка консистентності на основі гістограми.

10. Якщо воксель консистентний, включити його в кінцевий об'єм, інакше висікти цей воксель з об'єму.

2.3.2. Метрики консистентності

Часто перевіряють консистентність вокселя, використовуючи стандартне відхилення (σ) кольору в області зображення, що відповідає вокселю.

$$\text{consistent}(v) = \begin{cases} \text{true} & \text{якщо } \sigma(S_v) < t \\ \text{false} & \text{інакше} \end{cases}$$

Тут S_v — це набір пікселів на зображенні, який відповідає вокселю, що розглядається для перевірки консистентності.

Оскільки текстуровані області мають природно вищу дисперсію кольору, використання єдиного порогового значення t є неоптимальним для всієї сцени. Воксель, видимий на зображенні, зазвичай проектується на декілька пікселів, що охоплюють області з різною варіацією кольору [22]. Існує потреба в адаптивному порозі, який змінюється відповідно до варіації кольору, видимого з одного зображення.

Якщо S_i — це набір пікселів на зображенні i , з яких воксель V видимий, а σ_i — стандартне відхилення значень пікселів. Якщо $\bar{\sigma}$ — це середнє значення для зображень i , з яких V видимий, то адаптивна консистентність (ASDT) визначається наступним чином:

$$\text{consistent}(v) = \text{true} \quad \text{для всіх } i \in P_v, \text{ якщо } \sigma_i \leq t_1 \cdot \bar{\sigma} + t_2$$

$$\text{consistent}(v) = \text{false} \quad \text{інакше}$$

t_1 та t_2 — це пороги, встановлені вручну. Модель ASDT демонструє меншу кількість артефактів (отворів) порівняно з іншими моделями.

Фотоконсистентність може бути визначена за допомогою відстаней Мінковського L_p [15]. Якщо воксель v видимий з p ракурсів то відстань між кольоровими сутностями будь-яких двох наборів кольорів має бути обмеженою. Фотоконсистентність визначається співвідношенням:

$$\text{consistent}(v) = \begin{cases} \text{true} & \text{якщо } L_p(\pi_i, \pi_j) < t \\ \text{false} & \text{інакше} \end{cases} \quad \text{для всіх } i, j \in \{0, \dots, p-1\}$$

Перевага цього підходу полягає в тому, що якщо під час перевірки будь-який воксель виявиться неконсистентним (тобто, відстань між парою пікселів $> t$), подальше обчислення може бути зупинено.

Стандартне відхилення та L_1 відстань не завжди є надійними мірами, оскільки вони не мні на білому фоніють диференціювати між кластеризованими розподілами кольорів і суто випадковими розподілами. Перевірка консистентності гістограми здатна подолати це обмеження.

Отже, в розділі детально обговорено алгоритм фарбування вокселів та його поведінку залежно від варіацій роздільної здатності та кількості вхідних зображень. Підтверджено, що:

1. Збільшення роздільної здатності вокселів призводить до підвищення деталізації реконструйованої моделі, але супроводжується значним (нелінійним) зростанням часу обчислення.

2. Збільшення кількості вхідних зображень лінійно підвищує час обчислення, водночас накладаючи жорсткіші обмеження на фотоконсистентність, що, як правило, призводить до точнішої апроксимації об'єкта та, відповідно, до збільшення кількості фінальних вокселів.

У реконструкції з використанням фотоконсистентності, міра консистентності відіграє вирішальну роль. При використанні порогових методів (оригінальна та адаптивна перевірка консистентності), значення порогу критично впливає на кінцеву реконструкцію, що вимагає ретельної

перевірки різних значень. На противагу цьому, перевірка консистентності гістограми є методом, менш залежним від порогового значення.

2.4. Представлення та опис методології фарбування вокселів

На рисунку 2.4 візуалізовано повний процес методології фарбування вокселів (Voxel Coloring) — одного з провідних об'ємних підходів у тривимірній реконструкції на основі фотограмметричної консистентності. Цей процес дозволяє відновити просторову форму об'єкта (у даному випадку, вази) виключно на основі його двовимірних проєкцій.

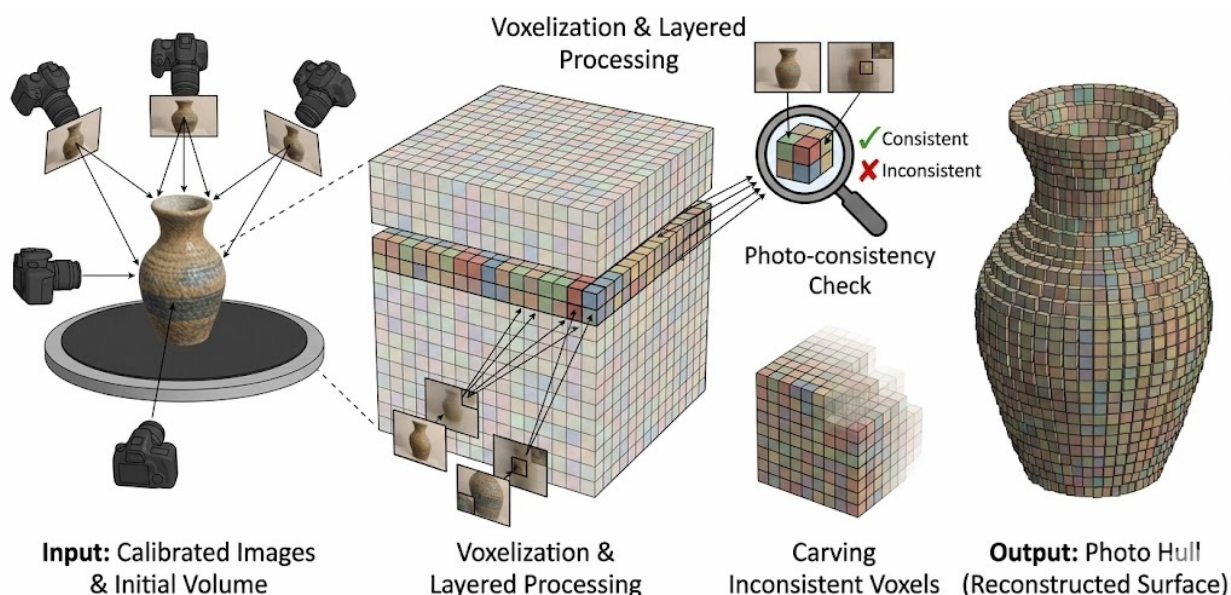


Рис. 2.4. Процеси методології фарбування вокселів

1. Етап вхідних даних (input) та ініціалізації

Процес починається із двох основних компонентів:

- калібровані зображення (calibrated images): об'єкт розташовано на поворотному столі (або його оточує масив камер), що дозволяє зібрати серію знімків з різних, точно визначених ракурсів. Для кожного зображення

необхідні точні параметри калібрування камери, що визначають її внутрішні та зовнішні характеристики.

- початковий об'єм (initial volume): навколо об'єкта визначається обмежувальний об'єм (зазвичай, куб), який гарантовано його охоплює. Цей об'єм є простором пошуку для реконструкції.

2. Вокселізація та пошарова обробка (voxelization & layered processing)

Початковий об'єм дискретизується на регулярну тривимірну сітку — вокселі (Voxels). Усі вокселі спочатку вважаються потенційно непрозорими (належать об'єкту).

Критичним для ефективності алгоритму Voxel Coloring є пошарова обробка (Layered Processing). Вокселі логічно розділяються на шари, що відповідають зростаючій глибині відносно певної камери (або групи камер). Це забезпечує топологічний порядок, при якому процес завжди перевіряє ближчі до камери вокселі першими. Таким чином, коли обробляється певний воксель, відомо, які саме сусідні чи попередні вокселі його закривають, що є необхідним для точного обчислення видимості.

3. Перевірка фотоконсистентності (photo-consistency check)

Цей етап є ядром методології. Кожен воксель на поточному шарі перевіряється з метою визначення, чи може він належати реальній поверхні об'єкта, припускаючи, що поверхня є ламбертівською (тобто, відбиває світло однаково у всіх напрямках).

1) проєкція та видимість: воксель проєктується на площини всіх каліброваних зображень, з яких він є видимим.

2) порівняння кольорів: зображення ілюструє збільшувальну лупу, де відбувається порівняння кольорів (текстури) вокселя на пікселях, на які він проєктується.

3) критерій консистентності:

- consistent (консистентний): якщо колір вокселя на всіх видимих зображеннях є ідентичним або знаходиться в межах мінімально допустимого

порогу (низьке стандартне відхилення), воксель маркується як консистентний і зберігається.

- inconsistent (неконсистентний): якщо кольори суттєво відрізняються, що вказує на те, що воксель не лежить на фактичній поверхні (наприклад, він знаходиться в порожньому просторі або є частиною тіні/відблиску, що порушує припущення про Ламбертову поверхню), він позначається як неконсистентний.

4. Висікання та фінальний результат (carving and output)

- висікання неконсистентних вокселів (carving inconsistent voxels): усі вокселі, визнані неконсистентними, негайно видаляються (висікаються) з обмежувального об'єму, залишаючи прозорий простір. зображення показує цей процес, де частина вокселів, які були у початковому кубі, зникає.

- вихід: фотовуаль (photo hull): результатом ітеративного висікання є фотовуаль (photo hull). це — об'єднання всіх фотоконсистентних реконструкцій сцени. фотовуаль складається лише з тих вокселів, які успішно пройшли перевірку консистентності, і являє собою реконструйовану поверхню (reconstructed surface) об'єкта з високим ступенем геометричної та фотореалістичної точності.

Таким чином, методологія Voxel Coloring ефективно відновлює 3D-форму, використовуючи колірну інформацію як жорстке геометричне обмеження.

Висновки до розділу

У другому розділі досліджено методологію об'ємної 3D-реконструкції з використанням фотоконсистентних критеріїв. Проаналізовано різні конфігурації збору вхідних даних та їх вплив на якість відновлення просторової структури об'єктів. Формалізовано постановку задачі реконструкції в об'ємному представленні, що дозволяє перейти від точкових моделей до щільних 3D-описів. Розглянуто основні об'ємні підходи, зокрема

реконструкцію візуальної оболонки та методи заповнення вокселів. Показано, що метод фарбування вокселів є ефективним інструментом для відновлення геометрії на основі узгодженості проєкцій у множині зображень. Детально описано етапи реалізації цього методу та умови його коректного застосування. Проаналізовано метрики фотоконсистентності, які відіграють ключову роль у визначенні достовірності реконструйованих об'ємів. Встановлено, що вибір метрик безпосередньо впливає на точність та стійкість результатів. Наголошено на обчислювальній складності об'ємних методів і їх чутливості до якості вхідних даних. Отримані висновки слугують підґрунтям для переходу до сучасних методів генерації 3D-моделей.

РОЗДІЛ 3. ІМПЛЕМЕНТАЦІЯ МОДЕЛЕЙ ТА МЕТОДІВ ГЕНЕРАЦІЇ ТРИВИМІРНИХ МОДЕЛЕЙ

3.1. Систематичний огляд генерації 3D-контенту

Автоматична генерація тривимірних (3D) моделей традиційно є однією з ключових, фундаментальних проблем у галузях комп'ютерного зору та комп'ютерної графіки. Це завдання викликає значний науковий інтерес, що зумовлено його широкими та критично важливими застосуваннями, зокрема у розробці відеоігор, кіноіндустрії, створенні віртуальних персонажів та імерсивних досвідів. Усі ці сфери, як правило, вимагають створення великої кількості деталізованих 3D-активів. Нещодавні успіхи в розробці нейронних представлень, зокрема нейронних полів випромінювання (Neural Radiance Fields, NeRFs) [12, 16], а також генеративних моделей, таких як дифузійні моделі [18], значно сприяли винятковим проривам у сфері генерації 3D-контенту.

3.1.1. Еволюція генеративних моделей у 2D-просторі

У площині генерації двовимірного (2D) вмісту, останні досягнення в генеративних моделях послідовно підвищували спроможність систем до генерації та редагування зображень, забезпечуючи дедалі різноманітніші та виключно високоякісні результати. Новаторські дослідження в галузі генеративно-змагальних мереж (GANs) [19, 20], варіаційних автокодувальників (VAEs) [21] та авторегресивних моделей [22] продемонстрували вражаючі наслідки для фотореалістичного синтезу.

Більше того, поява генеративного штучного інтелекту (ШІ) та дифузійних моделей знаменує собою парадигмальний зсув у методах маніпуляції зображеннями, зокрема в архітектурах Stable Diffusion [23], Imagen [24], Midjourney [25] та DALL-E 3 [26]. Ці генеративні моделі ШІ уможливають створення та редагування фотореалістичних або

стилізованих зображень, а також відео, вимагаючи при цьому лише мінімального вхідного сигналу, часто у формі текстових підказок. Як наслідок, вони здатні генерувати креативний вміст, що виходить за межі реального світу, розширюючи ліміти творчості та художнього вираження. Завдяки їхнім "емерджентним" можливостям, ці моделі переосмислили досяжні межі у сфері генерації вмісту, значно розширюючи горизонти креативності та художнього вираження.

3.1.2. Перехід до 3D-генерації

Потреба в розширенні парадигми генерації 2D-вмісту в 3D-простір стає дедалі більш критичною для сучасних застосувань, пов'язаних зі створенням 3D-активів або імерсивних досвідів, особливо з огляду на стрімкий розвиток метавсесвіту. Однак, перехід від 2D до 3D-генерації вмісту не може бути інтерпретований лише як технологічна еволюція. Насамперед, це є відповіддю на вимоги сучасних застосувань, які вимагають більш детальної та точної реплікації фізичного світу, що часто не може бути забезпечено 2D-представленнями. Цей зсув підкреслює фундаментальні обмеження 2D-контенту в додатках, які вимагають комплексного розуміння просторових відношень та адекватного сприйняття глибини.

У міру того, як значущість 3D-контенту стає все більш очевидною, спостерігається значний сплеск дослідницьких зусиль, присвячених цій галузі. Проте, перехід від 2D до 3D-генерації вмісту не є простим розширенням чи масштабуванням існуючих 2D-методологій. Навпаки, він передбачає необхідність подолання унікальних викликів та переоцінку представлення даних, формулювання завдання (задачі) та базових генеративних моделей для ефективного вирішення складнощів 3D-простору. Наприклад, залишається не очевидним, яким чином слід інтегрувати 3D-представлення сцени у 2D-генеративні моделі для обробки вищих вимірів, що є необхідним для 3D-генерації. На відміну від зображень чи відео, які легко збирати з мережі, 3D-активи є відносно дефіцитними та складними для

отримання. Крім того, оцінювання якості згенерованих 3D-моделей створює додаткові методологічні виклики, оскільки існує потреба у розробці кращих формулювань для функцій об'єктивності, зокрема при врахуванні багатовидової узгодженості у 3D-просторі. Ці складнощі вимагають інноваційних підходів та нових рішень для подолання концептуального та технологічного розриву між 2D- та 3D-генерацією вмісту.

Хоча 3D-генерація контенту не була такою помітною, як її 2D-аналог, вона неухильно прогресувала, досягаючи низки значних наукових результатів.

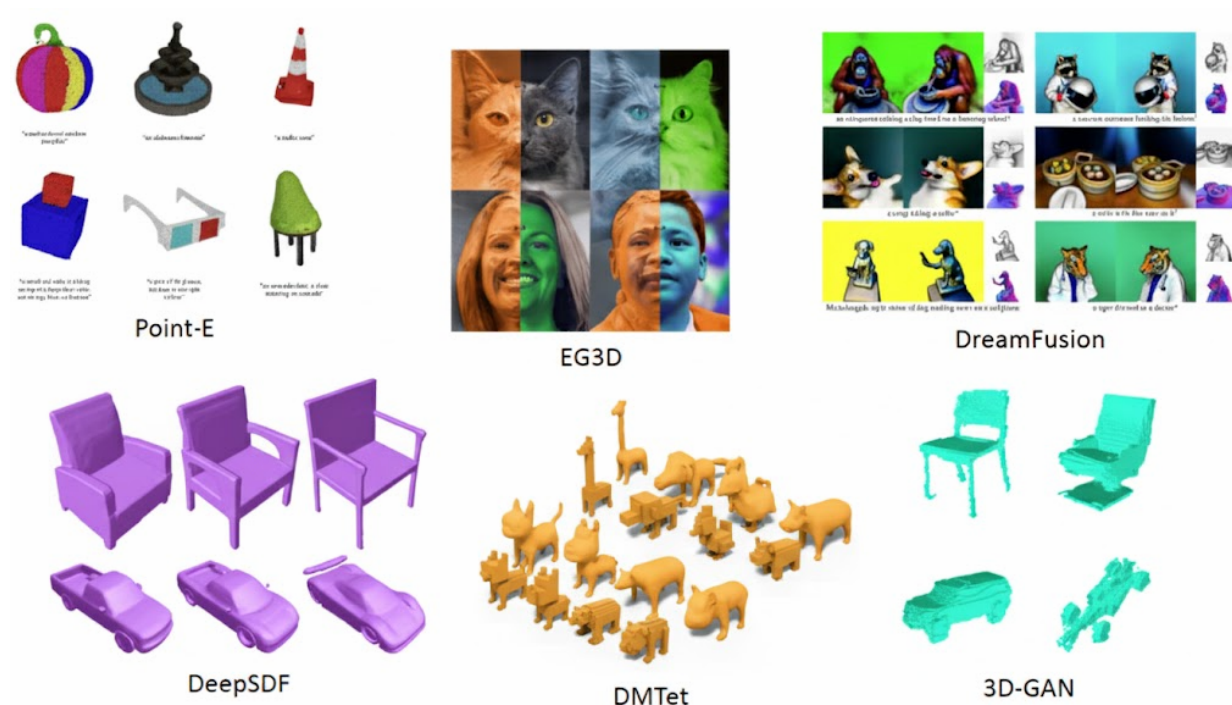


Рис. 3.1. Результати 3D-генерації, отримані за допомогою різних платформ

Репрезентативні приклади, які можуть бути продемонстровані на рис. 3.1, демонструють суттєві покращення як у якості синтезу, так і в різноманітності згенерованих об'єктів, переходячи від ранніх методів, як-от 3D-GAN, до новітніх підходів. Таким чином, це дослідження має на меті систематично вивчити стрімкі досягнення та багатогранні розробки, що

відбуваються у сфері генерації 3D-контенту. Ми представляємо структурований огляд та комплексну дорожню карту багатьох останніх робіт, зосереджуючись на ключових аспектах: 3D-представленнях, методах 3D-генерації, наборах даних та застосуваннях 3D-контенту, а також окреслюємо відкриті проблеми.

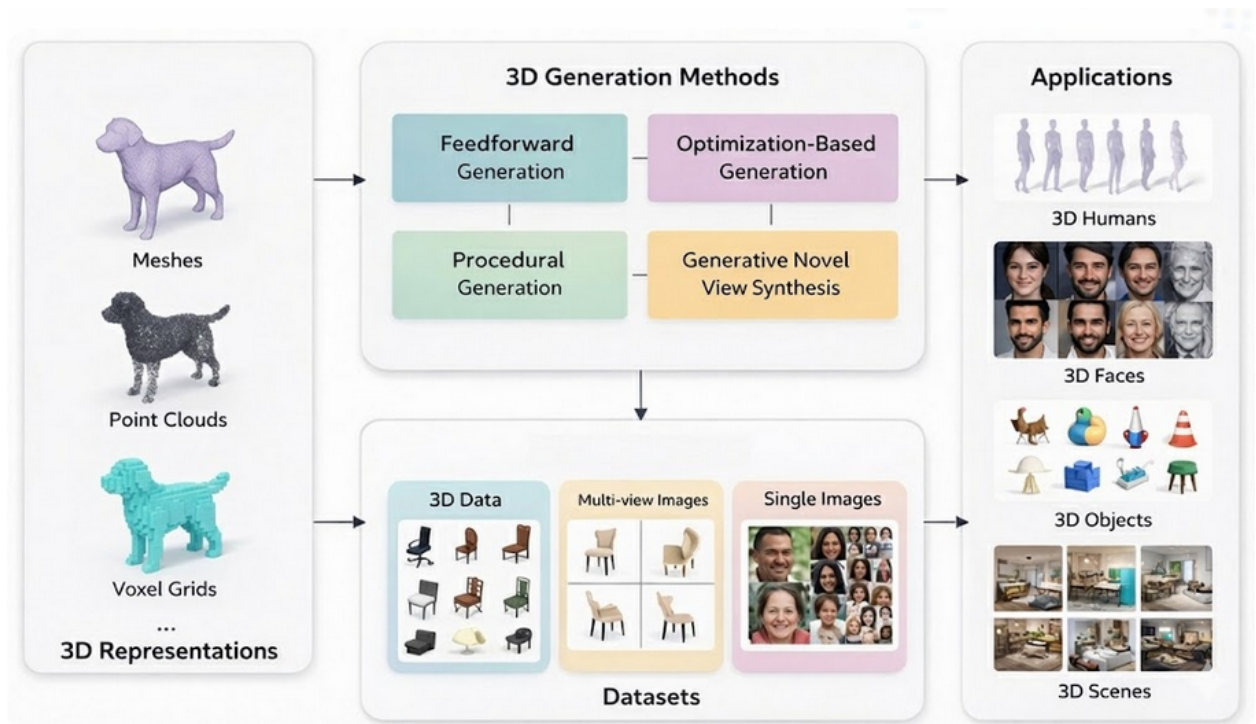


Рис. 3.2. Представлення методів тривимірної генерації моделей

На рисунку 3.2 подано 3D-представлення, методи 3D-генерації, набори даних та застосування. Зокрема, подаються 3D-представлення, які слугують основою для 3D-генерації. Крім того, ми проведемо комплексний огляд методів генерації, що стрімко зростає, категоризованих за типом алгоритмічних парадигм, включно з прямою генерацією (feedforward generation), генерацією на основі оптимізації, процедурною генерацією та генеративним синтезом нових видів (generative novel view synthesis).

В цьому розділі буде представлено основні парадигми представлення сцен та їхні відповідні функції рендерингу, що використовуються у 3D-генерації. Буде досліджено широкий спектр методів 3D-генерації, які можуть

бути класифіковані на чотири основні категорії за їхніми алгоритмічними методологіями. Також буде зображено еволюційне дерево цих методологій для ілюстрації їхніх основних відгалужень.

Оскільки накопичення даних відіграє життєво важливу роль у забезпеченні успіху моделей глибокого навчання, у наступному розділі ми представляємо відповідні набори даних, які використовуються для навчання методів 3D-генерації.

Насамкінець, ми включаємо коротке обговорення пов'язаних застосувань, таких як генерація 3D-людей та облич, окреслюємо відкриті проблеми та підсумовуємо це дослідження. В роботі ми надаємо вичерпний огляд з генерації 3D-контенту, маючи на меті запропонувати швидке розуміння архітектурного фреймворку 3D-генерації та його основних принципів. Ми пропонуємо багатоперспективну категоризацію методів 3D-генерації, швидко ідентифікувати релевантні роботи та сприяти кращому розумінню пов'язаних технік.

3.2. Деталізація методологій 3D-генерації

Згідно з представленою категоризацією, методи 3D-генерації можна розділити на чотири основні групи за їхніми фундаментальними алгоритмічними підходами. Ось детальний огляд цих категорій.

3.2.1. Метод прямої 3D генерації

Цей клас методів є найбільш прямим розширенням традиційних генеративних моделей (таких як GANs та VAEs) у 3D-простір.

Основний принцип полягає в наступному. Модель навчається відображати латентний вектор (latent vector) безпосередньо у цільове 3D-представлення (наприклад, вокселі, хмари точок, полігональні сітки або неявні функції). Процес генерації відбувається за один прохід (feedforward pass) через мережу.

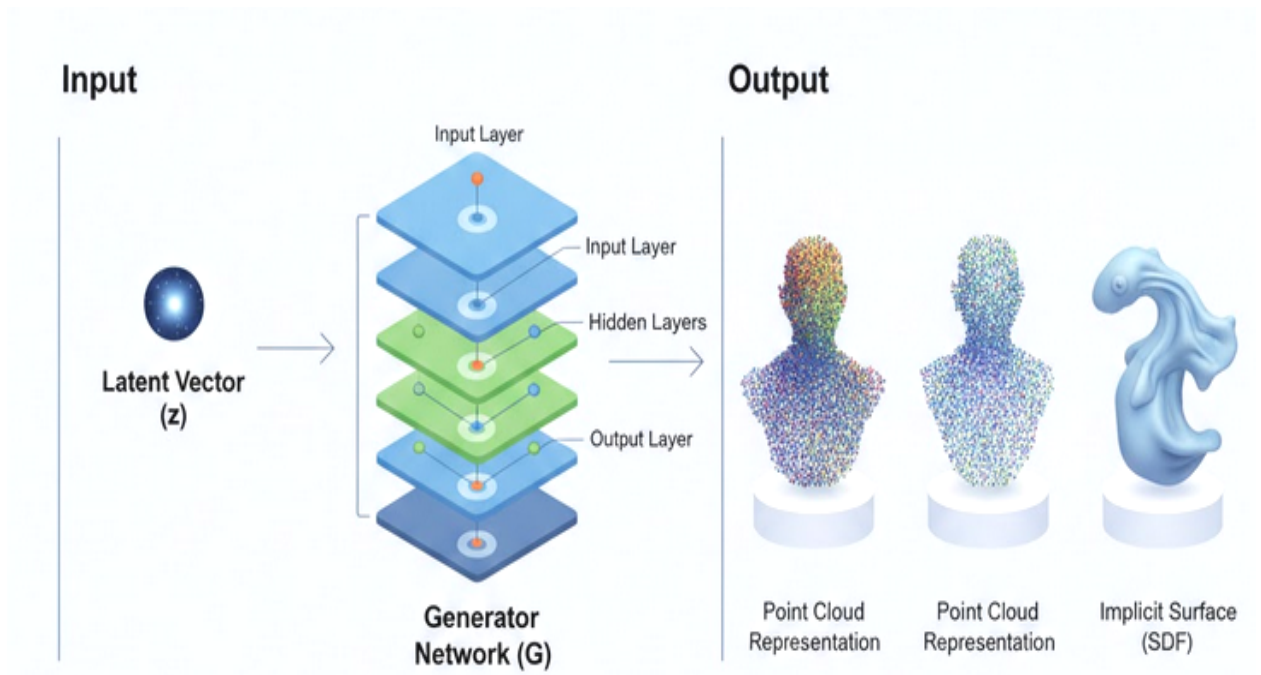


Рис. 3.3. Метод прямої 3D генерації

Методологія полягає в наступному:

- 3D-GANs / 3D-VAEs - навчаються генерувати 3D-дані в дискретизованому форматі (наприклад, воксельні сітки $V \in \{0,1\}N \times N \times N$). Через високу обчислювальну вартість високої роздільної здатності вокселів, ці підходи часто обмежені генерацією об'єктів з низькою деталізацією.

- Генерація на основі сіток (Mesh-based) - використовують архітектури, які генерують топологію сітки та її атрибути (вершини, грані) безпосередньо.

- Неявні представлення (Implicit Representations). Сучасніші методи використовують нейронні мережі для безпосередньої генерації параметрів неявних функцій, які описують 3D-форму (наприклад, поля знакового відстані - Signed Distance Fields, SDFs), що дозволяє генерувати об'єкти з гладкішими поверхнями.

Перевагою методу є висока швидкість генерації після навчання, оскільки не вимагає ітераційної оптимізації. До недоліку можна віднести уутливість до розподілу даних, складність масштабування до високої роздільної здатності для складних сцен.

3.2.2. 3D генерація на основі оптимізації

Цей підхід використовує попередньо навчену 2D-генеративну модель (часто дифузійну модель) як потужний "пріор" (prior) для синтезу 3D-моделі.

Основний принцип полягає в тому, що замість прямої генерації, 3D-представлення сцени (наприклад, NeRF) оптимізується ітераційно, щоб його 2D-проекції (рендеринги з різних ракурсів) були узгоджені з виходом, згенерованим потужною 2D-моделлю (зазвичай, за допомогою текстової підказки).

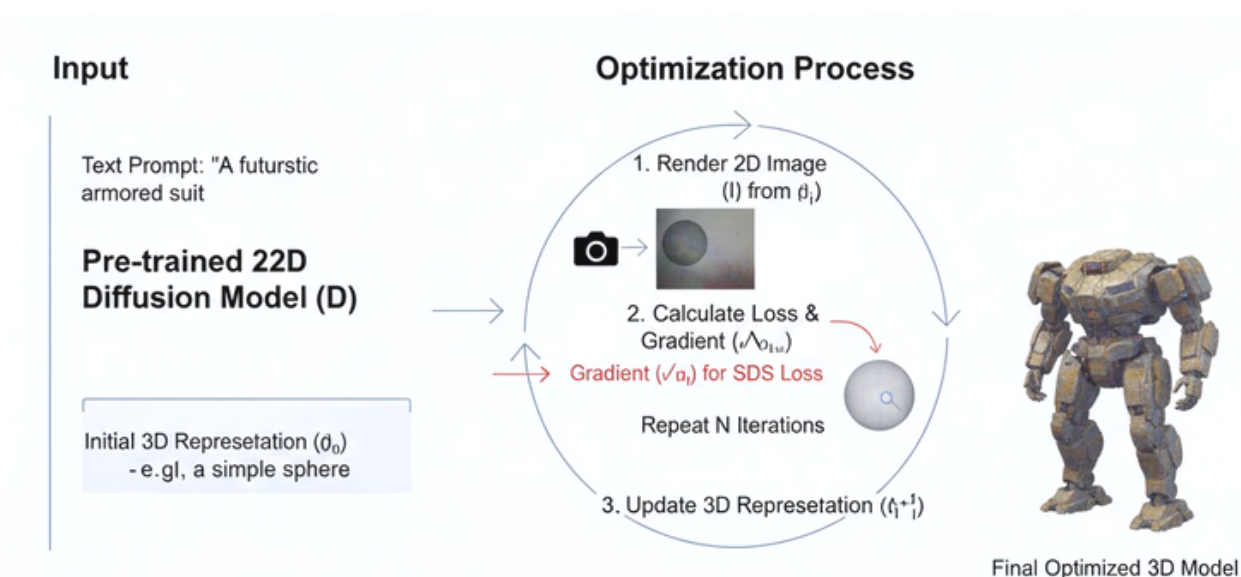


Рис. 3.4. Представлення методу 3D генерації на основі оптимізації

Методологія наступна. Score Distillation Sampling (SDS) - ключова техніка, яка використовує попередньо навчену дифузійну модель для створення функції втрат (loss function).

На кожній ітерації:

1. Рендериться 2D-зображення I з поточної 3D-моделі θ (наприклад, NeRF).

2. Розраховується градієнт, який вказує, як змінити 3D-модель θ , щоб I більше відповідало умові (наприклад, текстовій підказці) згідно з 2D-дифузійною моделлю.

3. 3D-модель θ оновлюється за допомогою градієнтного спуску.

Перевагою є здатність використовувати надзвичайну креативність і якість, властиві великим попередньо навченим 2D-дифузійним моделям, для генерації високоякісних 3D-об'єктів за текстовою підказкою. Недоліком є значна обчислювальна вартість через тривалий процес ітераційної оптимізації (потрібно багато кроків для збіжності); виклики з багатовидовою узгодженістю.

3.2.3. Процедурна 3D генерація

Ці методи спираються на набір формальних правил або алгоритмів для створення 3D-геометрії, часто з високим ступенем деталізації та структури. Генерація керується не великим набором даних, а визначеними правилами, які можуть бути ієрархічними або стохастичними.

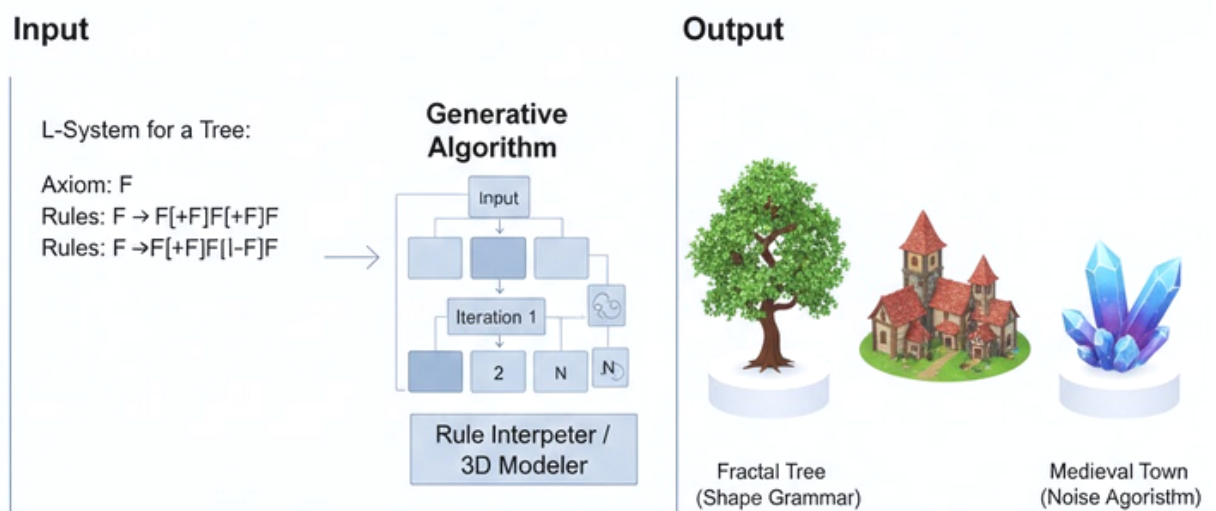


Рис. 3.5. Метод процедурної 3D генерації

Методологія наступна:

- L-системи (Lindenmayer Systems) - використовуються для генерації складних рослинних структур (дерева, куці) за допомогою рекурсивних правил заміщення.

- Генерація будівель - використання граматики форм (Shape Grammars) для автоматичного створення архітектурних елементів на основі стилю.

- Комбінування ШІ та правил. Сучасні підходи можуть використовувати ШІ для визначення високорівневих параметрів (стилю, розташування), які потім передаються процедурним генераторам для створення детальної геометрії.

Перевагою є висока структурованість, легкість контролю над результатами, створення дуже складних і реалістичних структур (особливо природних). Недолік - менша гнучкість порівняно з методами, керованими даними, оскільки результат обмежений заданими правилами.

3.2.4. Метод генеративного синтезу нових видів

Цей підхід сфокусований на генерації узгоджених нових ракурсів сцени на основі обмеженої кількості вхідних зображень або латентного простору. Хоча він тісно пов'язаний з NeRFs, його генеративна природа відрізняється від реконструкції.

Основний принцип полягає в навчання генеративної моделі для синтезу вигляду сцени з довільної, раніше небаченої точки огляду. Це часто передбачає навчання моделі, яка кодує всю 3D-сцену в компактному латентному просторі.



Рис. 3.6. Метод генеративного синтезу нових видів

Методологія:

- Generative NeRFs (G-NeRFs) - навчання VAE або GANs, які працюють безпосередньо в латентному просторі NeRF, дозволяючи генерувати нові, різноманітні сцени, які можна рендерити з будь-якого ракурсу.

- Implicit Generative Models - моделі, які використовують 2D-дані (зображення) як вхідні дані для інференції (виведення) 3D-представлення, яке потім може бути використане для рендерингу нових видів.

- Методи регуляризації включають механізми для забезпечення багатовидової узгодженості та фотореалістичної якості у згенерованих нових ракурсах.

Перевагами є висока якість рендерингу, природна підтримка 3D-узгодженості, здатність до генерації не тільки об'єктів, але й цілих сцен. Щодо недоліків, то обчислювально інтенсивний рендеринг (якщо використовуються класичні NeRF), високі вимоги до обсягу та якості навчальних даних.

3.3. Представлення сцени та рендеринг у генерації 3d-моделей на основі штучного інтелекту

У сфері тривимірного вмісту, генерованого штучним інтелектом (ШІ), вибір адекватного представлення 3D-моделей відіграє критично важливу роль. Процес генерації типово охоплює два ключові компоненти: ефективне представлення сцени та диференційований алгоритм рендерингу для синтезу 3D-моделей і подальшої візуалізації 2D-зображень. З іншого боку, згенеровані 3D-моделі або їхні 2D-проекції можуть контролюватися (оптимізуватися/навчатися) або у домені реконструкції (3D), або у домені зображення (2D), як ілюструє рисунок 3.7. Деякі методології безпосередньо контролюють параметри 3D-моделі (представлення сцени), тоді як інші візуалізують представлення сцени у простір зображення та контролюють отримані візуалізації.

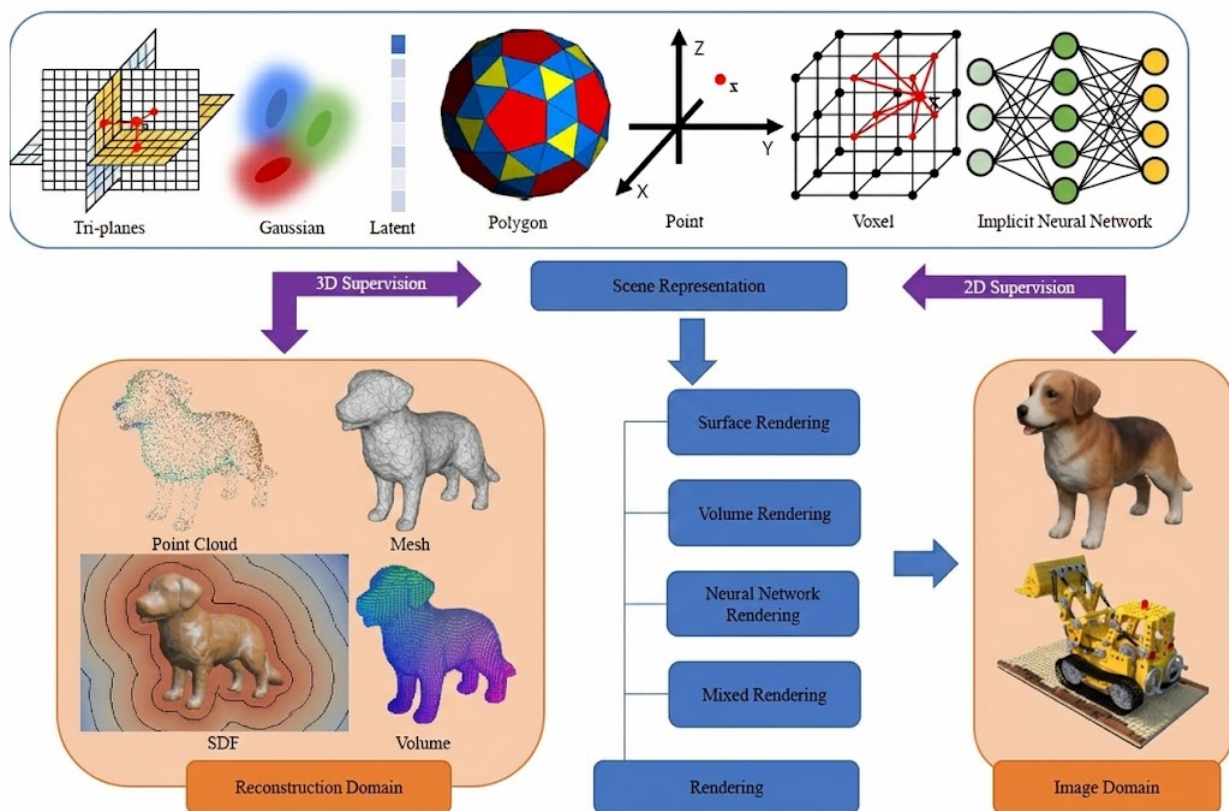


Рис. 3.7. Нейронні представлення сцени, які використовуються для 3D-генерації

3D-генерація передбачає використання представлень сцени та диференційованого алгоритму рендерингу для створення 3D-моделей або візуалізації 2D-зображень. З іншого боку, ці 3D-моделі або 2D-зображення можуть функціонувати як домен реконструкції або домен зображення, контролюючи 3D-генерацію представлень сцени.

Ми загалом класифікуємо представлення сцени на три основні групи:

- Експліцитні представлення сцени.
- Імпліцитні представлення.
- Гібридні представлення.

Важливо зазначити, що для забезпечення можливості оптимізації представлень сцени на основі градієнта з різних вхідних даних, застосовувані методи рендерингу (наприклад, трасування променів, об'ємний рендеринг, растеризація тощо) повинні бути диференційованими.

3.3.1. Експліцитні представлення

Експліцитні представлення сцени слугують фундаментальним модулем у комп'ютерній графіці та зорі, оскільки вони пропонують всебічний і прямий засіб опису 3D-сцен. Зображуючи сцени як сукупність базових примітивів (включно з точкоподібними примітивами, трикутниковими сітками та вдосконаленими параметричними поверхнями), ці представлення можуть створювати детальні та точні візуалізації різноманітних середовищ та об'єктів.

Хмара точок являє собою сукупність елементів у евклідовому просторі R^3 , що представляє дискретні точки P_i з додатковими атрибутами (наприклад, кольорами та нормальми) у тривимірному просторі. На додаток до простих точок, які можна вважати нескінченно малими фрагментами поверхні, можуть також використовуватися орієнтовані хмари точок із радіусом (сурфели).

Сурфели застосовуються в комп'ютерній графіці для рендерингу хмар точок (що називається розщепленням), які є диференційованими [17]. Це дозволяє дослідникам визначати диференційовані конвеєри рендерингу для коригування позицій і ознак хмари точок, таких як радіус або колір.

Такі методи, як нейронний рендеринг на основі точок (Neural Point-based Rendering) [18], SynSin [19], Pulsar та ADOP [20], використовують ознаки, що навчаються, для зберігання інформації про зовнішній вигляд та форму поверхні, що забезпечує більш точні та детальні результати рендерингу.

Кілька інших методів, наприклад, FVS, SVS [21] та FWD-Transformer, також застосовують ознаки, що навчаються, для покращення якості рендерингу. Ці методи зазвичай вбудовують ознаки в хмари точок і деформують їх до цільових видів для декодування значень кольору, що дозволяє більш точну та детальну реконструкцію сцени.

Шляхом інтеграції диференційованих рендерерів на основі хмар точок у процес 3D-генерації, дослідники можуть використовувати переваги хмар

точок, зберігаючи сумісність з методами оптимізації на основі градієнта. Цей процес загалом можна розділити на два різні способи:

- Розщеплення точок (point splitting) - змішування дискретних зразків з деякими локальними детермінованими ядрами розмиття [25].

- Звичайний точковий рендерер (conventional point renderer) [26].

Ці методи полегшують генерацію та маніпулювання 3D-моделями хмар точок, зберігаючи при цьому диференційовність, необхідну для навчання та оптимізації нейронних мереж у завданнях 3D-генерації.

3.3.2. Застосування сіток

Шляхом з'єднання багатьох вершин ребрами можуть бути сформовані більш складні геометричні структури (наприклад, каркаси та сітки) [28]. Ці структури можуть бути додатково вдосконалені за допомогою полігонів, зазвичай трикутників або чотирикутників, для створення реалістичних представлень об'єктів.

Сітки забезпечують універсальний та ефективний засіб представлення складних форм і структур, оскільки їх можна легко маніпулювати та візуалізувати комп'ютерними алгоритмами.

Більшість інструментів графічного редагування використовують трикутні сітки. Цей тип представлення є незамінним для будь-якого конвеєра створення цифрового контенту (DCC) з огляду на його широке визнання та сумісність.

Для безперешкодного узгодження з цими конвеєрами нейронні мережі можуть бути стратегічно навчені прогнозувати дискретні розташування вершин. Ця здатність дозволяє безпосередньо імпортувати ці розташування в будь-який конвеєр DCC, сприяючи плавному та ефективному робочому процесу.

На противагу прогнозуванню дискретних текстур, пропонуються методи неперервних текстур, оптимізовані через нейронні мережі, такі як поля текстур (texture fields) [26] та NeRF-Tex [27]. Це може забезпечити

більш витончену та детальну текстуру, підвищуючи загальну якість та реалізм згенерованих 3D-моделей.

Інтеграція сіткового представлення у 3D-генерацію вимагає використання диференційованих методів рендерингу на основі сіток, які дозволяють растеризувати сітки у спосіб, сумісний з оптимізацією на основі градієнта. Було запропоновано кілька таких методів, зокрема OpenDR [28], neural mesh renderer, Paparazzi та Soft Rasterizer [29]. Крім того, фізично коректні рендерери загального призначення, як-от Mitsuba 2 та Taichi [30], підтримують диференційований рендеринг на основі сіток через автоматичне диференціювання.

3.3.3. Багатошарові представлення

Використання багатьох напівпрозорих кольорових шарів для представлення сцен є популярною та успішною схемою у синтезі нових видів у режимі реального часу.

Представлення LDI (Layered Depth Image) [31] є помітним прикладом, розширюючи традиційні карти глибини шляхом включення багатьох шарів карт глибини $L_k(u,v)$ (де k — індекс шару), кожен із пов'язаними значеннями кольору. Кілька методів [32] черпали натхнення з представлення LDI та використовували досягнення глибокого навчання для створення мереж, здатних прогнозувати LDI.

В [33] первинно представлено представлення MPI, яке описує сцени за допомогою багатьох фронтально-паралельних напівпрозорих шарів, включаючи кольори та непрозорість, у фіксованих діапазонах глибини через об'єми площинного розгортання (plane sweep volumes). За допомогою об'ємного рендерингу та гомографічної проєкції може бути синтезовано новий вид у режимі реального часу.

Ґрунтуючись на цьому, різні методи [34] прийняли представлення MPI для покращення якості рендерингу. Оскільки дослідження в цій галузі продовжують розвиватися, можна очікувати подальших удосконалень у цих

методах, що призведе до більш ефективних та дієвих методів 3D-генерації для рендерингу в режимі реального часу.

3.4. Методологія та еволюція генеративних моделей у задачах синтезу 3D-моделей та сцен

Протягом останнього часового проміжку інтенсивний розвиток генеративних моделей у сегменті синтезу двовимірних зображень — зокрема генеративно-змагальних мереж (GANs), варіаційних автокодувальників (VAEs), авторегресійних та дифузійних моделей — зумовив їхню експансію у сферу тривимірного моделювання. Це призвело до формування складних гібридних архітектур, де вказані моделі інтегруються з сучасними представленнями сцен для вирішення завдань 3D-генерації.

Таблиця 3.1.

Порівняльна характеристика методів генерації 3D-контенту

| Метод | Генеративна модель | Простір генерації | Простір реконструкції | Рендеринг | Нагляд | Умова |
|-----------------------|---------------------|-------------------|-----------------------|--------------------|--------|--------------------------|
| Потоки / VAE | | | | | | |
| PointFlow | Нормалізуючий потік | Латентний код | Хмара точок | - | 3D | Без умов |
| 3dAAE | VAE | Латентний код | Хмара точок | - | 3D | Без умов |
| SDM-NET | VAE | Латентний код | Сітка (Mesh) | - | 3D | Без умов |
| Авторегресійні | | | | | | |
| AutoSDF | Авторегресійна | Воксель | SDF | - | 3D | Без умов |
| PolyGen | Авторегресійна | Полігон | Сітка (Mesh) | - | 3D | Без умов / Мітка / Зобр. |
| PointGrow | Авторегресійна | Точка | Хмара точок | - | 3D | Без умов / Мітка / Зобр. |
| GAN | | | | | | |
| EG3D | GAN | Латентний код | Три-площина | Змішаний рендеринг | 2D | Без умов |
| GIRAFFE | GAN | Латентний код | NeRF | Змішаний рендеринг | 2D | Без умов |

| Метод | Генеративна модель | Простір генерації | Простір реконструкції | Рендеринг | Нагляд | Умова |
|------------------|--------------------|-------------------|-----------------------|-----------------------|---------|--------------------------|
| BlockGAN | GAN | Латентний код | Воксельна сітка | Мережевий рендеринг | 2D | Без умов |
| gDNA | GAN | Латентний код | Поле зайнятості | Поверхневий рендеринг | 2D & 3D | Без умов |
| SurfGen | GAN | Латентний код | SDF | - | 3D | Без умов |
| tree-GAN | GAN | Латентний код | Хмара точок | - | 3D | Без умов |
| Дифузійні | | | | | | |
| HoloDiffusion | Дифузійна | Воксель | NeRF | Об'ємний рендеринг | 2D | Зображення |
| SSDNeRF | Дифузійна | Три-площина | NeRF | Об'ємний рендеринг | 2D | Без умов / Зображення |
| 3DShape2Vec Set | Дифузійна | Латентна множина | SDF | - | 3D | Без умов / Текст / Зобр. |
| Point-E | Дифузійна | Точка | Хмара точок | - | 3D | Текст |
| 3DGen | Дифузійна | Три-площина | Сітка (Mesh) | - | 3D | Текст / Зображення |
| DreamFusion | Дифузійна | - | NeRF | Об'ємний рендеринг | SDS | Текст |
| Make-It-3D | Дифузійна | - | Хмара точок | Мережевий рендеринг | SDS | Зображення |
| Zero-1-to-3 | Дифузійна | Піксель | - | - | 2D | Зображення |
| MVDream | Дифузійна | Піксель | - | - | 2D | Зображення |
| DMV3D | Дифузійна | Піксель | Три-площина | Об'ємний рендеринг | 2D | Текст / Зображення |

У таблиці 3.1 наведено репрезентативні приклади систем 3D-генерації, що базуються на синергії генеративних парадигм та специфічних представлень сцен. Дані методи оперують двома ключовими доменами: простором генерації, де безпосередньо формується представлення за допомогою генеративних мереж, та простором реконструкції, де отримується фінальний результат. Для ілюстрації: архітектура AutoSDF застосовує авторегресійну модель на базі трансформера для апроксимації функціональної воксельної сітки з подальшим декодуванням у функцію

знакового відстані (SDF). Система EG3D використовує GAN для генерації об'єктів у латентному просторі, впроваджуючи концепцію «три-площини» (tri-plane) для ефективного рендерингу. Аналогічно, SSDNeRF залучає дифузійну модель для продукування ознак три-площини з їх подальшою трансформацією у формат NeRF. Використання нейронних представлень у поєднанні з генеративними моделями дозволило досягти високого рівня фотореалізму та геометричної складності при збереженні суворої багатовидової узгодженості.

Сучасний ландшафт методів 3D-генерації можна класифікувати на чотири фундаментальні категорії відповідно до їхніх алгоритмічних парадигм:

- Генерація з прямим поширенням (Feedforward generation): формування результату здійснюється за один ітераційний прохід через мережу.

- Генерація на основі оптимізації (Optimization-based generation): передбачає індивідуальну ітераційну оптимізацію для кожного об'єкта безпосередньо під час тестування.

- Процедурна генерація (Procedural generation): базується на детермінованих або стохастичних наборах правил.

- Генеративний синтез нових видів (Generative novel view synthesis): концентрується на створенні багатовидових зображень, уникаючи експліцитного формування 3D-геометрії.

Еволюційна траєкторія цих методів відображена на рисунках 3.8, 3.10 – 3.12, де простежуються основні генеалогічні гілки розвитку технологій та взаємозв'язок між ключовими науковими працями.

Генерація з прямим поширенням (Feedforward Generation) - підхід є домінуючим технічним вектором, що дозволяє безпосередньо синтезувати 3D-представлення. Систематизація цих методів здійснюється на основі типу базової генеративної моделі (рис. 3.9), що включає GANs, дифузійні та авторегресійні моделі, VAEs, а також нормалізуючі потоки.

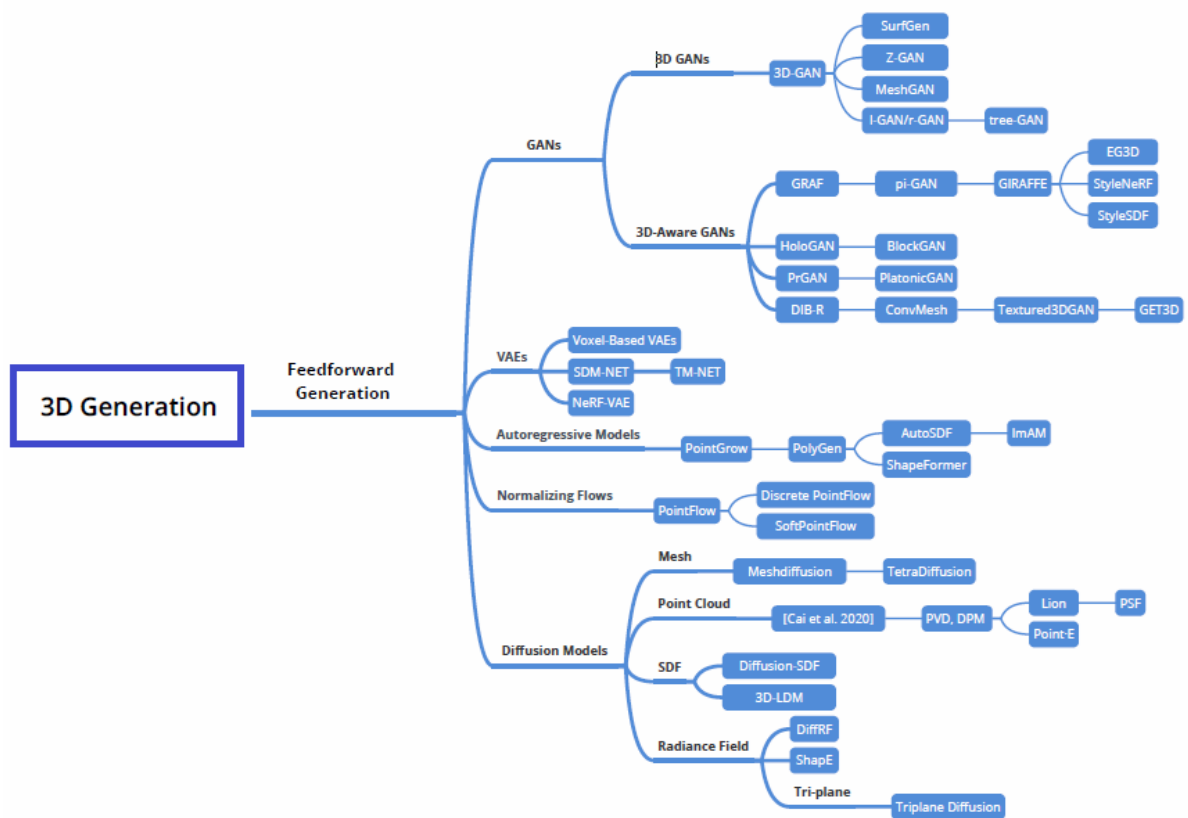


Рис. 3.8. Класифікація методів 3D генерації з прямим поширенням

Рисунок 3.9 ілюструє різноманітність архітектурних підходів, які дозволяють отримувати 3D-представлення за один прохід через мережу без необхідності ітераційної оптимізації під час тестування.

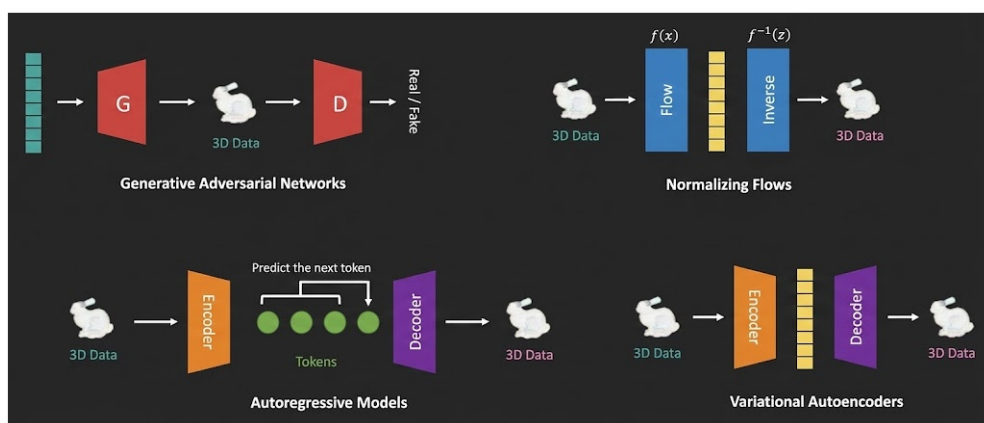


Рис. 3.9. Репрезентативні моделі 3D-генерації з прямим поширенням

На рис. 3.9 представлено кілька характерних технологічних конвеєрів (pipelines) моделей генерації тривимірного контенту за принципом прямого

поширення, зокрема: генеративно-змагальні мережі; авторегресійні моделі; варіаційні автокодувальники; нормалізуючі потоки.

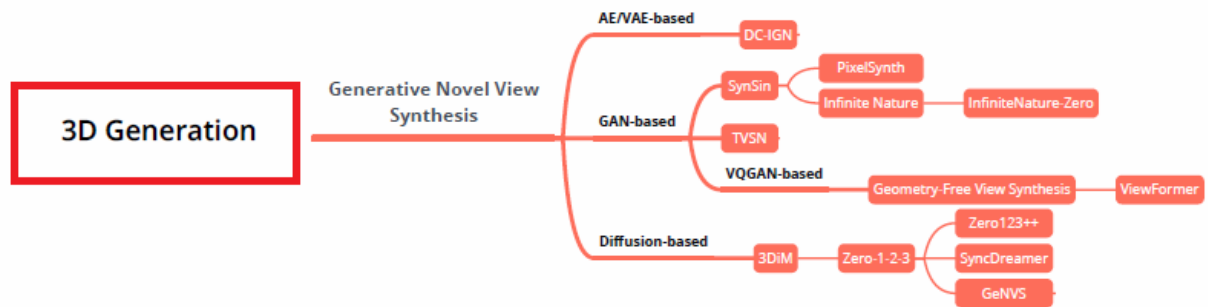


Рис. 3.10. Класифікація методів 3D генерації на основі генеративного синтезу НОВИХ ВИДІВ

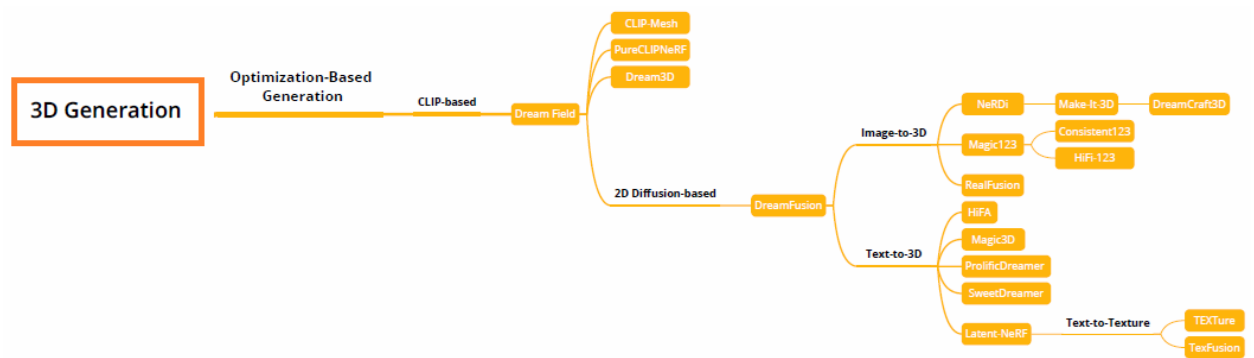


Рис. 3.11. Класифікація методів 3D генерації на основі оптимізації

Архітектура GAN продемонструвала виняткову ефективність у завданнях синтезу, базуючись на антагоністичній взаємодії двох компонентів: генератора $G(\cdot)$ та дискримінатора $D(\cdot)$. Генератор G трансформує вхідний латентний код z у синтетичні дані, тоді як дискримінатор D намагається диференціювати згенеровані зразки від автентичних даних з реальної вибірки. Спільна оптимізація G та D у процесі навчання стимулює генератор створювати контент, що за своїми статистичними характеристиками максимально наближений до реального.

Екстраполяція успіху 2D-GAN на тривимірний простір призвела до появи моделей, адаптованих під різні 3D-представлення:

- Хмари точок: l-GAN, r-GAN, tree-GAN.

- Воксельні сітки: 3D-GAN, Z-GAN.
- Полігональні сітки: MeshGAN.
- Функції знакового відстані (SDF): SurfGen, SDFStyleGAN.

Процес генерації в даному контексті розглядається як послідовність змагальних ітерацій. Ранні ітерації моделей (1-GAN, 3D-GAN) використовували безпосередній нагляд (supervision) на основі експліцитних тривимірних даних. Дискримінатори таких мереж аналізували безпосередньо 3D-структури, направляючи генератор до створення реалістичних вокселів або хмар точок. Більш прогресивні розробки, такі як SurfGen, використовують проміжні імпліцитні представлення, які згодом конвертуються в експліцитні формати, що забезпечує вищу якість фінальної геометрії.

Паралельно з методами прямого 3D-нагляду виник напрям 3D-обізнаних GAN (3D-aware GANs), що використовують 2D-нагляд через механізми диференційованого рендерингу. Це дозволяє моделям засвоювати просторові закономірності, спираючись на величезні масиви доступних 2D-зображень. У цій парадигмі генератор формує 2D-проекції з 3D-представлення, а дискримінатор порівнює ці рендеринги з реальними фотографіями.

Серед ключових архітектур варто виділити:

- HoloGAN та BlockGAN: використовують 3D-карти ознак, що проєктуються у 2D-простір відповідно до параметрів камери. BlockGAN додатково реалізує концепцію декомпозиції сцени на окремі об'єкти та фон.
- PrGAN та PlatonicGAN: базуються на воксельних структурах із застосуванням спеціалізованих шарів рендерингу.
- GET3D, DIBR, ConvMesh: спеціалізуються на генерації трикутників сіток та текстур виключно за допомогою 2D-зображень.
- GRAF та pi-GAN: інтегрують GAN з полями радіації (NeRF) та періодичними функціями активації (SIREN) для досягнення високої роздільної здатності та керованості синтезу.

- GIRAFFE: представляє сцену як композицію незалежних генеративних нейронних полів.

- EG3D: впроваджує гібридне представлення «три-площини», яке поєднує обчислювальну ефективність із високою експресивністю, ставши фундаментом для багатьох подальших досліджень у галузі 3D-зору.

3.5. Математичні парадигми імовірнісного моделювання та послідовного синтезу в 3D-генерації

У контексті синтезу тривимірних об'єктів, 3D-модель можна інтерпретувати як спільну ймовірність просторового розподілу множини елементарних компонентів:

$$p(x_0, x_1, \dots, x_n)$$

де x_i позначає i -й елемент структури, наприклад, координату точки або воксельний дескриптор. Оскільки пряма оцінка спільної ймовірності для великої кількості змінних є обчислювально складною, її факторизують у добуток умовних ймовірностей:

$$p(x_0, x_1, \dots, x_n) = p(x_0) \prod_{i=1}^n p(x_i | x_{<i})$$

Така декомпозиція дозволяє апроксимувати складні розподіли шляхом послідовного навчання та вибірки (sampling). Авторегресійні моделі визначають поточний стан системи на основі ретроспективних даних: припускаючи наявність упорядкованої послідовності $\{x_0, \dots, x_n\}$, функція моделі f навчається передбачати ймовірність появи наступного елемента x_i за умови попередньо згенерованого контексту $x_{<i}$:

$$p(x_i|x_{<i}) = f(x_0, \dots, x_{i-1})$$

Ця парадигма навчання відома як «teacher forcing». Під час інференції генерація здійснюється ітераційно за принципом максимізації ймовірності: $x_i = \arg \max p(x_i|x_{<i})$.

Сучасні архітектури, зокрема трансформери типу GPT, продемонстрували високу ефективність у задачах обробки природної мови та візуального синтезу, що стимулювало їх впровадження у 3D-генерацію.

Ключові дослідження в цій галузі включають:

- PointGrow: застосовує авторегресійну мережу з механізмами самоуваги (self-attention) для поточної генерації хмар точок, де координати кожної осі передбачаються через три спеціалізовані гілки.

- PolyGen: використовує каскад двох трансформерів — перший ітераційно генерує координати вершин, а другий формує топологію граней (faces), спираючись на отриманий контекст вершин.

- AutoSDF: оперує представленням у вигляді обрізаної функції знакової відстані (T-SDF). Використовуючи VQ-VAE для квантування локальних регіонів T-SDF, модель навчає трансформер передбачати токени кодексу в не послідовному авторегресійному режимі, що дозволяє відновлювати форми за текстовими або візуальними підказками.

- ShapeFormer: базується на квантованій глибокій неявній функції (VQDIF) для реконструкції повних поверхонь з неповних або зашумлених хмар точок.

- ImAM (Improved Autoregressive Model): пропонує проєкцію 3D-вокселів у три вирівняні за осями площини з подальшим стисненням в одновимірний простір дискретних токенів, що значно підвищує ефективність навчання.

Варіаційні автокодувальники являють собою ймовірнісні архітектури, що інтегрують кодер, який відображає дані у латентний простір параметрів

варіаційного розподілу, та декодер, що здійснює зворотну трансформацію для синтезу нових зразків. Навчання реалізується через оптимізацію доказової нижньої межі (ELBO) з використанням методу перепараметризації (reparameterization trick).

SDM-Net та TMNet фокусуються на структурній генерації сіток, моделюючи об'єкт як сукупність деформованих частин, що забезпечує кращу інтерпретованість геометрії.

Переваги та недоліки: Завдяки цільовій функції, орієнтованій на реконструкцію, VAE демонструють вищу стабільність навчання порівняно з GAN, проте часто генерують менш деталізовані, «розмиті» результати.

Методи нормалізуючих потоків базуються на послідовності оборотних диференційованих перетворень, що трансформують простий базовий розподіл (наприклад, ізотропний гаусівський) у складний цільовий розподіл даних. Це дозволяє точно обчислювати логарифмічну правдоподібність та виконувати ефективну вибірку.

- PointFlow: використовує безперервні нормалізуючі потоки для моделювання спільного розподілу форм та точок, дозволяючи генерувати довільну кількість точок для кожної вибірки форми.

- Discrete PointFlow (DPF): оптимізує швидкість навчання та інференції шляхом переходу до дискретних потоків.

- SoftFlow: вирішує проблему моделювання тонких структур на низьковимірних многовидах шляхом оцінки умовного розподілу збурених (perturbated) вхідних даних.

3.6. Генерація на основі оптимізації

Парадигма генерації на основі оптимізації базується на ітераційному синтезі 3D-моделей безпосередньо під час виконання (test-time optimization). В основі цих методів лежить використання попередньо навчених мультимодальних нейронних мереж як потужних геометричних та візуальних

«пріорів». Головним завданням є забезпечення семантичної відповідності між вхідними запитами (промптами) та генерованим контентом при збереженні високої деталізації та варіативності. Нижче розглянуто основні підходи залежно від модальності керуючого сигналу.

3.6.1. Текстово-орієнтована генерація

Природна мова є найбільш гнучким інструментом дескрипції сцен. Методи Text-to-3D фокусуються на узгодженні текстових дескрипторів із зображеннями, отриманими за допомогою диференційованого рендерингу.

TANGO: Використовує модель CLIP для нагляду за диференційованим фізично коректним рендерингом (PBR), що дозволяє синтезувати високоякісні текстурні карти на фіксованих поверхнях.

DreamFusion та Score Distillation Sampling (SDS): Знаковим етапом стала розробка втрати SDS, яка трансформує мінімізацію похибки рендерингу в дистиляцію ймовірнісної щільності. Це дозволяє використовувати 2D-дифузійні моделі для оптимізації 3D-представлень (наприклад, NeRF або сіток) без необхідності у 3D-датасетах. Аналогічний підхід використовує Score Jacobian Chaining (SJC), інтерпретуючи передбачення дифузії як функцію оцінки (score function) логарифмічної ймовірності.

Прогресивні стратегії (Magic3D, Fantasia3D): Вводять концепцію «від грубого до точного» (coarse-to-fine). Magic3D спочатку оптимізує об'ємне представлення, а потім вдосконалює витягнуту сітку за допомогою диференційованого нейронного рендерингу. Fantasia3D використовує гібридне представлення DM Tet для явної оптимізації геометрії та розділення текстури й освітлення.

VSD та MVDream: ProlificDreamer впроваджує варіаційну дистиляцію оцінки (VSD), моделюючи розподіл сцен як систему часток. Це дозволяє подолати проблему надмірної насиченості та низької різноманітності, властиву SDS. MVDream вирішує проблему «ефекту Януса» (багатоликості)

шляхом тонкого налаштування багатовидових дифузійних моделей, що забезпечує просторову узгодженість.

3.6.2. Візуально-орієнтована генерація

Зображення надають більш прецизійну інформацію про деталі та зовнішній вигляд об'єктів порівняно з текстом. Сучасні методи Image-to-3D спрямовані на реконструкцію 3D-моделей, які максимально точно зберігають візуальні характеристики вхідного референсу.

NeuralLift-360, RealFusion та NeRDі: Використовують монокулярну оцінку глибини та текстову інверсію (textual inversion) для адаптації дифузійних пріорів під конкретне зображення, «піднімаючи» 2D-дані у 3D-сцену на базі NeRF.

Magic123 та Make-it-3D: Magic123 комбінує 2D-пріори з 3D-обізнаними пріорами моделі Zero-1-to-3 для створення текстурованих сіток. Make-it-3D фокусується на створенні високодеталізованих хмар точок на фінальному етапі вдосконалення.

DreamGaussian та 3DGS: Впровадження методу 3D Gaussian Splatting (3DGS) дозволило досягти рендерингу та генерації в реальному часі. DreamGaussian використовує ефективну двостадійну структуру: спочатку оптимізуються 3D-гаусіани за допомогою SDS, а потім витягується сітка з вдосконаленими текстурами в UV-просторі.

3.7. Методи синтезу тривимірних моделей з використанням генеративного моделювання антропоцентричних об'єктів

У контексті розвитку концепції метавсесвіту та вдосконалення технологій віртуальної соціальної взаємодії, галузь цифрової реконструкції та генерації 3D-аватарів людини набула значного імпульсу. На відміну від загальних методів 3D-генерації, що оперують об'єктами з простою топологією, спеціалізовані алгоритми для моделювання людини спрямовані

на вирішення складних завдань: артикуляції поз та відтворення деталізованої геометрії одягу.

Систематизація сучасних методів базується на вхідних умовах (латентні вектори, референтні зображення, текстові промпти) та форматі вихідних даних. За типом виходу методи класифікують на:

- Генерація форми без текстури: акцент на анатомічній точності та складках одягу.

- Генерація текстурованого тіла: створення фотореалістичних цифрових двійників у одязі.

Генерація геометрії без текстури

Ранні підходи базувалися на оцінці параметрів моделі SMPL (Skinned Multi-Person Linear model) для виведення низькополігональної сітки. Проте фіксована топологія SMPL не дозволяє адекватно моделювати вільний одяг. Для вирішення цієї проблеми було запропоновано низку вдосконалень:

- CAPE: генеративна модель, що трансформує сітку SMPL залежно від параметрів пози та типу одягу.

- gDNA: фреймворк, що використовує змагальну функцію втрат для вивчення статистики складних деталей одягу зі сканованих датасетів.

- SMPLicit: імпліцитна модель, що генерує одяг як окремий шар поверх тіла SMPL.

- Chura: застосовує дифузійні моделі для генерації 2D-карт нормалей з наступною 3D-реконструкцією, що забезпечує безпрецедентний рівень деталізації поверхні.

Для створення повноцінних цифрових активів використовуються методи, що поєднують геометричну точність із фотореалістичними текстурами:

- Латентно-керовані методи: StylePeople (синергія StyleGAN та нейронного рендерингу), AvatarGen та GNARF.

- Текстово-керована генерація: Інтеграція пріорів із моделей text-to-image у 3D-конвеєри. Це дозволяє обійти обмеженість навчальних 3D-даних.

- Генерація за одним зображенням: Методи типу PIFu та PHORHUM демонструють високу точність на основі навчених нейронних функцій дистанції, тоді як TeSH та HumanRef використовують дифузійні моделі для «дофантазовування» невидимих частин тіла без потреби у специфічних 3D-датасетах.

Популярні завдання можна вільно класифікувати на три основні категорії, включаючи створення персоналізованих аватарів голови (наприклад, генерація говорячих голів), нейронні неявні 3D-морфовані моделі (3DMM) та генеративні 3D-моделі обличчя, які показані на рис. 3.12 і рис. 3.13.

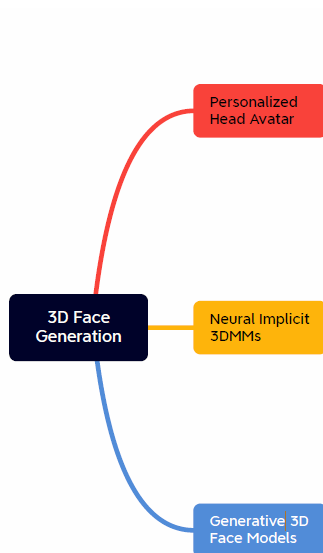


Рис. 3.12. Репрезентативні методи та прикладні аспекти генерації тривимірних облич

Центральним завданням у цій сфері є синтез багатовидово-узгоджених зображень обличчя з високою роздільною здатністю. Дослідження структуровані за трьома основними напрямками:

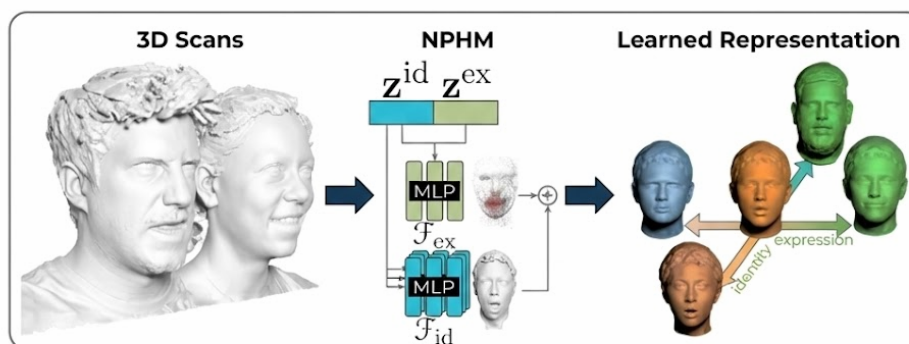
- Персоналізовані аватари голови: Створення анімованих «говорячих голів» на основі монокулярного відео. Сучасні підходи використовують нейронні поля радіації (NeRF) для забезпечення стабільності при великих кутах повороту голови. Якість оцінюється за метриками PSNR, SSIM та LPIPS.

- Нейронні імпліцитні 3DMM: На зміну традиційним лінійним морфованим моделям обличчя (3DMM) прийшли неперервні нейронні представлення (INR). Моделі типу NPHM здатні відтворювати мікроміміку та складні зачіски, що є недоступним для сіток із фіксованою топологією.

- Генеративні 3D-моделі обличчя (3D GANs): Розвиток від воксельних методів (HoloGAN) до гібридних архітектур (EG3D). Використання представлень три-площини дозволило досягти високої роздільної здатності та суворій геометричній узгодженості при зміні пози камери.



Generative 3D face models



Neural Implicit 3DMMs

Рис. 3.13. Репрезентативні завдання генерації тривимірних облич

Ці технології є фундаментом для інструментів 3D-редагування та створення цифрових акторів, суттєво знижуючи залежність від об'ємних наборів навчальних даних.

Аналіз сучасного стану галузі генерації тривимірною контенту дозволяє констатувати стрімку еволюцію методів: від класичних полігональних та воксельних представлень до складних нейронних

імпліцитних архітектур. На основі проведеного огляду можна зробити такі ключові висновки:

1. Домінування гібридних архітектур.

Найбільшу ефективність демонструють методи, що поєднують імовірнісні моделі (дифузійні моделі, GAN) з нейронними представленнями сцен (NeRF, три-площини). Зокрема, використання диференційованого рендерингу дозволило подолати дефіцит 3D-даних, використовуючи масиви 2D-зображень для навчання об'ємно-узгоджених моделей.

2. Зміна парадигм нагляду.

Спостерігається перехід від прямого 3D-нагляду до методів дистиляції ознак, таких як Score Distillation Sampling (SDS). Це відкрило шлях до створення високодеталізованих 3D-активів на основі текстових запитів та поодиноких зображень, що раніше було неможливим без складного ручного моделювання.

3. Антропоцентричне моделювання.

Генерація цифрових людей та облич виділилася в окремий високотехнологічний сегмент. Впровадження нейронних морфованих моделей (Neural Implicit 3DMM) дозволило подолати обмеження фіксованої топології традиційних методів, забезпечуючи реалістичну анімацію міміки та складних структур, таких як волосся та елементи одягу.

4. Прогрес у реальному часі.

Поява технології 3D Gaussian Splatting (3DGS) ознаменувала новий етап у розвитку галузі, дозволяючи поєднувати якість нейронних полів із продуктивністю класичної растрової графіки. Це робить можливим інтеграцію генеративних моделей у системи віртуальної та доповненої реальності в реальному часі.

5. Перспективи розвитку.

Основними викликами залишаються забезпечення повної редагованості згенерованих об'єктів (зокрема, топологічна гнучкість сіток) та покращення семантичної узгодженості при генерації складних багатооб'єктних сцен.

Подальші дослідження, імовірно, будуть зосереджені на інтеграції мультимодальних великих мовних моделей (LLMs) для більш точного керування геометрією та фізичними властивостями матеріалів.

Висновки до розділу

У третьому розділі виконано комплексний аналіз сучасних підходів до генерації тривимірних моделей із використанням генеративного моделювання. Розглянуто еволюцію генеративних моделей від двовимірного до тривимірного простору та визначено ключові етапи цього переходу. Проаналізовано методи прямої 3D-генерації, процедурні підходи та генеративний синтез нових видів як альтернативні стратегії створення 3D-контенту. Показано, що вибір методу генерації тісно пов'язаний зі способом представлення сцени. Детально охарактеризовано експліцитні, сіткові та багатосарові представлення, а також їх роль у задачах рендерингу. Розкрито значення імовірнісних моделей і послідовного синтезу для підвищення узагальнювальної здатності генеративних систем. Проаналізовано підходи до генерації на основі оптимізації, зокрема текстово- та візуально-орієнтовані методи. Окрему увагу приділено специфіці синтезу антропоцентричних об'єктів, що характеризуються високими вимогами до геометричної та семантичної узгодженості. Показано, що сучасні генеративні моделі здатні доповнювати класичні методи реконструкції. Узагальнення результатів розділу підтверджує перспективність інтеграції традиційних і неймережевих підходів у задачах генерації 3D-моделей.

ВИСНОВКИ

У магістерській роботі досліджено теоретичні, математичні та алгоритмічні засади генерації тривимірних моделей із двовимірних зображень, а також сучасні підходи до синтезу 3D-контенту з використанням класичних методів комп'ютерного зору та генеративних моделей штучного інтелекту. Отримані результати підтверджують актуальність обраної теми в контексті розвитку комп'ютерної графіки, доповненої та віртуальної реальності, цифрових двійників і автоматизованих систем проєктування.

У першому розділі здійснено ґрунтовний аналіз предметної області 3D-реконструкції із зображень. Розглянуто фундаментальні принципи комп'ютерного зору, що лежать в основі вилучення просторової структури сцени, зокрема методи реконструкції за множиною зображень (Structure from Motion, Multi-View Stereo). Детально проаналізовано основні етапи 3D-реконструкції, включно з детекцією та зіставленням ознак, оцінюванням параметрів камер і відновленням просторових координат. Особливу увагу приділено підходам, спрямованим на уникнення або мінімізацію необхідності попередньої калібрування камер, що підвищує практичну застосовність методів у реальних умовах. Обґрунтовано роль математичного апарату, зокрема епіпольярної геометрії та алгоритму RANSAC, як ключових інструментів забезпечення стійкості та точності реконструкції в умовах шуму й наявності викидів. У результаті першого розділу сформовано теоретичну та методологічну базу для подальшого дослідження об'ємних методів генерації 3D-моделей.

У другому розділі зосереджено увагу на методології фотоконсистентних критеріїв та об'ємних підходах до генерації тривимірних моделей об'єктів. Проведено аналіз конфігурацій збору вхідних даних і формалізовано задачу 3D-реконструкції в об'ємному просторі. Розглянуто класичні об'ємні методи, зокрема реконструкцію візуальної оболонки, заповнення вокселів і метод фарбування вокселів, що дозволяють

відновлювати геометрію сцени на основі узгодженості проєкцій у множині зображень. Детально описано етапи реалізації методу фарбування вокселів та проаналізовано метрики фотоконсистентності, які визначають якість та достовірність реконструйованої моделі. Результати другого розділу демонструють, що фотоконсистентні критерії є ефективним інструментом для побудови щільних об'ємних моделей, однак потребують значних обчислювальних ресурсів і чутливі до умов освітлення та якості вхідних даних.

У третьому розділі виконано систематичний огляд та імплементаційний аналіз сучасних методів генерації 3D-контенту. Розкрито еволюцію генеративних моделей від двовимірного до тривимірного простору та охарактеризовано основні парадигми 3D-генерації, зокрема пряму генерацію геометрії, процедурну генерацію, генеративний синтез нових видів та підходи, засновані на оптимізації. Значну увагу приділено способам представлення сцени, включно з експліцитними моделями, сітками та багатшаровими (неявними) представленнями, а також їх ролі в задачах рендерингу та синтезу 3D-моделей. Проаналізовано математичні основи імовірнісного моделювання та послідовного синтезу, що лежать в основі сучасних генеративних моделей. Окремо розглянуто текстово- та візуально-орієнтовані методи генерації, а також специфіку синтезу антропоцентричних об'єктів, що має важливе значення для застосувань у сфері цифрових аватарів, ігор та віртуальних середовищ.

Загалом результати магістерської роботи дозволяють зробити висновок, що генерація тривимірних моделей із зображень є міждисциплінарною задачею, яка поєднує методи комп'ютерного зору, математичного моделювання та штучного інтелекту. Класичні алгоритми 3D-реконструкції забезпечують геометричну коректність і інтерпретованість результатів, тоді як сучасні генеративні підходи розширюють можливості синтезу складних і семантично насичених 3D-сцен. Практичне значення отриманих результатів полягає у можливості їх використання при розробці

систем автоматизованої 3D-реконструкції та генерації контенту для комп'ютерної графіки, віртуальної реальності, цифрового дизайну та суміжних галузей.

Перспективи подальших досліджень пов'язані з інтеграцією фотоконсистентних методів із сучасними генеративними моделями, підвищенням масштабованості та ефективності алгоритмів, а також адаптацією розглянутих підходів до роботи в умовах обмежених або неповних вхідних даних.

ПЕРЕЛІК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Efficient Geometry-aware 3D Generative Adversarial Networks / E. R. Chan et al. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2022. P. 16123–16133.
2. DreamFusion: Text-to-3D using 2D Diffusion / B. Poole et al. International Conference on Learning Representations (ICLR). 2023. URL: <https://arxiv.org/abs/2209.14988>.
3. 3D Gaussian Splatting for Real-Time Radiance Field Rendering / B. Kerbl et al. ACM Transactions on Graphics. 2023. Vol. 42, No. 4. P. 1–14.
4. AutoSDF: Shape Priors for 3D Completion, Reconstruction and Generation / P. Mittal et al. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2022. P. 306–315.
5. PointFlow: 3D Point Cloud Generation with Continuous Normalizing Flows / G. Yang et al. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). 2019. P. 4541–4550.
6. ProlificDreamer: High-Fidelity and Diverse Text-to-3D Generation with Variational Score Distillation / Z. Wang et al. Advances in Neural Information Processing Systems (NeurIPS). 2023. Vol. 36.
7. PolyGen: An Autoregressive Model for 3D Mesh Generation / C. Nash et al. International Conference on Machine Learning (ICML). 2020. P. 7220–7229.
8. PIFu: Pixel-aligned Implicit Function for High-resolution Clothed Human Digitization / S. Saito et al. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). 2019. P. 2304–2314.
9. Magic3D: High-Resolution Text-to-3D Content Creation / C. Lin et al. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2023. P. 300–309.

10. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis / B. Mildenhall et al. European Conference on Computer Vision (ECCV). 2020. P. 405–421.
11. SMPL: A Skinned Multi-Person Linear Model / M. Loper et al. ACM Transactions on Graphics (TOG). 2015. Vol. 34, No. 6. P. 1–16.
12. DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation / J. J. Park et al. CVPR. 2019. P. 165–174.
13. GET3D: A Generative Model of High-Quality 3D Textured Shapes Learned from 2D Images / J. Gao et al. NeurIPS. 2022.
14. GIRAFFE: Controllable Image Synthesis with Compositional Neural Radiance Fields / M. Niemeyer, A. Geiger. CVPR. 2021. P. 11435–11445.
15. HoloGAN: Unsupervised Learning of 3D Representations from Natural Images / T. Nguyen-Phuoc et al. ICCV. 2019. P. 7588–7597.
16. StylePeople: A Generative Model of Full-Body Human Appearance / S. Grigorev et al. CVPR. 2021. P. 5129–5139.
17. AvatarCLIP: Zero-Shot Text-Driven Generation and Animation of 3D Avatars / Z. Hong et al. ACM Transactions on Graphics. 2022. Vol. 41, No. 4.
18. DreamGaussian: Generative 3D Gaussian Splatting Guided by Image Diffusion / J. Tang et al. arXiv preprint arXiv:2309.16657. 2023.
19. Magic123: One Image to High-Quality 3D Object Generation / G. Qian et al. ICCV. 2023.
20. Make-It-3D: High-Fidelity 3D Creation from a Single Image / J. Tang et al. ICCV. 2023.
21. MVDream: Multi-view Diffusion for 3D Generation / X. Pan et al. arXiv preprint arXiv:2308.16512. 2023.
22. NPHM: Neural Parametric Head Models / S. Geng et al. CVPR. 2023. P. 22474–22484.
23. EG3D: Efficient Geometry-aware 3D Generative Adversarial Networks / E. R. Chan et al. CVPR. 2022.

24. Fantasia3D: Disentangling Geometry and Appearance for High-quality Text-to-3D Content Creation / R. Chen et al. ICCV. 2023.
25. 3DMM: A Morphable Model for the Synthesis of 3D Faces / V. Blanz, T. Vetter. SIGGRAPH. 1999. P. 187–194.
26. PointGrow: Autoregressive Point Cloud Generation with Contextual Awareness / J. Sun et al. CVPR. 2020. P. 12620–12629.
27. SoftFlow: Probabilistic Framework for Normalizing Flow on Manifolds / H. Kim et al. NeurIPS. 2020.
28. Occupancy Networks: Learning 3D Reconstruction in Function Space / L. Mescheder et al. CVPR. 2019. P. 4460–4470.
29. TeCH: Text-guided Reconstruction of Lifelike Clothed Humans / Y. Huang et al. International Conference on 3D Vision (3DV). 2024.
30. HumanRef: High-Fidelity 3D Human Reconstruction from Sparse Views with Semantic Prior / L. Liu et al. CVPR. 2023.
31. TADA! Text to Animatable Digital Avatars / T. Liao et al. International Conference on 3D Vision (3DV). 2024.
32. Score Jacobian Chaining: Lifting Pretrained 2D Diffusion Models for 3D Generation / H. Wang et al. CVPR. 2023.
33. SJC: Score Jacobian Chaining for 3D Generation / J. Wang et al. CVPR. 2023.
34. RealFusion: 3D Reconstruction of Any Object from a Single Image / L. Melas-Kyriazi et al. CVPR. 2023.
35. NeuralLift-360: Lifting an In-the-wild 2D Photo to a 3D Object / J. Xu et al. CVPR. 2023.