

2. АНАЛІЗ МЕТОДІВ 3D РЕКОНСТРУКЦІЇ СЕРЕДОВИЩА ДЛЯ ДОПОВНЕНОЇ РЕАЛЬНОСТІ

Володимир Савін, аспірант
Кафедра математичного моделювання і аналізу даних
Навчально-науковий Фізико-технічний інститут
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

vladimir.savin@gmail.com

ВСТУП

В останнє десятиліття спостерігається надзвичайно швидкий розвиток технологій доповненої реальності (ДР), які відкривають новий етап взаємодії людини з інформацією та оточуючим середовищем. Сценарії використання цих технологій стрімко еволюціонують, переходячи з мобільних на самостійні носимі пристрої. Продукти, які дозволяють споживати сценарії ДР, виходять зі стадії обмежених та дорогих платформ розробки і набувають масового розповсюдження. Розмір світового ринку ДР наразі становить 25,1 млрд \$ (2023 рік). При прогнозованому загальному річному темпі зростання (CAGR) у 23,2% він має перевищити 70 млрд \$ у 2028 році [1].

Сучасні тренди у галузі доповненої реальності визначаються не лише зростанням її популярності, але й постійним удосконаленням технологічного арсеналу. Все більше компаній і дослідницьких груп зосереджують свою увагу на розробці інноваційних методів та рішень спрямованих на підвищення реалістичності взаємодії об'єктів ДР з реальним світом та користувача з об'єктами ДР. Інтенсивність розвитку цього напрямку не тільки збільшує можливості технологій ДР, але і відкриває нові перспективи для їхнього впровадження у різноманітних сферах життя, від освіти та розваг до промисловості та медицини.

Один із ключових напрямків цього еволюційного процесу є 3D реконструкція оточуючого середовища та його об'єктів, що стає невід'ємною складовою при досягненні максимальної імерсії

та реалізації потенціалу ДР. Мета 3D реконструкції полягає у відтворенні тривимірної моделі оточення на основі доступних даних: дво- та тривимірні зображення, внутрішні та зовнішні параметри камери, дані з IMU сенсорів, або систем позиціонування і т.д. 3D реконструкція є своєрідним каркасом для доповненої реальності. Якість отриманої моделі суттєво впливає на реалістичність сприйняття сценаріїв доповненої реальності користувачем. Справжній виклик полягає в тому, як забезпечити точність та достовірність тривимірної моделі. В цьому контексті, доволі часто, важливу роль відіграє оцінка карти глибини. Методи реконструкції карти глибини приймають на вхід монокулярне зображення, стереозображення, або їх послідовність, та повертають результуюче зображення, де кожний піксель представляє собою відносну, або абсолютну відстань від камери до відповідної точки простору. Оцінка глибини дозволяє системам реконструкції ефективно розміщувати об'єкти в просторі, враховуючи їхнє точне положення та взаємодію. Це стає критичним у віртуальних і доповнених середовищах, де недостовірність 3D реконструкції може призвести до неправильної взаємодії з оточенням та втрати реалістичності.

Цей науковий огляд покликаний систематизувати знання, щодо доповненої реальності, 3D реконструкції та оцінки глибини, а також проаналізувати поточний стан досліджень у цих інноваційних напрямках.

2.1. ТРИВИМІРНА РЕКОНСТРУКЦІЯ ЯК ВАЖЛИВА СКЛАДОВА ДОПОВНЕНОЇ РЕАЛЬНОСТІ

Доповнена реальність представляє собою технологію, яка дозволяє в режимі реального часу відображати штучний 2D/3D контент та інформацію поверх або поруч фізичних об'єктів реального світу. За рахунок алгоритмічного стеку, який включає в себе: супроводження об'єктів та відстеження позиції пристрою ДР в 6-ти ступенях свободи, аналіз карти глибини та 3D реконструкцію, розпізнавання об'єктів та семантичне розуміння сцени, виявлення джерел освітлення та рендерингу штучних тіней, а також багатьох інших технологій, доповнені об'єкти природньо інтегруються у фізичне середовище. Це дає змогу підвищити реалістичність сценаріїв доповненої реальності та сприймати штучні об'єкти як частину оточення.

1.2. Аналіз методів 3D реконструкції середовища для доповненої...

На відміну від віртуальної реальності, де користувач занурюється у повністю штучне середовище, доповнена реальність надає можливість взаємодії зі штучним контентом не втрачаючи зв'язок з оточенням. Доповнені 2D/3D об'єкти прив'язуються до певних фізичних об'єктів, взаємодіють з ними та з іншими предметами сцени за рахунок оклюзій та колізій (рис. 1). Реалістичність таких взаємодій залежить від якості «каркасу», на який вони накладаються. Цим каркасом виступає 3D реконструкція середовища.

Останнім часом доповнена реальність широко застосовується в різноманітних галузях, що включають: освіту, медицину, інженерію, зв'язок та віддалену підтримку, сферу розваг та багатьох інших напрямків. Також, доповнена реальність використовується і у військових цілях.

На сьогоднішній день найдоступнішим пристроєм, який дозволяє споживати сценарії ДР, виступає мобільний телефон. Зображення, отримане з камери, ретранслюється на екран та доповнюється штучними об'єктами. Основні недоліки такої експлуатації: телефон треба тримати у руках, користувач спостерігає лише невеличку область сцени, відсутність реалістичного сприйняття середовища.



Рис. 1. Приклад сценарію ДР з урахуванням оклюзій та колізій

Стрімкий розвиток носимих пристроїв ДР націлений на позбавлення вищезгаданих недоліків та на популяризацію технології. За формфактором носимі пристрої ДР можна розділити на:

- Формфактор шолому віртуальної реальності (рис. 2, а). Пара камер ретранслює зображення середовища, яке доповнюється, на пару екранів високої роздільної здатності та частоти, що знаходяться в середині шолому перед очима

користувача. Додатковою перевагою цієї конструкції є можливість споживання сценаріїв віртуальної реальності. Основні недоліки: великий розмір та вага, що впливають на ергономічність використання.

- Формфактор окулярів (рис. 2, б). Зображення з камер передаються на прозорі дисплеї, наприклад Waveguide [2], Transparent MICRO LED [3], або інші. Користувач спостерігає оточуюче середовище крізь них. Доповнений контент відображається на дисплеях поверх реальних об'єктів. Основний недолік: мініатюрний розмір впливає на можливість розміщення апаратних засобів та батарей високої ємності.



Рис. 2. Приклад носимих пристроїв ДР: формфактор шолому віртуальної реальності (а) – Meta Quest Pro, формфактор окулярів (б) – XREAL Air 2 Pro

За критерієм розміщення апаратно-обчислювальних потужностей носимі пристрої ДР можна розділити на:

- Самостійні (рис. 2). Апаратно-обчислювальні потужності розміщені на самому пристрої.
- Пов'язані (рис. 3). Апаратно-обчислювальні потужності частково або повністю розміщені на пристрої-компаньйоні. Пристрої можуть з'єднуватись як провідним, так і безпроводним шляхом.



Рис. 3. Приклад пов'язаних пристроїв ДР (Magic Leap 2)

Сучасна тенденція розвитку пристроїв ДР спрямована на мініатюризацію з метою перетворити громіздкий гаджет на ергономічний повсякденний аксесуар та помічник. Однак зменшення розмірів змушує використовувати батареї невеликої ємності. Це, в свою чергу, впливає на можливість встановити або активно використовувати сенсори з підвищеним енергоспоживанням, такі як: ToF (Time of Flight) [4], LIDAR [5] та сенсори структурного світла [6]. Отже, доволі часто, реконструкцію карт глибини та/або 3D реконструкцію сцени доводиться проводити на основі даних з монокулярної, або стереокамери, що є доцільніше з точки зору оптимізації енергоспоживання.

2.2. МЕТОДИ РЕКОНСТРУКЦІЇ КАРТИ ГЛИБИНИ ПО 2D ДАНИМ

Важливою складовою 3D реконструкції є розуміння карти глибини (відстані до кожної точки спостереження). Людині зазвичай легко сприймати інформацію про тривимірну структуру об'єкта або сцени та оцінювати відстань до об'єктів. Але визначення карти глибини по зображенню або їх серії є складною задачею комп'ютерного зору, оскільки під час зйомки відбувається проєкція сцени на площину, що призводить до втрати третього виміру.

Методи реконструкції карти глибини з 2D даних можна умовно поділити на два класи в залежності від кількості вхідних зображень:

- методи, що використовують одне вхідне нерухоме зображення та спираються на монокулярні ознаки глибини;
- методи, що базуються на аналізі двох і більше зображень та оперують багатоокулярними (multi-ocular) ознаками глибини.

У другому випадку два або більше вхідних зображень можуть бути зроблені кількома фіксованими камерами з різних кутів огляду або однією камерою з динамічними об'єктами на сцені.

Таблиця 2. Основні ознаки глибини, що використовуються методами реконструкції глибини по 2D даним

Кількість вхідних зображень	Ознаки глибини
Одне зображення	Лінійна перспектива
	Атмосферне розсіювання
	Розуміння форми по затіненню
Два і більше зображень	Бінокулярний диспаратет
	Паралакс руху
	Розмиття зображення
	Силует
	Структура з руху

2.2.1. ЛІНІЙНА ПЕРСПЕКТИВА

В основі лінійної перспективи лежить ідея, що паралельні лінії, такі як дороги або стежки, збігаються вдалині. Точки перетину цих ліній менш помітні, ніж точки ліній, наближені до спостерігача. Підхід, запропонований Battiato, Curti та ін. [7], працює для зображень, що містять поверхні з жорсткою геометрією. Точка з найбільшою кількістю перетинів у певному районі вважається точкою зникнення. Основні лінії поблизу точки зникнення позначаються як лінії зникнення (рис. 4).

Між кожною парою сусідніх ліній зникнення призначається набір градієнтних площин, кожна з яких відповідає окремому рівню глибини. Пікселі ближче до точок зникнення отримують більше значення глибини, і щільність градієнтних площин вища (рис. 5).



Рис. 4. Детектування ліній і точки зникнення [7]

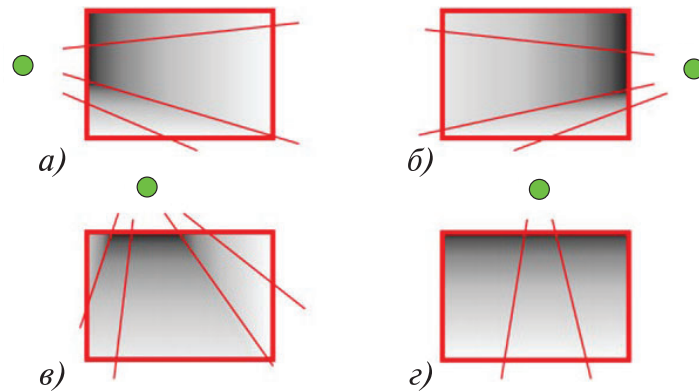


Рис. 5. Приклади евристичних правил для створення градієнтних площин глибини, де зелене коло – точка зникнення [7]

2.2.2. АТМОСФЕРНЕ РОЗСІЮВАННЯ

Підхід атмосферного розсіювання ґрунтується на тому, що потужність і напрям світла змінюються, коли світло проходить через атмосферу через наявність у ній дрібних частинок. Об'єкти, які знаходяться ближче до камери, виглядають чіткіше, тоді як більш віддалені об'єкти – розмиті. У 1997 році Krotkov та Cozman [8] представили аналіз цього перетворення, заснований на фізичній моделі розсіювання лорда Релея 1871 року. Їх алгоритм підходить для оцінки глибини зображень на відкритому повітрі, які містять частину неба (рис. 6).

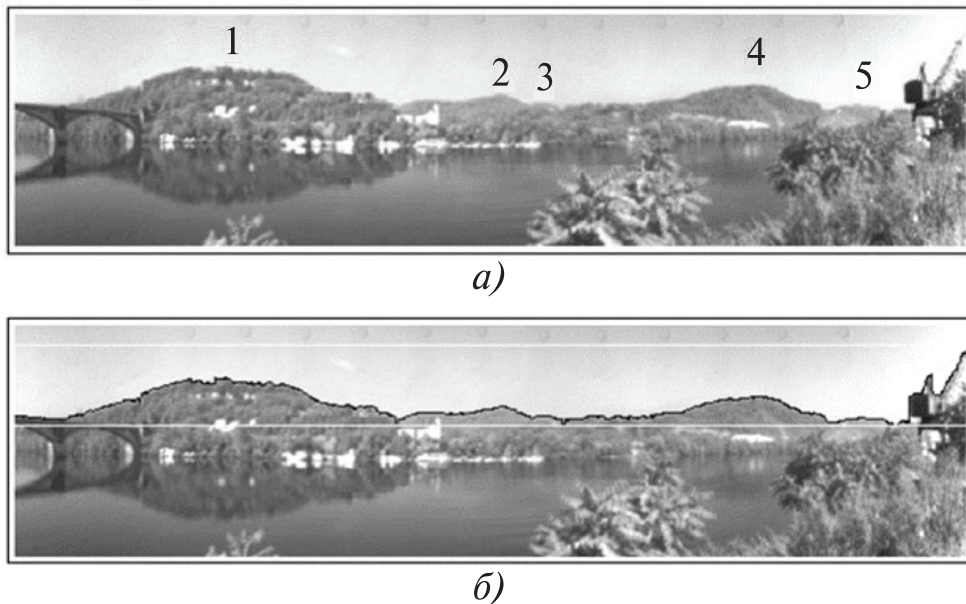


Рис. 6. Приклад детектування горизонту та зон атмосферного розсіювання [8]

2.2.3. РОЗУМІННЯ ФОРМИ ПО ЗАТІНЕННЮ

Метод розуміння форми по затіненню дозволяє визначити нормаль поверхні об'єкта, спостерігаючи за відбивною здатністю світла на цьому об'єкті. Кількість світла, яка відбивається від поверхні об'єкта, залежить від його орієнтації. Вперше цю ідею представив Woodham у 1980 році. Підхід розуміння форми по затіненню (shape from shading), використовується для аналізу одного вхідного зображення та був представлений В. К. Horn в 1989 році. Фотометричний стереоаналіз відтоді був узагальнений для багатьох інших ситуацій, таких як, наприклад, не Ламбертові поверхні. Процес реконструкції карти глибини по затіненню на основі зображень світлового поля представлений на рис. 7.

По декільком зображенням об'єкта при різному освітленні можна провести оцінку векторів нормалей у кожному пікселі [9].

Метод потребує специфічного технічного обладнання.

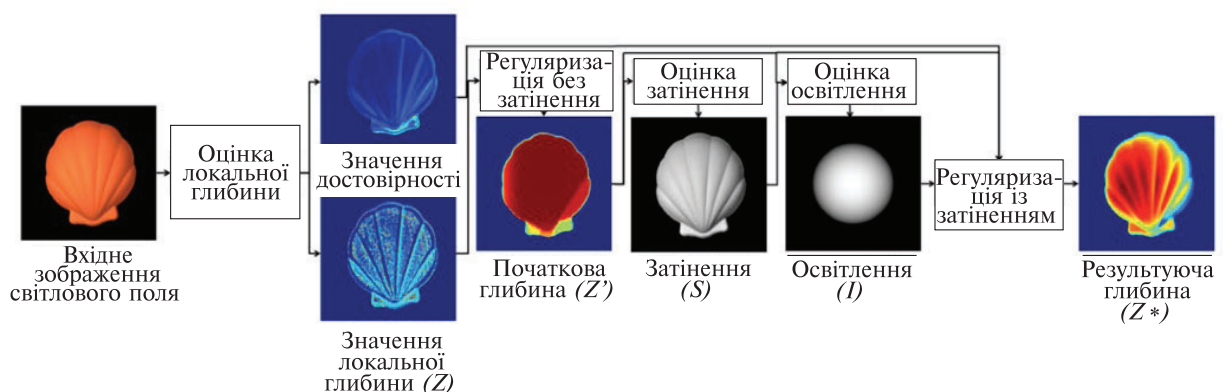


Рис. 7. Ілюстрація підходу відтворення глибини по затіненню на основі зображень світлового поля [9]

2.2.4. БІНОКУЛЯРНИЙ ДИСПАРИТЕТ

Використовуючи два зображення однієї сцени, зроблені одночасно з дещо різних точок спостереження, можна відновити глибину точки, що присутня на обох зображеннях. Спочатку знаходиться відповідний набір точок на обох зображеннях. Потім, для кожного набору точок використовується метод триангуляції для визначення глибини відповідної фізичної точки, що була спроектована на пару зображень [10, 11].

На рис. 8 проілюстрована система стереоскопічного зору для якої проводиться обрахунок бінокулярного диспаритету, де:

1.2. Аналіз методів 3D реконструкції середовища для доповненої...

- P – точка у просторі, що належить фізичному об'єкту;
- C_l та C_r – ліва та права камера стереосистеми;
- P_l та P_r – проєкції точки на матриці лівої (C_l) та правої камер (C_r). P_l та P_r знаходяться на епіполярній лінії;
- x_l та x_r – відповідні зсуви проєкції точки від початку системи координат зображення (x координати точок P_l та P_r);
- Z – значення глибини (відстань до P);
- f – фокальна відстань;
- B – відстань між камерами стереопари (baseline).

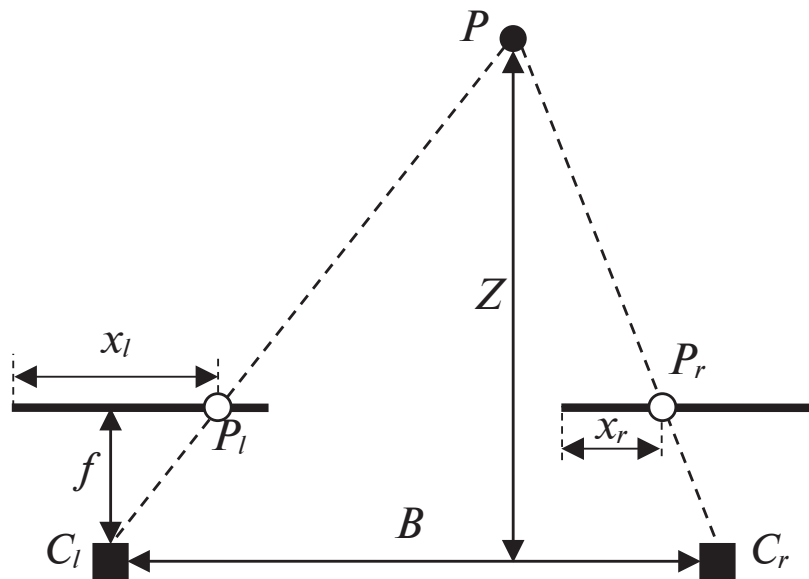


Рис. 8. Ілюстрація бінокулярного диспаритету

Значення глибини Z можна розрахувати спираючись на принципи триангуляції, що застосовуються в системах стереоскопічного зору:

$$Z = f \frac{B}{D} \quad (1)$$

де D – значення диспаритету, який розраховується як різниця між x_l та x_r .

2.2.5. ПАРАЛАКС РУХУ

Відносний рух між камерою і сценою надає важливі ознаки для сприйняття глибини. Об'єкти, які знаходяться ближче до камери, рухаються швидше, ніж ті, що знаходяться далі. Відновлення тривимірних структур називається реконструкцією структури з руху. Рух можна розглядати як форму диспаритету

протягом часу, що представлено поняттям поля руху. Поле руху — це двовимірні вектори швидкості точок зображення та спостережуваної сцени. Основні припущення для структури з руху полягають у тому, що об'єкти не деформуються і їхні рухи є лінійними. Ці властивості було використано у кількох методах, таких як «стереоскопія погойдування (wiggle stereoscopy)» [12], де паралакс руху використовується як представлення для стереоскопічних зображень, або «стереоскопія прокручування (parallax scrolling)» [13], що широко використовується в комп'ютерній графіці, де шляхом руху переднього і заднього планів з різною швидкістю викликається відчуття глибини. Вплив цієї ознаки глибини є відносно сильним в порівнянні з іншими монокулярними ознаками глибини, а також у порівнянні з бінокулярним диспаратетом.

2.2.6. РОЗМИТТЯ ЗОБРАЖЕННЯ

Методи визначення глибини за розмиттям зображення дозволяють реконструювати карту глибини на основі ступеня розмиття, присутнього на зображеннях. У системі з тонкою лінзою об'єкти, які знаходяться у фокусі, відображаються чітко, тоді як об'єкти на інших відстанях розфокусовані, тобто розмиті. На рис. 9 показана модель тонкої лінзи з реальною точкою P , що знаходиться за межами фокусної відстані лінзи. Відповідна проєкція на площину зображення є кругова розмита пляма зі сталою яскравістю, центрованою у точці P'' та радіусом розмиття σ .

Для того щоб оцінити глибину u , нам необхідні наступні рівняння. Основне рівняння, що описує співвідношення між u , v та f для тонких лінз (2):

$$\frac{1}{u} + \frac{1}{v} = \frac{1}{f}. \quad (2)$$

Pentland [14] вивів співвідношення між відстанню u та розмиттям σ у рівнянні (3):

$$u = \begin{cases} \frac{fs}{s - f - kf\sigma}, & u > v, \\ \frac{fs}{s - f + kf\sigma}, & u < v, \end{cases} \quad (3)$$

1.2. Аналіз методів 3D реконструкції середовища для доповненої...

де: u – глибина, v – відстань між лінзою та точкою ідеального фокусу, s – відстань між лінзою та площиною зображення, f – фокусна відстань лінзи, k – константа, визначена оптичною системою, σ – радіус розмиття.

Задача обчислення глибини u перетворюється на задачу оцінки параметрів камери (s , f та k) та параметра розмиття σ . Параметри камери можуть бути отримані шляхом її калібрування. Глибину u можна обчислити з рівняння (3), якщо відомий параметр розмиття σ .

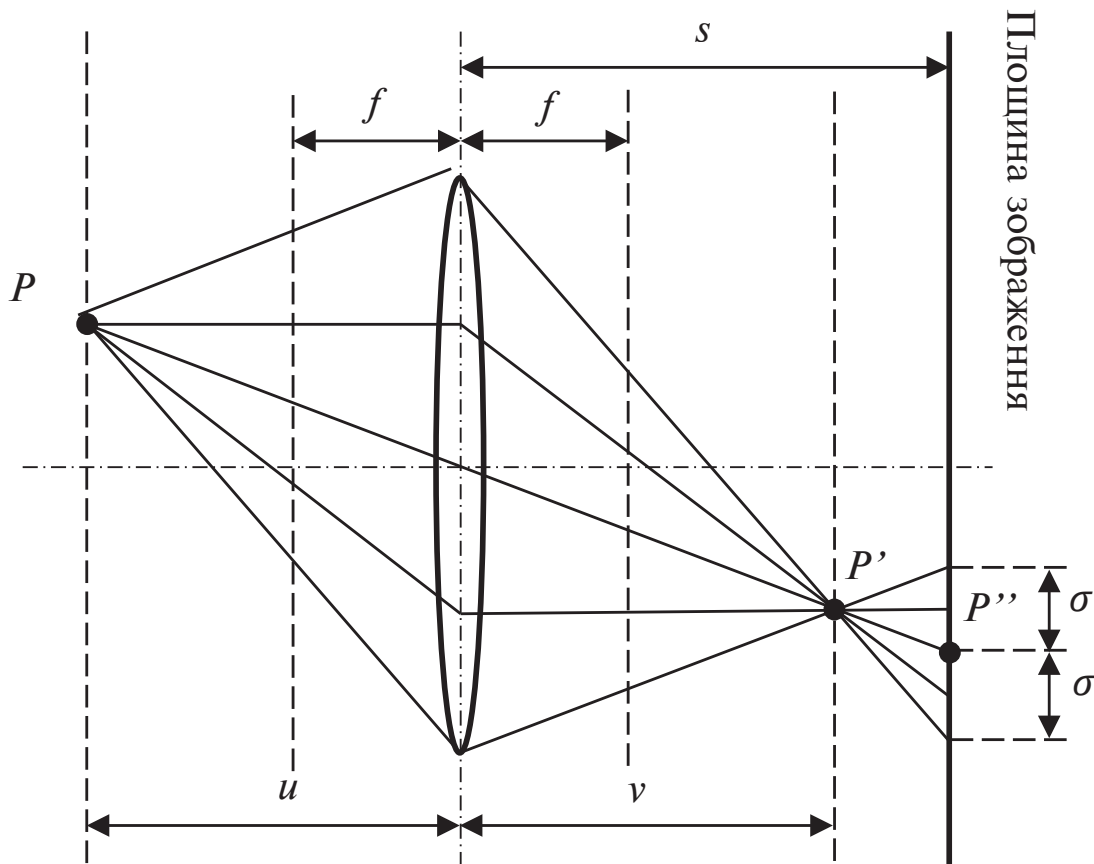


Рис. 9. Ілюстрація моделі тонкої лінзи

2.2.7. СИЛУЕТ (SHAPE-FROM-SILHOUETTE)

Силует об'єкта на зображенні відноситься до контуру, який відділяє об'єкт від фону. Методи визначення форми за силуетом вимагають кілька видів сцени, знятих камерами з різних точок спостереження. Такий процес разом із правильною текстуризацією створює повну 3D модель об'єктів у сцені. Метод відтворення форми за силуетом потребує точної калібровки камери.

Процедура 3D реконструкції, що базується на аналізі силуетів називається відтворення форми за силуетом (shape-from-silhouette) описана в [15]. Для кожного зображення силует цільового об'єкту сегментується за допомогою віднімання фону. Отримані силуети проєктуються назад у загальний 3D простір з проєкційними центрами, рівними положенням камер. Зворотне проєктування силуету створює конусоподібний об'єм. Перетин усіх конусів утворює візуальну оболонку цільового 3D об'єкта, який часто зберігається у воксельному вигляді.

На рис. 10 C позначено куб, який є прикладом 3D об'єкта; S позначає двовимірний екран; P_A та P_B – точки спостереження в 3D просторі; D_A – двовимірний багатокутник на екрані, який є силуетом куба; V_A та V_B – конусоподібний об'єм, зворотно спроектований з точок спостереження P_A та P_B .

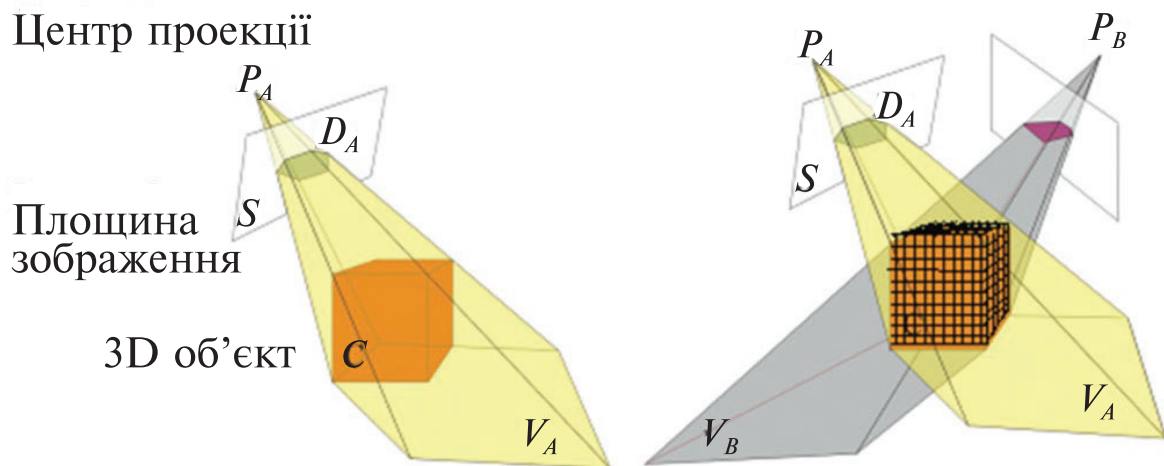


Рис. 10. Перетин об'ємів силуетів [15]

2.2.8. СТРУКТУРА З РУХУ (STRUCTURE FROM MOTION – SFM)

Метод відновлення структури з руху (SfM) використовує серію двовимірних зображень сцени або об'єкта для відновлення їх розрідженого об'єму точок та положень камери (позиція та орієнтація).

SfM базується на тих же принципах, що й стереоскопічна фотограмметрія. У стереофотограмметрії використовується триангуляція для обчислення відносних тривимірних позицій (x, y, z) об'єктів з пар стереозображень. Такі методи вимагають дорожчого спеціалізованого обладнання (стереокамера) та відповідного програмного забезпечення. На відміну від них,

1.2. Аналіз методів 3D реконструкції середовища для доповненої...

стандартні камери добре підходять для методів SfM. Зображення часто знімають під час руху однієї або декількох камери з різних точок спостереження.

Метод відновлення структури з руху часто використовують як початковий крок 3D реконструкції, на якому відбувається генерація розрідженого об'єму точок та положень камери для наступних методів відновлення тривимірної структури сцени. Незважаючи на це, SfM може створювати 3D моделі на основі об'ємів точок з високою роздільною здатністю. Тому, більш детально розглянемо його як окремий повноцінний метод 3D реконструкції сцени.

2.3. КЛАСИЧНІ МЕТОДИ 3D РЕКОНСТРУКЦІЇ СЕРЕДОВИЩА

3D реконструкцію в режимі реального часу можна визначити як процес, при якому відбувається відновлення віртуальної тривимірної моделі сцени, або об'єкта на ній, по зображенням з камери. Окрім даних з камери необхідними є параметри камери та її положення під час зйомки. Зазвичай ця інформація відома, або обчислюється на послідовності зображень.

3D реконструкцію сцени можна зробити, використовуючи або одне зображення, або кілька знімків, зроблених з різних позицій камери. Високу популярність набрала 3D реконструкція на основі множини зображень, де використовуються такі стандартні підходи як стереозір, структура з руху (Structure from Motion – SfM) та підходи на основі даних з довільних точок спостереження (Multi-View Stereo – MVS). Активний розвиток глибинного навчання дозволив проводити реконструкцію середовища навіть по одному зображенню.

2.3.1. ВІДНОВЛЕННЯ СТРУКТУРИ З РУХУ ДЛЯ ЗАДАЧІ 3D РЕКОНСТРУКЦІЇ

Метод відновлення структури з руху (SfM) використовує серію двовимірних зображень сцени або об'єкта для відновлення їх 3D структури на базі знайденого розрідженого об'єму точок. Для створення 3D реконструкції із застосуванням SfM необхідно мати багато зображень області або об'єкта з високим ступенем перекриття, зроблених з різних точок спостереження. SfM проілюстровано на рис. 11. Алгоритм включає три основні етапи:

Співставлення відповідних ознак і вимірювання відстаней між ними на площині зображення камери d та d' . Алгоритм Scale Invariant Feature Transform (SIFT) [16] дозволяє порівнювати відповідні ознаки навіть за великих варіацій у масштабі та куті огляду, а також за умов часткового перекриття і зміни освітлення.

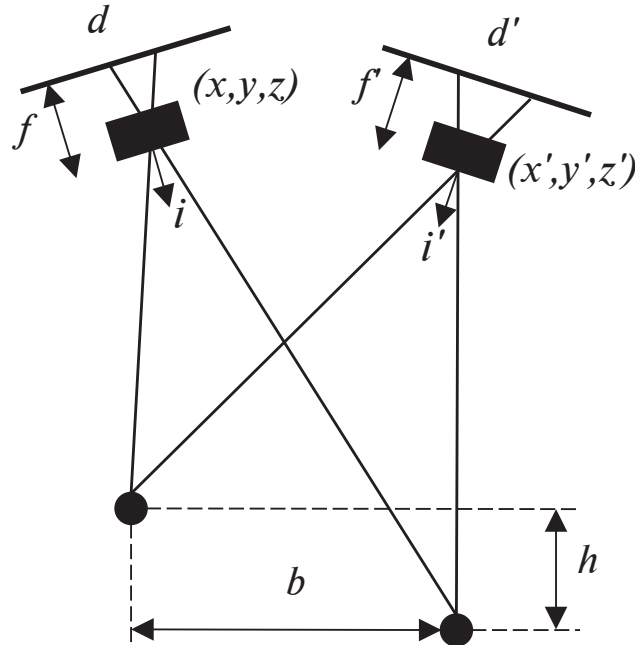


Рис. 11. Ілюстрація методу структура з руху (SfM)

При наявності відповідних місць розташування декількох точок на двох або більше зображеннях, зазвичай існує лише одне математичне рішення для визначення місця, де були зроблені фотографії. Тому можна обчислити позиції камер (x, y, z) , (x', y', z') , орієнтації i та i' , фокусні відстані f та f' , і відносні позиції відповідних ознак b та h в одному кроці, який називається пакетна корекція даних (bundle adjustment). Звідси походить термін «структура з руху». Структура сцени включає всі ці параметри, а рух відноситься до руху камери.

Наступним кроком є визначення щільного об'єму точок і 3D поверхні, використовуючи параметри камери і точки SfM.

SfM зазвичай поділяється на три типи: глобальний, інкрементний та гібридний.

Глобальний SfM [17] представлений на рис. 12. Підхід оптимізує всі положення камери одночасно, використовуючи всі доступні переміщення. У глобальному SfM початкові положення камер оцінюються на основі принципів стереозору та епіпольярної геометрії. Пакетна корекція даних (bundle adjustment) [18]

1.2. Аналіз методів 3D реконструкції середовища для доповненої...

виконується лише один раз, що призводить до поліпшення ефективності системи.

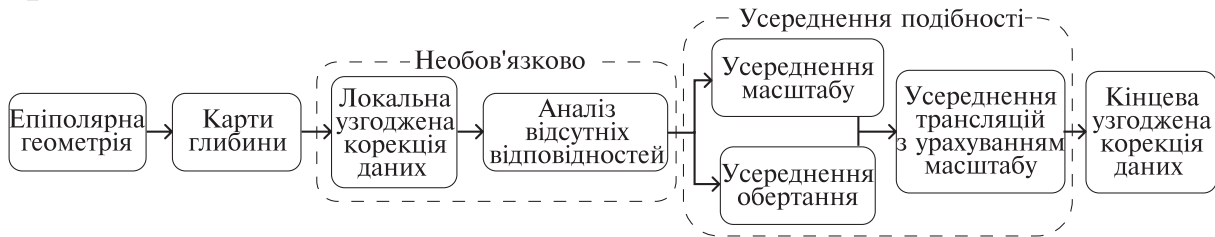


Рис. 12. Приклад алгоритму глобального SfM [17]

Процес глобального SfM включає, в основному, два етапи: оптимізацію положень камери та оптимізацію її орієнтацій. Точність оптимізації орієнтацій камери залежить від точності початкового розрахунку параметрів епіполярного геометричного графу.

Основні компоненти епіполярного геометричного графу:

- Центри камер (C_1 та C_2): Позиції двох камер у просторі.
- Точка в 3D просторі (P): Точка 3D сцени, яку спостерігають обидві камери.
- Точки зображення (p_1 та p_2): Проєкції тривимірної точки P на площини зображень двох камер.
- Базисна лінія: Лінія, що з'єднує центри двох камер.
- Епіполярна площина: Площина, що проходить через тривимірну точку P та центри камер.
- Епіполі (e_1 та e_2): Точки, де базисна лінія перетинає площини зображень.
- Епіполярні лінії: Лінії на площинах зображень, вздовж яких лежать відповідні точки.

Інкрементний SfM [19] здатний обробляти великі обсяги даних. Цей підхід, зазвичай, використовується саме для реконструкції розріджених об'ємів точок з неупорядкованого набору зображень. Метод включає кілька етапів, таких як калібрування камери, відстеження необхідних ознак, визначення положень камери, триангуляція та пакетна корекція даних (bundle adjustment). Алгоритм можна описати наступним чином.

1. Ініціалізація вхідних зображень з урахуванням приблизної фокусної відстані.

2. Використання SIFT (Scale Invariant Feature Transform) дескриптору [16] для пошуку необхідних ключових точок на зображенні.

3. Застосування порівняльного аналізу для визначення схожості знайдених ключових точок.

4. Визначення положення камери для першої пари зображень на основі п'яти точок.

5. Здійснення тріангуляції для знаходження перших 3D точок, що формують об'єми точок у тривимірному просторі та виконання пакетної корекції даних (bundle adjustment).

6. Використання методу прямої лінійної трансформації (Direct Linear Transform – DLT) для визначення положення кожної використаної камери.

7. Здійснення тріангуляції для додавання нових точок до об'ємів точок з подальшою пакетною корекцією даних для уточнення результатів.

8. Повторення кроків 5 та 6 до тих пір, поки не завершиться додавання нових камер. Для уточнення 3D об'єму точок модель оптимізується за допомогою глобальної пакетної корекції даних (global bundle adjustment) [18].

Гібридний SfM використовує глобальний SfM для знаходження параметрів камери та інкрементний SfM для реконструкції розріджених об'ємів точок.

SfM підхід часто застосовують на початковому етапі 3D реконструкції. Отримані положення камери та розріджені об'єми точок подаються на вхід наступних методів, які уточнюють отримані дані для відновлення щільної та деталізованої тривимірної моделі середовища. Один з таких методів – багатовидове стереозіставлення (Multi-View Stereo - MVS).

2.3.2. БАГАТОВИДОВЕ СТЕРЕОСПІВСТАВЛЕННЯ

Алгоритми багатовидового стереоспівставлення (Multi-View Stereo — MVS) [20] приймають на вхід велику кількість зображень з даними про їх відповідні положення камери та повертають щільні тривимірні моделі з покращеною точністю. MVS підхід заснований на патчах (Patch-based Multi-view Stereo – PMVS) [21] і дозволяє ефективно отримувати щільні об'єми точок. Метод включає три етапи: ініціалізацію, розширення та фільтрацію. Основна ідея PMVS полягає у відтворенні набору вирівняних патчів, що повністю покривають поверхню об'єкта. Модель патча є ключовим елементом PMVS. Процес реконструкції моделі за допомогою PMVS продемонстровано на рис. 13. Патч P описується як

1.2. Аналіз методів 3D реконструкції середовища для доповненої...

тривимірний прямокутник, для якого конфігурація визначається центральними координатами, одиничним вектором нормалі, екземпляром зображення та колекцією зображень, на яких P розпізнається.



Рис. 13. Ілюстрація проміжних результатів MVS підходу заснованого на патчах

Ініціалізація патча передбачає наступні етапи:

1. Використання детектор Харріса та різницю гаусіанів (Difference of Gaussian – DoG), щоб виявити ключові точки на послідовності зображень.

2. Застосування техніки епіполярного порівняння для зіставлення ключових точок, що дають початкові відповідності.

3. Формування початкових патчів з початкових відповідностей за допомогою триангуляції.

4. Створення колекції розпізнаваних зображень для патча P за допомогою використання кутового тесту, що порівнює вектор спостереження направлений від камери до патча та вектор нормалі відповідної поверхні.

5. Оновлення центру та нормалі шляхом зменшення оцінки фотометричної невідповідності.

6. Використання фотометричного диспаритету для оцінки ефективності створення патча P .

Розповсюдження патча здійснюється поділом кожного зображення на систематичні сітки розміром $N \times N$ пікселів.

1. Збір всіх сусідніх клітинок зображення з кожного розпізнаного зображення в патчі P .

2. Створення нового патча P для зібраних клітинок зображення.

3. Оптимізація шляхом зменшення значення фотометричного диспаритету.

4. Оптимізація нового розпізнаного набору зображень за допомогою порівняння по глибині.

5. Визначення ефективності розширення нового патча.

Фільтрація патча виконується для оптимізації створених щільних об'ємів точок.

2.3.3. РЕКОНСТРУКЦІЯ ПОВЕРХНІ

Алгоритм відтворення поверхні за допомогою методу Пуассона (Poisson Surface Reconstruction – PSR) [22], використовується для створення цілісної структури. Цей метод розглядає завдання відтворення поверхні як просторову задачу Пуассона [22] і може бути описаний наступними етапами: дискретизація задачі, визначення векторного поля, вирішення рівняння Пуассона та відновлення ізоповерхні. Основні етапи відтворення поверхні за методом Пуассона включають наступні.

1. Використання дерева октантів (octree) для ілюстрації ступеню функції розмірів функцій та дискретизації задачі вирішення структури Пуассона.

2. Для визначення векторного поля використовується функція вузла дерева октантів (octree), яка схожа на градієнтне поле функції індикатора.

3. Для вирішення рівняння Пуассона використовується описане векторне поле так, щоб градієнтне поле функції індикатора було суміжним з векторним полем.

4. Для того, щоб зобразити відповідну ізоповерхню використовується оцінена функція індикатора.

2.3.4. ПОЄДНАННЯ МЕТОДІВ SfM, MVS ТА РЕКОНСТРУКЦІЇ ПОВЕРХНІ

У [23] запропонований підхід 3D реконструкції, який поєднує в собі SfM, MVS та реконструкцію поверхні. Алгоритм умовно зображений на рис. 14. Процедура 3D реконструкції дозволяє отримувати окремі 3D моделі, разом з розрідженим та щільним об'ємом 3D точок, а також грубі структури з колекції зображень.

1.2. Аналіз методів 3D реконструкції середовища для доповненої...

Цей метод можна використовувати в різних сферах та для виконання різноманітних завдань.



Рис. 14. Ілюстрація роботи методу, що поєднує SfM, MVS та реконструкцію поверхні [23]

Інкрементальний підхід SfM використовується для отримання розріджених об'ємів точок з вхідних зображень, при цьому обчислюються різні параметри для кожного зображення. Цей підхід включає етапи калібрування камери, відстеження особливостей, обчислення позицій камери, триангуляції та пакетного коригування даних (bundle adjustment). Для отримання щільних об'ємів точок використовується підхід MVS на основі патчів, що включає ініціалізацію, розширення та фільтрацію патчів. Після отримання щільних об'ємів точок для створення моделі використовується підхід реконструкції поверхні за методом Пуассона, яка включає етапи дискретизації задачі, визначення векторного поля, розв'язання рівняння Пуассона та побудову ізоповерхні.

2.4. СУЧАСНІ МЕТОДИ 3D РЕКОНСТРУКЦІЇ СЕРЕДОВИЩА

У сучасному світі нейронні методи для відновлення карт глибини та 3D реконструкції стали більш популярними порівняно зі стандартними та класичними підходами, такими як Structure from Motion (SfM) та Multi-View Stereo (MVS). Причини цього включають кілька ключових аспектів:

- *Висока точність і деталізація.* Нейронні мережі, зокрема глибокі нейронні мережі, можуть навчатися на великих наборах даних і знаходити складні, нелінійні залежності між зображеннями та їх тривимірними представленнями. Це дозволяє їм досягати більш високої точності та деталізації в

порівнянні з традиційними методами, які часто базуються на простих геометричних припущеннях. Наприклад, на відміну від класичних підходів, сучасні методи дозволяють отримувати якісні карти глибин та 3D реконструкцію навіть в областях з однорідними та періодичними текстурями.

- *Автоматизація і узагальнення.* Глибокі нейронні мережі здатні автоматично витягати особливості та патерни з вхідних даних, що зменшує потребу в ручному налаштуванні параметрів та попередньому обробленні даних, як це часто потрібно у традиційних методах. Це робить їх більш універсальними та здатними до узагальнення на нові сцени та умови зйомки.
- *Зменшення чутливості до шумів.* Нейронні методи мають вбудовану здатність до зменшення впливу шумів та артефактів в даних завдяки використанню навчання на великих і різноманітних наборах даних. Це дозволяє отримувати більш стабільні результати навіть при наявності шумів у вхідних зображеннях.
- *Інтеграція різних джерел даних.* Нейронні мережі можуть інтегрувати інформацію з різних даних, таких як RGB-зображення, глибинні карти, та інші сенсорні дані, для створення більш комплексних і точних 3D моделей.

2.4.1. HIGHRES-MVSNET

R. Weilharter та ін. [24] використали архітектуру глибокого навчання для 3D реконструкції із зображень високої роздільної здатності. Традиційні MVS-техніки для уточнення значення глибини кожного пікселя використовують обчислені ознаки та зв'язки між кількома видами сцени [25]. Використання ознак, виділених нейронною мережею під час навчання, є достойною альтернативою. Архітектура HighRes-MVSNet зосереджується на зниженні вимог до пам'яті, необхідної для використання великої кількості даних, які доступні завдяки сучасним камерами з високою роздільною здатністю зображень. Підхід використовує архітектуру енкодера-декодера, зображену на рис. 15. В енкодері система спочатку виконує 3 згортки у початковому шарі, після чого йде шар об'єднання регіонів даних (pooling) і ще один згортковий шар. Таким чином, розмір обсягу ознак зменшується до однієї восьмої від вхідних даних зображення.

1.2. Аналіз методів 3D реконструкції середовища для доповненої...

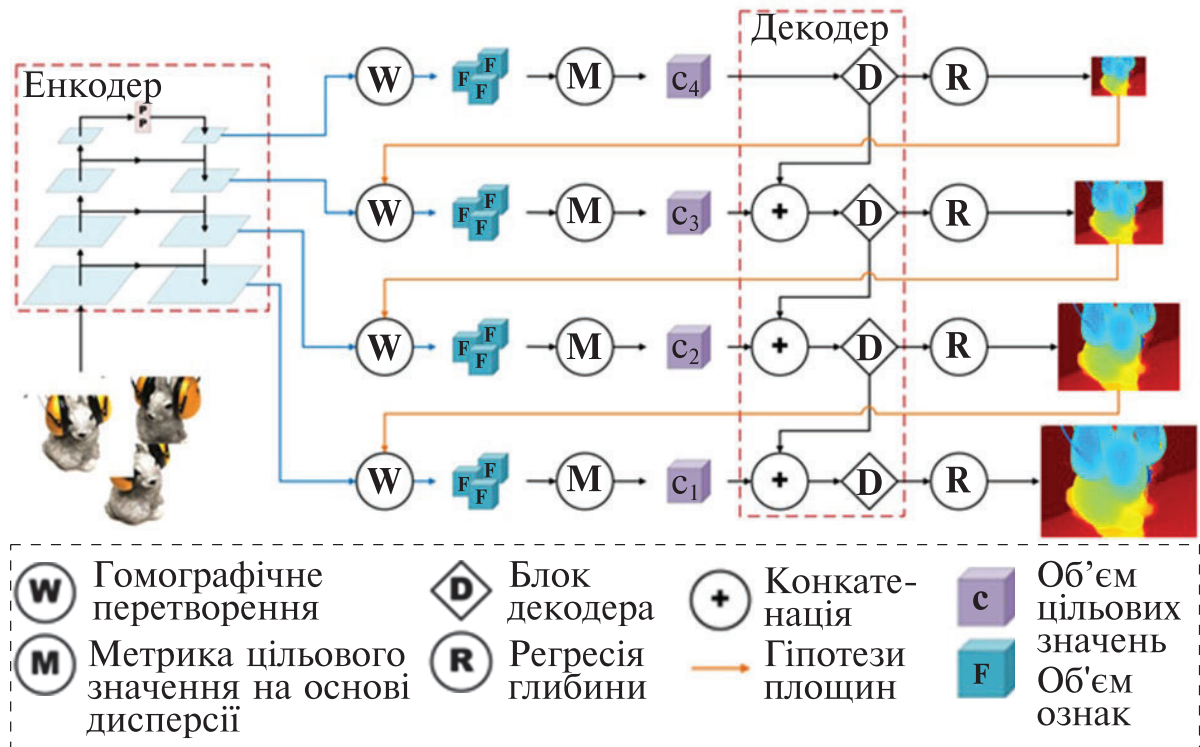


Рис. 15. Архітектура мережі HighRes-MVSNet [24]

Потім використовується архітектура U-Net для вилучення ознак на декількох масштабах, і на кожному масштабі ознаки проєктуються в нижчий вимірний підпростір для контролю розміру вихідної карти ознак. Диференційована функція гомографічного викривлення використовується для побудови 3D об'ємів цільових значень з попередньо вилучених ознак на кожному масштабі. Функція викривлення задається наступним чином:

$$H_i(d) = K_i \cdot R_i \cdot \left(I - \frac{t_0 - t_i \cdot n_0^T}{d} \right) \cdot R_0^T \cdot K_0^T, \quad (4)$$

де $H_i(d)$ є гомографією між i -тою картою ознак і картою ознак на глибині d . K_i , R_i та t_i відповідають параметрам камери, що вказують на поточний кадр, n_0 – головна вісь поточної камери. N об'ємів ознак F_i збираються у об'єм цільових значень C за допомогою метрики втрат на основі варіації:

$$C = \frac{\sum_{i=1}^N (F_i - \bar{F}_i)^2}{N}, \quad (5)$$

де \bar{F}_i – усереднений об'єм ознак.

Об'єми цільових значень декодуються від грубого до детального рівня та об'єднують усі виходи з найдетальнішим

масштабом. Декодерна система має чотири блоки, кожен з яких відповідає за результат на одному з чотирьох етапів. Кожен блок декодера складається з шести 3D-згорткових блоків, кожен з яких містить дві 3D-згортки з залишковим з'єднанням. Декодер надає два виходи: об'єм цільових значень, який об'єднується з вхідними даними наступного етапу, та класифікований об'єм цільових значень, який обробляється шаром softmax і регресією глибини для створення карти глибини, що ініціює об'єми ознак наступного етапу.

Класифікація здійснюється за допомогою шару 3D-згортки, за яким іде шар ReLU та ще один шар 3D-згортки. На найбільш грубому етапі результат створюється лише з сирого об'єму цільових значень. Використовуючи концепцію каскадного об'єму цільових значень, ці об'єми цільових значень формуються в більш точні діапазони глибин залежно від попередньої оцінки. Глибина обчислюється на кожному масштабі шляхом збільшення роздільної здатності класифікованих об'ємів цільових значень до необхідного розміру виходу перед використанням регресії глибини.

2.4.2. 3D-FHNET

Тривимірні ієрархічна мережа об'єднання була представлена в [26]. В її основі лежать ідеї використання як техніки комбінування ознак з кількох видів, так і стратегія ієрархічного прогнозування. Це забезпечує можливість об'єднання реконструкції з однієї та декількох точок спостереження з метою отримання точних результатів. Техніка комбінування ознак спрямована на постійне підвищення якості реконструкції моделі зі збільшенням кількості нових ракурсів. Ієрархічна стратегія прогнозування впроваджена в мережу для точного відтворення дрібних деталей об'єктів. Модель представлена у вигляді воксельного подання, де кожен воксель позначається нулем або одиницею для кожної воксельної сітки. На рис. 16 показана архітектура системи моделі.

Вхідними даними моделі можуть бути як одне, так і кілька зображень, на яких присутній об'єкт представлений з різних точок спостереження. На етапі навчання моделі можна подавати численні зображення одного й того самого об'єкта або сцени. Ці зображення проходять через мережу енкодера, яка вилучає ознаки з поданих даних. Мережа використовується для вилучення ознак

1.2. Аналіз методів 3D реконструкції середовища для доповненої...

із зображень з роздільною здатністю 128×128 пікселів. Енкодер використовує залишковий двовимірний енкодер. Кожне вхідне зображення проходить через шість залишкових згорткових блоків енкодера. Кожен із цих блоків виконує три згортки та одну операцію об'єднання регіонів даних (pooling). Для кожного блоку дані проходять одночасно двома шляхами. Один шлях обробляє дві згортки, а другий шлях — згортку розміром 1×1 . Ці процеси супроводжуються функцією активації ReLU. Ознаки, отримані з обох шляхів, об'єднуються та проходять через шар max-pooling. Нарешті, дані перетворюються у вектор ознак.

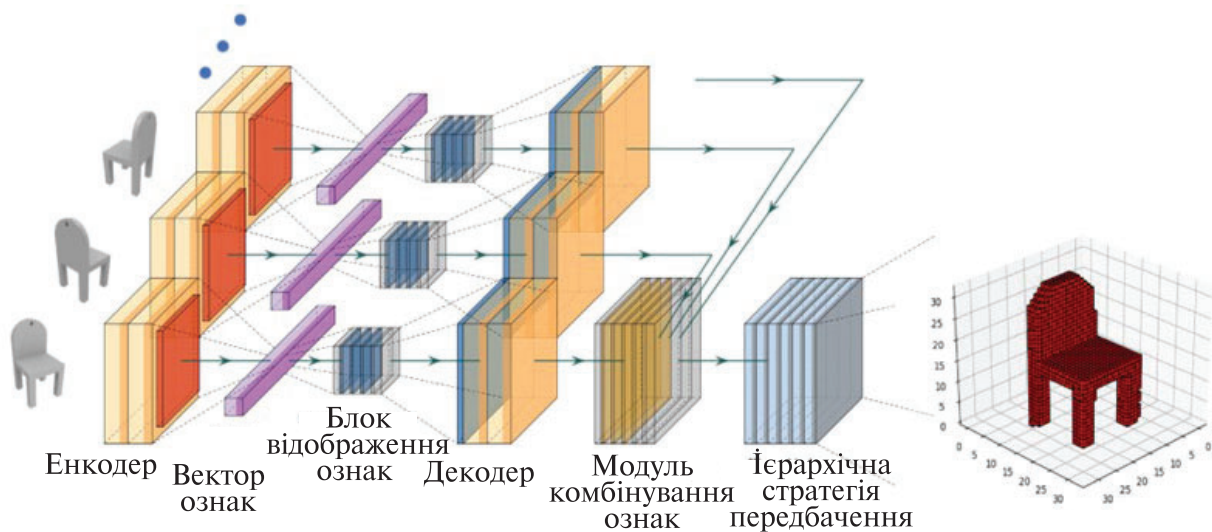


Рис. 16. Архітектура моделі 3D-FHNet [26]

Модуль відображення ознак використовується для перетворення двовимірних ознак у тривимірні, відображаючи вектори ознак, отримані з енкодера. Кожен вектор ознак проходить через повнозв'язний шар, а потім через модуль відображення ознак.

Тривимірна інформація декодується за допомогою залишкового тривимірного декодера. Декодер перетворює тривимірні ознаки в тривимірні об'єми. Він приймає ознаки з модуля відображення ознак та пропускає їх через шість тривимірних залишкових декодерних блоків. Результат потім нормалізується шаром softmax до тривимірного об'єму передбаченої ймовірності. Кожен блок містить три операції зворотної згортки та одну операцію розпакування (unpooling). Дані одночасно проходять двома шляхами. Один із шляхів містить дві тривимірні згортки, а інший шлях — згортку $1 \times 1 \times 1$. Усі ці процеси супроводжуються функцією активації ReLU. Дані надходять до наступного операційного блоку після об'єднання та проходження

через шар розпакування (unpooling). Для об'єднання ознак усіх отриманих зображень використовується модуль інтеграції ознак. Людина може отримати уявлення про об'єкт, рухаючись навколо нього та спостерігаючи його з різних ракурсів. Аналогічно, модель може отримати поточну передбачену ймовірність зайнятості вокселів, розглядаючи зображення. Коли кількість вхідних видів об'єкта збільшується, кількість воксельних сіток також зростає, відповідно точність моделі покращується.

Модель використовує ієрархічну стратегію прогнозування для виведення зайнятості вокселів у виді 0-1. Значення 0 вказує на відсутність зайнятості, а значення 1 вказує на зайнятість відповідного вокселю. Спочатку модель встановлює поріг, і воксельна сітка з передбаченою ймовірністю, рівною або більшою за поріг, буде класифікована як зайнята, а ті, що мають менше значення, як незайняті. Це також допомагає відновити дрібні деталі об'єкта. Таку метрику якості, як точність, можна обчислити, порівнюючи передбачену зайнятість вокселів 0-1 з фактичною зайнятістю вокселів.

2.4.3. ATLAS

Одним з варіантів представлення 3D об'єкта, або площини в комп'ютерній графіці, є усічена знакова функція відстані (Truncated Signed Distance Function - TSDF). Метод 3D-реконструкції сцени, що безпосередньо регресує TSDF із набору RGB-зображень та відповідних положень камери, представлений у [27].

ATLAS приймає на вхід послідовність RGB-зображень довільної довжини, внутрішні параметри камери та положення камери відповідні до кожного кадру. Зображення проходять через основу 2D CNN для вилучення ознак. Потім ці ознаки зворотно проєктуються в 3D-воксельний об'єм і акумулюються та усереднюються з вже існуючими даними. Після того як ознаки зображень об'єдналися в 3D, відбувається безпосередня регресія TSDF за допомогою 3D CNN (рис. 17).

Після акумуляції ознак у воксельному об'ємі використовується тривимірна згорткова мережу енкодера-декодера для уточнення ознак і регресії вихідного TSDF. Кожен шар енкодера і декодера використовує набір залишкових блоків розміром $3 \times 3 \times 3$. Зменшення масштабу реалізується за допомогою згортки $3 \times 3 \times 3$ з кроком 2, а збільшення масштабу використовує

1.2. Аналіз методів 3D реконструкції середовища для доповненої...

трилінійну інтерполяцію з наступним застосуванням згортки $1 \times 1 \times 1$ для зміни розміру ознак. Розмір ознак подвоюється при кожному зменшенні масштабу і зменшується вдвічі при кожному збільшенні масштабу. Усі згорткові шари супроводжуються batchnorm та активацією ReLU.

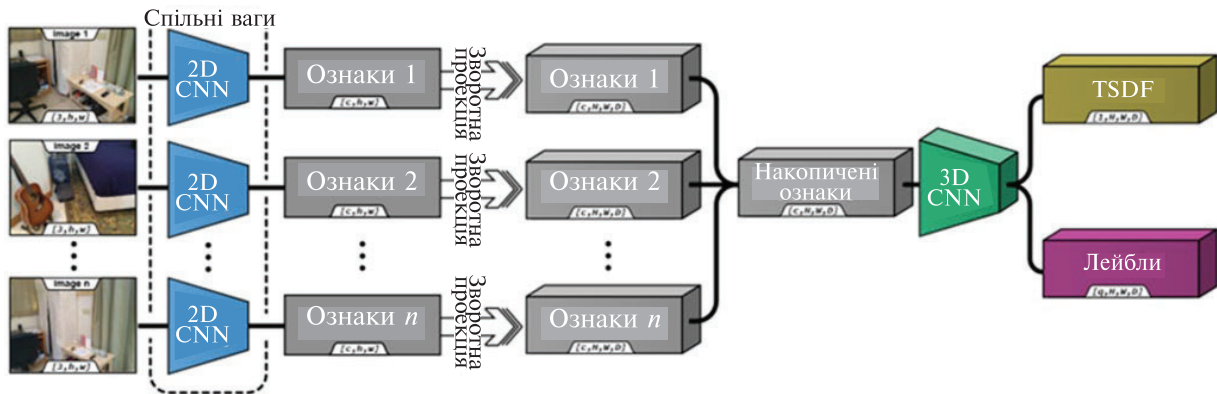


Рис. 17. Архітектура моделі ATLAS [27]

На верхньому шарі енкодера-декодера використовується згортка $1 \times 1 \times 1$, що супроводжується функцією активації \tanh для регресії кінцевих значень TSDF.

Крім того, модель має проміжні виходи на кожній декодованій роздільній здатності перед збільшенням масштабу. Ці додаткові виходи використовуються як для проміжного контролю, щоб допомогти мережі швидше навчатися, так і для покращення якості наступних ітерацій.

2.4.4. SIMPLERECON

На відміну від багатьох сучасних методів 3D реконструкції, що застосовують важкі тривимірні згорткові шари, тим самим обмежуючи своє застосування в умовах обмежених обчислювальних ресурсів, SimpleRecon [28] робить акцент на високоякісному прогнозуванні глибини. Традиційний та доступний підхід об'єднання карт глибини у поєднанні із запропонованим якісним відновленням глибини з кількох ракурсів призводить до точних 3D реконструкцій.

На вхід методу подаються: опорне зображення I_0 , набір вхідних зображень I_n , де $n \in \{1, \dots, N - 1\}$, а також внутрішні параметри камер та відносні позиції камер. Під час навчання використовуються карти глибини D_{gt} , що відповідають вхідним RGB-зображенням. Під час тестування модель виконує

передбачення щільних карт глибини \hat{D} для кожного опорного зображення.

SimpleRecon доповнює архітектуру енкодера-декодера для прогнозування глибини об'ємом цільових значень (рис. 18).

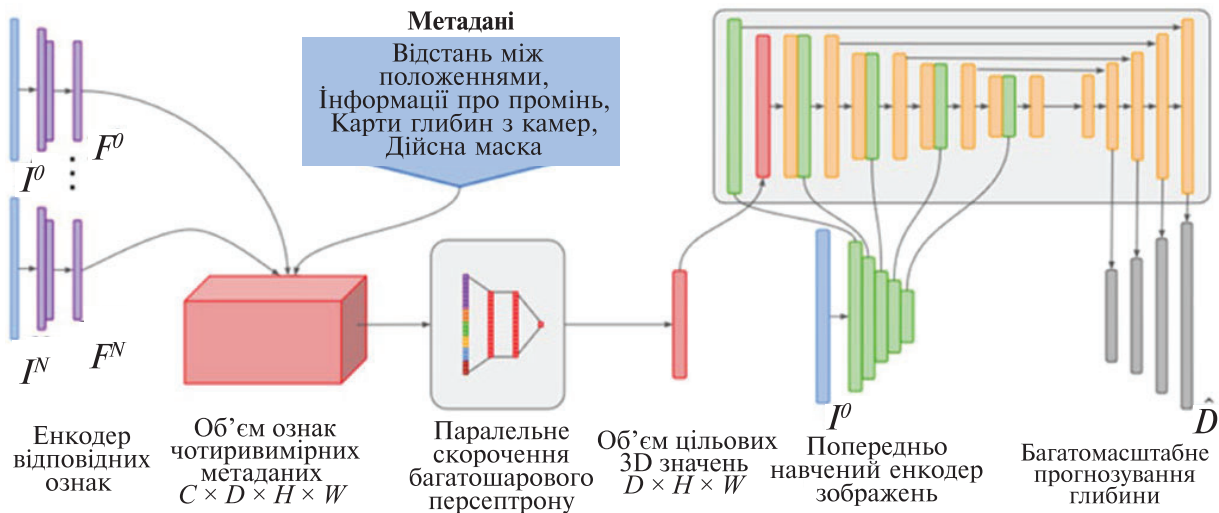


Рис. 18. Архітектура моделі SimpleRecon [28]

Енкодер зображень вилучає відповідні ознаки з опорних зображень, що подаються у об'єм цільових значень. Вихід об'єму цільових значень обробляється за допомогою двовимірної згорткової мережі енкодера-декодера, доповненої ознаками вилученими за допомогою окремого попередньо навченого енкодера зображень. Ключова ідея підходу полягає у включенні доступних метаданих до об'єму цільових значень разом із типовими глибинними ознаками зображення, що дозволяє мережі використовувати корисну інформацію, таку як геометричні дані та відносна позиція камер. Ці метадані дозволяють мережі краще визначати відносну важливість кожного вхідного зображення для оцінки глибини для певного пікселя.

Об'єм цільових значень є чотиривимірним тензором розмірністю $C \times D \times H \times W$, де для кожного просторового положення (k, i, j) існує вектор ознак розмірністю C , де k – індекс площини глибини. Вектор ознак C складається з ознак опорного зображення та набору ознак вхідних зображень, що деформується враховуючи специфічні метадані.

Мережа базується на двовимірній згортковій архітектурі енкодера-декодера. Об'єднання об'єму цільових значень відбувається за рахунок підсумовування результатів добутку точок між опорним зображенням і кожним вхідним зображенням та дає

1.2. Аналіз методів 3D реконструкції середовища для доповненої...

результати, які конкурентоспроможні з сучасними методами оцінки глибини.

В якості енкодера зображень та енкодера відповідності ознак використовується невеликий, але потужніший енкодер EfficientNetv2 S [29]. Для ефективного створення карт відповідності ознак застосовуються перші два блоки ResNet18 [30].

В SimpleRecon об'єднання ознак зображення в енкодер об'єму цільових значень відбувається на зразок DeepVideoMVS [31]. Глибинні ознаки зображення об'єднуються на кількох масштабах, додаючи пропуски з'єднання між енкодером зображення та енкодером об'єму цільових значень на всіх роздільних здатностях.

Функція втрат представлена у вигляді:

$$L = L_{depth} + \alpha_{grad}L_{grad} + \alpha_{normals}L_{normals} + \alpha_{mv}L_{mv}, \quad (6)$$

де: L_{depth} – функція втрат регресії глибини; L_{grad} – функція втрат багатомасштабного градієнту та нормалей; L_{mv} – функція втрат багатовидової регресії глибини; $\alpha_{grad} = \alpha_{normals} = 1.0$ та $\alpha_{mv} = 0.2$ – коефіцієнти підібрані експериментально на валідаційному наборі даних.

2.4.5. MARIGOLD

Підхід Marigold [32], запропонований Bingxin Ke та ін., базується на використанні дифузійної моделі та пропонує протокол тонкого налаштування для монокулярної реконструкції карти глибини. Враховуючи щільну карту глибини, що повертає алгоритм, та її високу точність, підхід також застосовується і для задач 3D реконструкції (рис. 19).

Основний принцип нового підходу полягає у використанні багатих візуальних знань, що зберігаються в сучасних генеративних візуальних моделях. Marigold розроблена на основі попередньо навченої Stable Diffusion моделі та доналаштована з використанням синтетичних даних.

Однією з основних цілей Marigold є ефективність навчання, оскільки дифузійні моделі зазвичай потребують багато ресурсів для навчання. Тому запропонована модель базується на попередньо навченій text-to-image LDM (Stable Diffusion v2 [33]), яка навчилася дуже хорошим апостеріорним ознакам зображень на наборі даних LAION-5B [34].



Вхідне
зображення

Карта глибини

3D реконструкція

Рис. 19. Приклади 3D реконструкції із застосуванням Marigold [32]

Для кодування зображення x та його відповідної карти глибини d в латентний простір, з метою навчання умовного розшумлювача, використовується оригінальний зафіксований варіаційний автоенкодер (VAE) Stable Diffusion. Карта глибини дублюється в три канали, щоб імітувати RGB-зображення, та подається на енкодер. Доналаштується лише U-Net, оптимізуючи стандартну дифузійну задачу відносно латентного коду глибини. Відповідність співставлення карти глибини до зображення досягається шляхом об'єднання двох латентних кодів перед їх подачею в U-Net. Перший шар U-Net модифіковано для прийняття об'єднаних латентних кодів. Підхід доналаштування Marigold представлений на рис. 20. Загальна схема виконання підходу представлена на рис. 21.

Вхідне зображення x кодується за допомогою оригінального варіаційного автоенкодера (VAE) Stable Diffusion у латентний код $z^{(x)}$ і об'єднується з латентним кодом глибини $z_t^{(d)}$ перед подачею в модифіковану доналаштовану U-Net мережу на кожній ітерації розшумлення. Після T ітерацій розшумлення, отриманий

1.2. Аналіз методів 3D реконструкції середовища для доповненої...

латентний код глибини $z_0^{(d)}$ декодується в зображення, три канали якого усереднюються для отримання кінцевої оцінки \hat{d} .

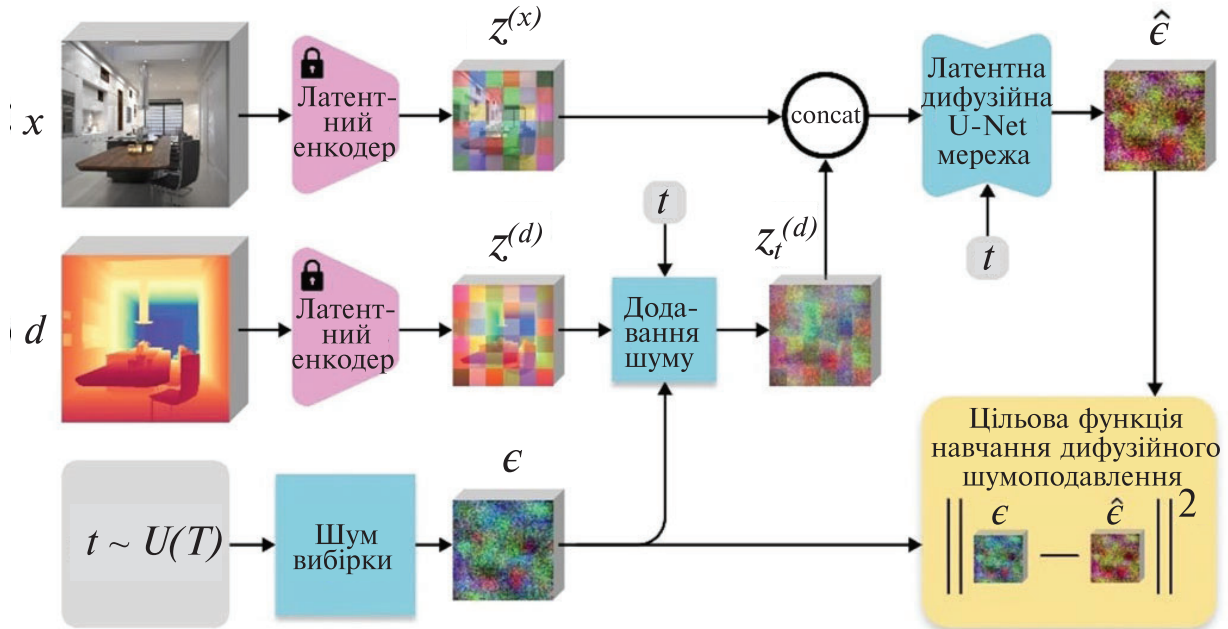


Рис. 20. Огляд підходу доналаштування Marigold [32]

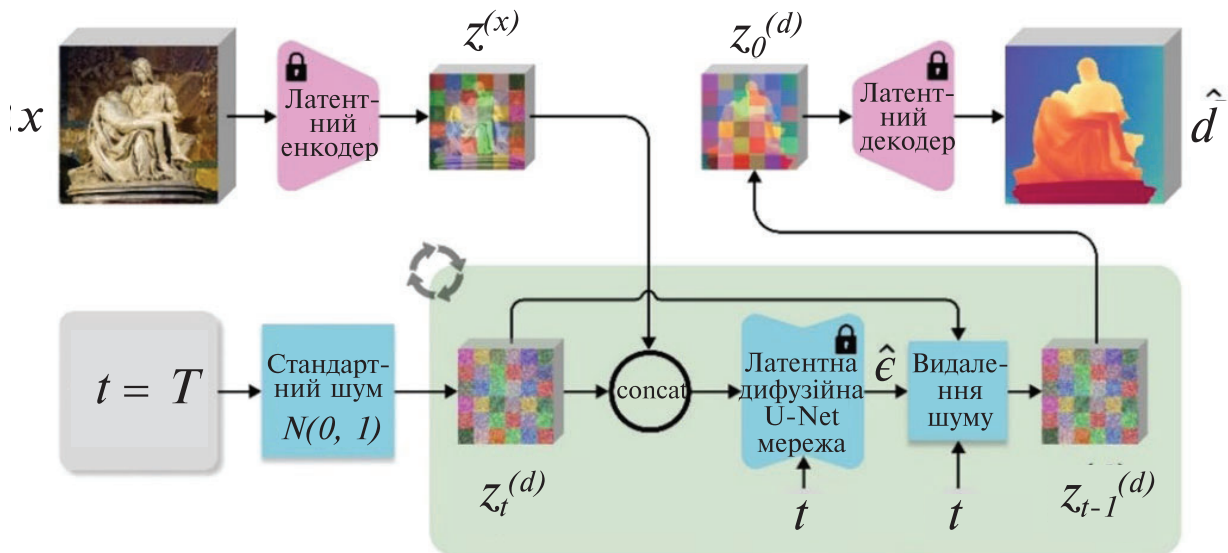


Рис. 21. Загальна схема Marigold для реконструкції карти глибини [32]

Основний недолік Marigold – тривалий час виконання, що робить цей підхід непридатним для систем, які оперують в умовах реального часу.

ВИСНОВКИ

Представлений аналіз дозволяє систематизувати знання та оцінити поточний стан розвитку методів доповненої реальності, 3D реконструкції та оцінки глибини по 2D даним. Розглянуті основні підходи та технології, що використовуються для побудови тривимірних моделей, демонструють стрімкий розвиток даного наукового та інноваційного напрямку. Незважаючи на це залишається низка відкритих питань, які потребують подальших досліджень та удосконалення з метою покращення природності та імерсивності досвіду користувача в доповненій реальності.

По-перше, необхідно розширювати та адаптувати існуючі набори даних до конкретних конфігурацій камер кінцевих пристроїв. Кожний носимий пристрій доповненої реальності має свої специфічні сенсори та їх параметри. Зазвичай наявні набори даних не покривають необхідний кінцевий домен. Для покращення точності та надійності методів 3D реконструкції на кінцевому пристрої необхідно включати набори даних, що відповідають конкретним типам камер та їх характеристикам. Збір та аотація даних — це доволі тривала по часу та витратна процедура. Тому виникає необхідність адоптації існуючих наборів даних під кінцевий домен.

По-друге, для повсякденного використання пристроїв доповненої реальності, важливо зосередити зусилля на покращенні точності реконструкції сцени в складних умовах. До таких сцен можна віднести динамічні сцени з великою кількістю рухомих об'єктів та сцени, в яких присутні відбивні поверхні. Подолання цих викликів сприятиме покращенню якості 3D реконструкції середовища і матиме суттєвий вплив на якість сприйняття користувачем сценаріїв доповненої реальності.

Нарешті, пристрої споживання сценаріїв доповненої реальності, такі як смартфони та носимі окуляри ДР, обмежені з точки зору наявних обчислювальних ресурсів та ємності батареї. Якісні методи 3D реконструкції, зазвичай, мають високу обчислювальну складність і не відповідають вимогам обробки в реальному часі та вимогам по енергоспоживанню. Адаптація, портування, оптимізація методів 3D реконструкції та квантизація відповідних моделей під кінцевий пристрій без суттєвого погіршення їх якості є критичною задачею для забезпечення ефективної роботи сценаріїв ДР, збільшення часу експлуатації носимих пристроїв та їх подальшої мініатюризації.

1.2. Аналіз методів 3D реконструкції середовища для доповненої...

Таким чином, подальші дослідження в напрямку адаптації існуючих наборів даних, покращення якості реконструкції складних сцен та оптимізація рішень для виконання на носимих пристроях є ключовими напрямками для досягнення нових висот у галузі 3D реконструкції середовища та популяризації використання доповненої реальності в повсякденному житті.

ПЕРЕЛІК ПОСИЛАНЬ

1. «Market Research Report: Augmented & Virtual Reality Market by Enterprise (Small, Medium, Large), Technology (AR and VR), Offering (Hardware, Software), Device Type (AR, VR Devices), Application (AR Application, VR Application) and Region - Global Forecast,» 10/2023. URL: <https://www.marketsandmarkets.com/Market-Reports/augmented-reality-virtual-reality-market-1185.html>.

2. Grayson C., «Holographic Waveguides: What You Need To Know To Understand The Smartglasses Market,» 4/5/2017. URL: <https://www.uploadvr.com/waveguides-smartglasses/>.

3. «Samsung's New Transparent MICRO LED Display Blurs the Boundaries Between Content and Reality,» 8/1/2024. URL: <https://news.samsung.com/global/video-ces-2024-samsungs-new-transparent-micro-led-display-blurs-the-boundaries-between-content-and-reality>.

4. Kumar P., «What is a ToF sensor?,» 25/11/2021. URL: <https://www.e-consystems.com/blog/camera/technology/what-is-a-time-of-flight-sensor-what-are-the-key-components-of-a-time-of-flight-camera/>.

5. «What is LiDAR?,» URL: <https://www.synopsys.com/glossary/what-is-lidar.html>.

6. «What is Structured Light Imaging?,» 2018. URL: <https://www.roboticstomorrow.com/article/2018/04/what-is-structured-light-imaging/11821>.

7. Battiato S., S. Curti, M. Cascia & M. Tortora, «Depth map generation by image classification», SPIE Proc. Vol 5302, EI2004 conference 'Threedimensional image capture and applications VI', 2004.

8. F. Cozman & E. Krotkov, «Depth from scattering», IEEE Computer society conference on Computer Vision and Pattern Recognition, Proceedings, Pages: 801–806, 1997.

9. M. W. Tao, P. P. Srinivasan, J. Malik, S. Rusinkiewicz & R. Ramamoorthi, «Depth from shading, defocus, and correspondence

using light-field angular coherence», IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.

10. E. Trucco & A. Verri, Introductory techniques for 3-D computer vision, Prentice Hall, 1998.

11. D. Scharstein, R. Szeliski & R. Zabih, «A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms», International Journal of Computer Vision, № 47, pp. 7-42, 2002.

12. «Stereokinetic Phenomenon from Michael Bach's "Optical Illusions & Visual Phenomena"», URL: <https://michaelbach.de/ot/mot-ske/index.html>.

13. «Parallax scrolling,» Wikipedia, the free encyclopedia, URL: https://en.wikipedia.org/wiki/Parallax_scrolling.

14. A. P. Pentland, «Depth of Scene from Depth of Field», IEEE Transactions on Pattern Analysis and Machine Intelligence, v. 9, № 4, pp. 523-531, 1987.

15. T. Matsuyama, «Exploitation of 3D video technologies», International Conference Informatics Research for Development of Knowledge Society Infrastructure (ICKS), 2004.

16. D. G. Lowe, «Object Recognition from Local Scale-Invariant Features», Proc. of the International Conference on Computer Vision, 1999.

17. Z. Cui & P. Tan, «Global Structure-from-Motion by Similarity Averaging», IEEE International Conference on Computer Vision (ICCV), pp. 864-872, 2015.

18. S. E. Hazzat, A. Saaidi & K. Sat, «Structure from motion for 3D object reconstruction based on local and global bundle adjustment», 2015 Third World Conference on Complex Systems (WCCS), pp. 1-6, 2015.

19. H. Yin & H. Yu, «Incremental SfM 3D reconstruction based on monocular», 13th International Symposium on Computational Intelligence and Design (ISCID), pp. 17-21, 2020.

20. Y. Furukawa & C. Hernández, «Multi-View Stereo: A Tutorial», Foundations and Trends® in Computer Graphics and Vision, v. 9, pp. 1-148, 2015.

21. Y. Furukawa & J. Ponce, «Accurate, Dense, and Robust Multi-View Stereopsis», IEEE Conference on Computer Vision and Pattern Recognition, 2007.

22. J. M. Juszczak, A. Wijata, J. Czajkowska, M. Krecichwost, M. Rudzki, M. Biesok, B. Pyciński, J. Majewski, J. Kostecki & E. Pietka, «Wound 3D Geometrical Feature Estimation Using Poisson Reconstruction», IEEE Access, v. 9, pp. 7894-7909, 2021.

23. Y. Cai, M. Cao, L. Li & X. Liu, «An End-to-End Approach to Reconstructing 3D Model From Image Set», IEEE Access, v. 8, pp. 192268-193284, 2020.

24. R. Weilharter & F. Fraundorfer, «HighRes-MVSNet: A Fast Multi-View Stereo Network for Dense 3D Reconstruction From High-Resolution Images», IEEE Access, v. 9, pp. 11306-11315, 2021.

25. P.-H. Chen, H.-C. Yang, K.-W. Chen & Y.-S. Chen, «MVSNet++: Learning DepthBased Attention Pyramid Features for Multi-View Stereo», IEEE Transactions on Image Processing, v. 29, pp. 7261-7273, 2020.

26. Q. LU, Y. LU, M. XIAO, X. YUAN & W. JIA, «D-FHNet: ThreeDimensional Fusion Hierarchical Reconstruction Method for Any Number of Views», IEEE Access, v. 7, pp. 172902-172912, 2019.

27. Z. Murez, T. As, J. Bartolozzi, A. Sinha, V. Badrinarayanan & A. Rabinovich, Atlas: End-to-End 3D Scene Reconstruction from Posed Images, 2020.

28. M. Sayed, J. Gibson, J. Watson, V. Prisacariu, M. Firman & C. Godard, «SimpleRecon: 3D Reconstruction Without 3D Convolutions», ECCV 2022, 2022.

29. M. Tan & Q. V. Le, «Efficientnetv2: Smaller models and faster training», ICML, 2021.

30. K. He, X. Zhang, S. Ren & J. Sun, «Deep Residual Learning for Image Recognition», IEEE Conference on Computer Vision and Pattern Recognition, 2016.

31. A. Duzceker, S. Galliani, C. Vogel, P. Speciale, M. Dusmanu & M. Pollefeys, «Deepvideomvs: Multi-view stereo on video with recurrent spatio-temporal fusion», в CVPR, 2021.

32. B. Ke, A. Obukhov, S. Huang, N. Metzger, R. C. Daudt & K. Schindler, «Marigold: Repurposing Diffusion-Based Image Generators for Monocular Depth Estimation», The IEEE/CVF Conference on Computer Vision and Pattern Recognition 2024, 2024.

33. R. Rombach, A. Blattmann, D. Lorenz, P. Esser & B. Ommer, «High-resolution image synthesis with latent diffusion models», Computer Vision and Pattern Recognition, 2021.

34. C. Schuhmann, R. Beaumont, R. Vencu, C. Gordon, R. Wightman, M. Cherti, T. Coombes, A. Katta, C. Mullis & M. Wortsman, «LAION-5B: An open large-scale dataset for training next generation image-text models», NeurIPS, 2022.