Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет Кібербезпеки
(повна назва)

Кафедра Інформаційно-мережної інженерії
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА
Пояснювальна записка

рівень вищої освіти перший (бакалаврський)
Дослідження методів відновлення тривимірної структури нерухомої сцени з набору двовимірних зображень
(тема)

Виконав:
здобувач 4 року навчання,
групи ТРІМІ-22-2
Роман Компанієць
(власне ім'я, прізвище)

Спеціальність 172 Телекомунікації та радіотехніка
(код і повна назва спеціальності)
Тип програми освітньо-професійна
Освітня програма Інформаційно-мережна інженерія
(повна назва освітньої програми)

Керівник ст. викл. Олексій Федоров
(посада, власне ім'я, прізвище)

Допускається до захисту

В. о. зав. кафедри ІМІ _____ Дарія Чеботарьова
(підпис) (власне ім'я, прізвище)

2026 р.

Не містить відомостей, заборонених до відкритого публікування

Студент _____ / *Компанієць Р.О.* /
(підпис) (прізвище та ініціали)

Керівник _____ / *Федоров О.В* /
(підпис) (прізвище та ініціали)

Харківський національний університет радіоелектроніки

Факультет Кібербезпеки

Кафедра Інформаційно мережної інжерепії

Рівень вищої освіти перший (бакалаврський)

Спеціальність 172 Телекомунікації та Радіотехніка
(код і повна назва)

Тип програми освітньо-професійна

Освітня програма Інформаційно-мережна інженерія
(повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри _____
(підпис)
«____»_____ 20 ___ р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

здобувачеві Компанійцу Роману Олексійовичу
(прізвище, ім'я, по батькові)

1. Тема роботи Дослідження методів відновлення тривимірної структури нерухомої сцени з набору двовимірних зображень

затверджена наказом університету від 02 червня 2026 р. № 466Ст

2. Термін подання здобувачем роботи до екзаменаційної комісії 17 липня 2026 р.

3. Вихідні дані до роботи Дослідити сучасні методи відновлення тривимірної структури нерухомої сцени за сукупністю двовимірних зображень. В якості вхідних зображень використовувати пари зображень, отримані з різних ракурсів. Окрему увагу приділити пошуку відповідностей на двох зображеннях і застосуванню цих відповідностей для відтворення тривимірної структури сцени. Розглянути математичну модель камери–обскури та особливості проективної геометрії.

4. Перелік питань, що потрібно опрацювати в роботі
Вступ
1 Поняття про комп'ютерний зір та зорову систему людини
2 Основи проективної геометрії
3 Відтворення структурованих сцен на основі двох зображень
Висновки

5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій (п.5 включається до завдання за рішенням випускової кафедри) _____
____слайди презентації в форматі MS Power Point_____
_____
_____
_____
_____
_____
_____
_____
_____

6. Консультанти розділів роботи (п.6 включається до завдання за наявності консультантів згідно з наказом, зазначеним у п.1)

Найменування розділу	Консультант (посада, прізвище, ім'я, по батькові)	Позначка консультанта про виконання розділу	
		підпис	дата

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Строк / терміни виконання етапів роботи	Примітка
1.	Ознайомлення із завданням. Уточнення ТЗ.	08.06.2026	вик.
2.	Підбір літератури за темою роботи	08.06 – 03.07.2026	вик.
3.	Виконання розділу 1	10.06 – 25.06.2026	вик.
4.	Виконання розділу 2	20.06 – 03.07.2026	вик.
5.	Виконання розділу 3	16.06 – 05.07.2026	вик.
6.	Оформлення презентаційного матеріалу,	05.07 – 16.07.2026	вик.
	підготовка до захисту в ЕК		

Дата видачі завдання _08_ ___червня____2026р.

Здобувач _____
(підпис)

Керівник роботи _____ __ст. викл. Олексій Федоров_
(підпис) (посада, власне ім'я, прізвище)

РЕФЕРАТ

Пояснювальна записка: 52 с., 14 рис., 16 джерел, 1 додаток.

Об’єктом дослідження є методи відновлення тривимірної структури нерухомої сцени за парами зображень, отриманих з різних ракурсів.

Мета кваліфікаційної роботи – вивчення методів проективної геометрії та застосування цих методів для відтворення тривимірної структури сцени за набором двовимірних зображень.

В кваліфікаційній роботі досліджується методи відновлення структури та руху, тобто процес вилучення тривимірної структури сцени, а також руху камери шляхом аналізу послідовності зображень. Зокрема проаналізовано базові принципи комп’ютерного зору, такі як геометрія зображень створених фотокамерою, її калібрування, геометрія з одним та двома ракурсами.

3-ВИМІРНА РЕКОНСТРУКЦІЯ, ПРОЕКТИВНА ГЕОМЕТРІЯ, КАМЕРА, СЦЕНА, РУХ, СТЕРЕО ПАРА, ЗІР

ABSTRACT

Explanatory note: 52 p., 14 fig., 16 sources, 1 app.

Object of study is methods for reconstructing the 3D structure of a still scene using pairs of images taken from different positions.

Purpose of the qualification work is to study methods of projective geometry and to apply these methods to reconstruct the 3D structure of a scene from a set of two-dimensional images.

This thesis investigates methods for reconstructing structure from motion, that is, the process of extracting the 3D structure of a scene as well as camera motion by analyzing a sequence of images. In particular, the basic principles of computer vision, such as the geometry of images captured by a camera, camera calibration, as well as single- and dual-view geometry were analyzed.

3D RECONSTRUCTION, PROJECTIVE GEOMETRY, CAMERA, SCENE, MOTION, STEREO PAIR, VISION

ЗМІСТ

ПЕРЕЛІК СКОРОЧЕНЬ, УМОВНИХ ПОЗНАК,
СИМВОЛІВ, ОДИНИЦЬ І ТЕРМІНІВ 7
ВСТУП 8
1 ПОНЯТТЯ ПРО КОМП'ЮТЕРНИЙ ЗІР ТА ЗОРОВУ
СИСТЕМУ ЛЮДИНИ 9
1.1 Уявлення про комп'ютерний зір 9
1.2 Особливості зорової система людини 11
2 ОСНОВИ ПРОЕКТИВНОЇ ГЕОМЕТРІЇ 22
2.1 Математичні моделі камер 22
2.2 Геометрія зображення з одним ракурсом 27
2.2.1 Дія проекційної матриці на прямі 27
2.2.2 Дія проекційної матриці на площинах 28
2.2.3 Дія проекційної матриці на конічні перерізи 29
3 ВІДТВОРЕННЯ СТРУКТУРОВАНИХ СЦЕН
НА ОСНОВІ ДВОХ ЗОБРАЖЕНЬ 31
3.1 Геометрія зображення з двох ракурсів 31
3.1.1 Епіполярна геометрія та фундаментальна матриця 31
3.1.2 Тривимірна реконструкція 34
3.2 Реконструкція структурованих сцен з двох зображень 36
3.2.1 Стратегія виявлення площин 37
3.2.1.1 Груба оцінка гомографії площини 37
3.2.1.2 Зіставлення ліній на основі гомографії 38
3.2.1.3 Удосконалення гомографії 40
ВИСНОВКИ 43
ПЕРЕЛІК ПОСИЛАНЬ 45
ДОДАТОК А СЛАЙДИ ПРЕЗЕНТАЦІЇ 47

ПЕРЕЛІК СКОРОЧЕНЬ, УМОВНИХ ПОЗНАК, СИМВОЛІВ, ОДИНИЦЬ І ТЕРМІНІВ

CCD (charge-coupled device) - Прилад із зарядовим зв'язком.

SfM (structure from motion) - Відтворення форми (структури) на основі руху.

Гороптер - ділянка простору перед спостерігачем, усі точки якої не бачаться подвоєними, оскільки ці точки утворюють зображення у парних ідентичних місцях сітківки.

Епіполярна геометрія — це внутрішня геометрія між двома ракурсами.

Епіполярна площина – це площина, що проходить через два центри камер.

Епіполь – це точка перетину базової лінії та площини зображення.

Епіполярна лінія — це перетин епіполярної площини з площиною зображення.

Фовеа - це невелика центральна ямка в сітківці ока, розташована в центрі жовтої плями.

ВСТУП

В останні десятиліття спостерігається стрімке зростання наукових досліджень у багатьох галузях інформатики, зокрема в галузі комп’ютерного зору. Це зумовлено як природним інтересом дослідників, так і потребами промисловості та суспільства у якісно нових можливостях, які можуть надати комп’ютери. Однією з особливо бажаних можливостей є автоматична реконструкція та аналіз навколишнього тривимірного середовища, а також розпізнавання об’єктів у цьому просторі. Ефективні методи та реалізації 3D-комп'ютерного зору відкриють нові можливості, такі як автоматична навігація роботів і транспортних засобів, спостереження та моніторинг обстановки (що дозволяє автоматично розпізнавати несподівану поведінку людей або інших об'єктів, наприклад, автомобілів у повсякденному дорожньому русі), медичне міркування, дистанційна хірургія та багато-багато іншого.

Відновлення структури та руху — це процес вилучення тривимірної структури сцени, а також руху камери шляхом аналізу послідовності зображень. Це важлива тема в галузі комп’ютерного зору, і за останні три десятиліття було досягнуто значного прогресу як у теорії, так і на практиці.

Кваліфікаційна робота присвячена аналізу базових принципів комп’ютерного зору, таких як геометрія зображення камери, калібрування, геометрія з одним та двома ракурсами, 3D-реконструкція.

1 ПОНЯТТЯ ПРО КОМП'ЮТЕРНИЙ ЗІР ТА ЗОРОВУ СИСТЕМУ ЛЮДИНИ

1.1 Уявлення про комп'ютерний зір

Комп'ютерний зір — це галузь інформатики, метою якої є надання комп'ютерам можливості обробляти та ідентифікувати зображення й відео так само, як це робить людський зір. Комп'ютерний зір прагне імітувати людську зорову систему. Завдання полягає в тому, щоб створити штучні системи, здатні видобувати інформацію із зображень, тобто зробити так, щоб комп'ютери «розуміли» зображення та відео. Даними зображень можуть бути послідовності відео, карти глибини, види з декількох камер або багатовимірні дані з датчиків зображення. Основна мета комп'ютерного зору — описати сцену реального світу на одному або кількох зображеннях, а також ідентифікувати та реконструювати її властивості, такі як колірні характеристики, інформація про форму, текстурні особливості, освітленість сцени тощо [1].

Давайте розглянемо схожість між людською зоровою системою та системою комп'ютерного зору. Як проілюстровано на рис. 1.1, основний принцип роботи цих двох систем майже однаковий: перетворення світла на корисні сигнали/інформацію, на основі яких будуються точні моделі фізичного світу. Подібним чином, на високому рівні структури людського та комп'ютерного зору дещо схожі: обидві мають світлові датчики, які перетворюють фотони на сигнал (зображення), етап обробки та, нарешті, механізм інтерпретації сигналу (розпізнавання об'єктів) [1].

Відмінність між комп'ютерним зором, обробкою зображень та комп'ютерною графікою можна підсумувати так [1]:

1. У комп'ютерному зорі (аналіз зображень, інтерпретація зображень, розуміння сцени) входом є зображення, а виходом — інтерпретація сцени. Аналіз зображень спрямований на проведення кількісних вимірювань зображення для надання його опису.

2. В обробці зображень (відновлення, реконструкція, фільтрація, стиснення, візуалізація) входом є зображення, і виходом також є зображення.

3. Нарешті, у комп'ютерній графіці входом є будь-яка сцена реального світу, а виходом — зображення.

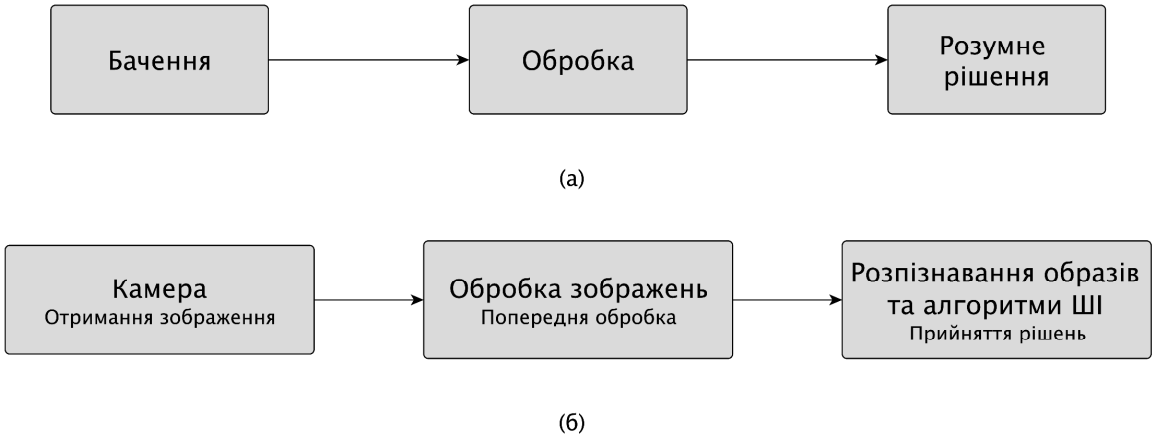


Рисунок 1.1 - Зір людини (а) та комп'ютерний зір (б)

Комп'ютерний зір створює модель із зображень (аналіз), тоді як комп'ютерна графіка бере модель як вхідні дані та перетворює її на зображення (синтез). З точки зору теореми Баєса цю концепцію можна пояснити так [1]:

$$P(\text{World} \mid \text{Image}) = \frac{P(\text{Image} \mid \text{World}) \times P(\text{World})}{P(\text{Image})}$$

де, $P(\text{World} \mid \text{Image})$ - це комп'ютерний зір, $P(\text{World})$ означає моделювання об'єктів у світі, а $P(\text{Image} \mid \text{World})$ - комп'ютерна графіка. Цей аспект веде до статистичних методів навчання. Отже, комп'ютерний зір займається створенням машин, які можуть бачити та взаємодіяти зі світом. Ця концепція наочно продемонстрована на рис. 1.2 [1].

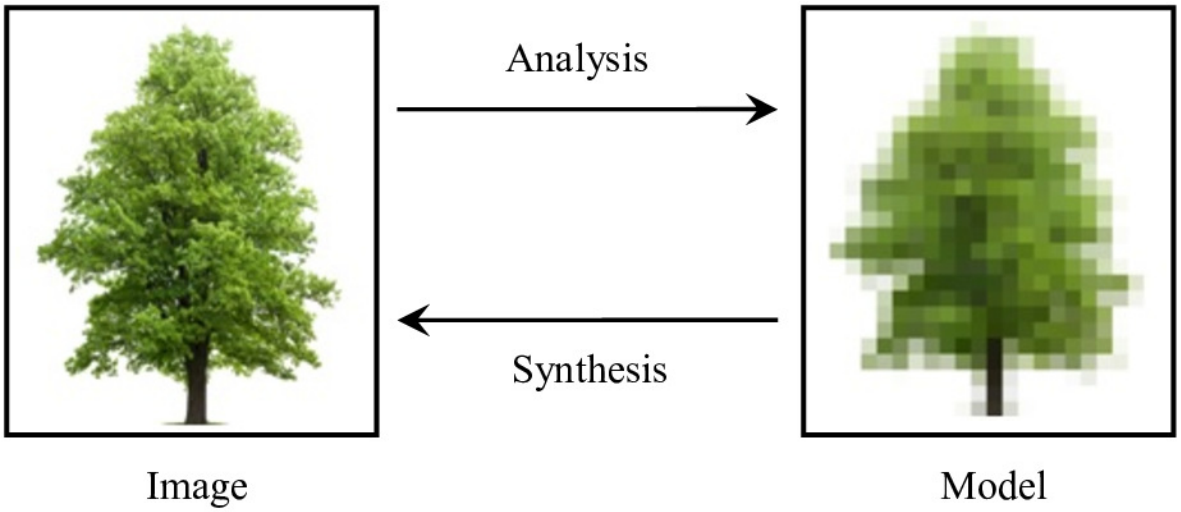


Рисунок 1.2 - Зв'язок комп’ютерного зору з комп’ютерною графікою

Комп’ютерний зір зараз використовується в багатьох реальних завданнях, які включають машинну інспекцію, оптичне розпізнавання символів (OCR), побудову 3D-моделей (фотограмметрія), аналіз медичних зображень, автоматичне відеоспостереження, біометрію, злиття та склеювання зображень, 3D-моделювання тощо. Як показано на рис. 1.3, комп’ютерний зір пов’язаний із багатьма важливими дослідницькими напрямами [1].

Після того, як ми отримали загальне уявлення про комп’ютерний зір, розглянемо особливості будови ока людини і бінокулярного зору загалом.

1.2 Особливості зорової система людини

Мільйони років еволюції сформували людську зорову систему (ЛЗС), а в її межах — найвишуканіший, недосяжний і таємничий механізм стереоскопічного сприйняття глибини на планеті Земля. Процес зору починається в оці, схема якого зображена на рисунку. 1.4 [2].

Падаюче світло спочатку проходить крізь зіницю, яка регулює кількість світла, що надходить на кришталик ока. Розмір апертури зіниці контролюється м’язами сфінктера райдужної оболонки (радинки). Чим більшою стає ця апертура, тим більшою є сферична аберація і меншою — глибина різкості ока.

Зорова вісь з’єднує точку фіксації та центральну ямку (фовеа). Хоча око не є обертально-симетричним, приблизну оптичну вісь можна визначити як лінію, що з’єднує центр кривизни рогівки та центр кришталика. Кут між цими двома осями становить близько 5°. Слід зазначити, що саме око не є окремим органом, а 150-міліметровим виростом мозку. У контексті комп’ютерного зору найважливішою частиною ока є сітківка — місце обміну, яке перетворює вхідний потік фотонів на відповідні нервові збудження [2].

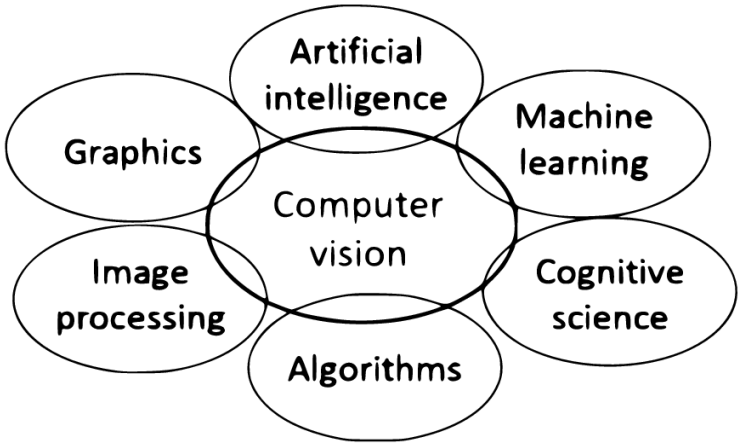


Рисунок 1.3 - Комп’ютерний зір та суміжні дисципліни

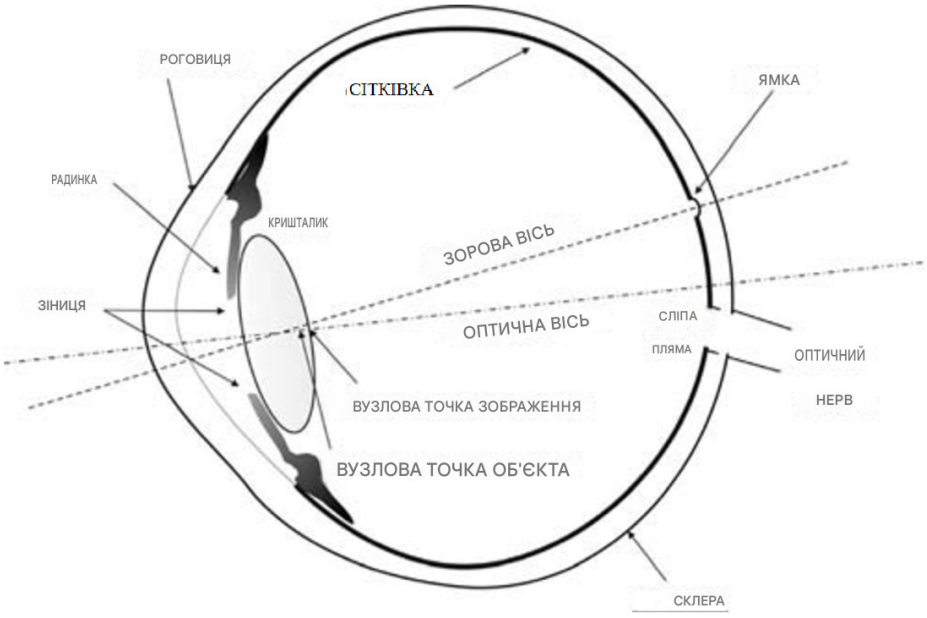


Рисунок 1.4 – Будова людського ока

У контексті бінокулярного зору та стереоскопічного сприйняття глибини важливо, щоб очі синхронізувались так, щоб одна й та сама область сцени проектувалася на відповідні фовеа кожного ока. Рисунок 1.5 представляє модель бінокулярного зору: зображення певної точки H створюється в обох очах, точно в центрах їхніх фовеа [2].

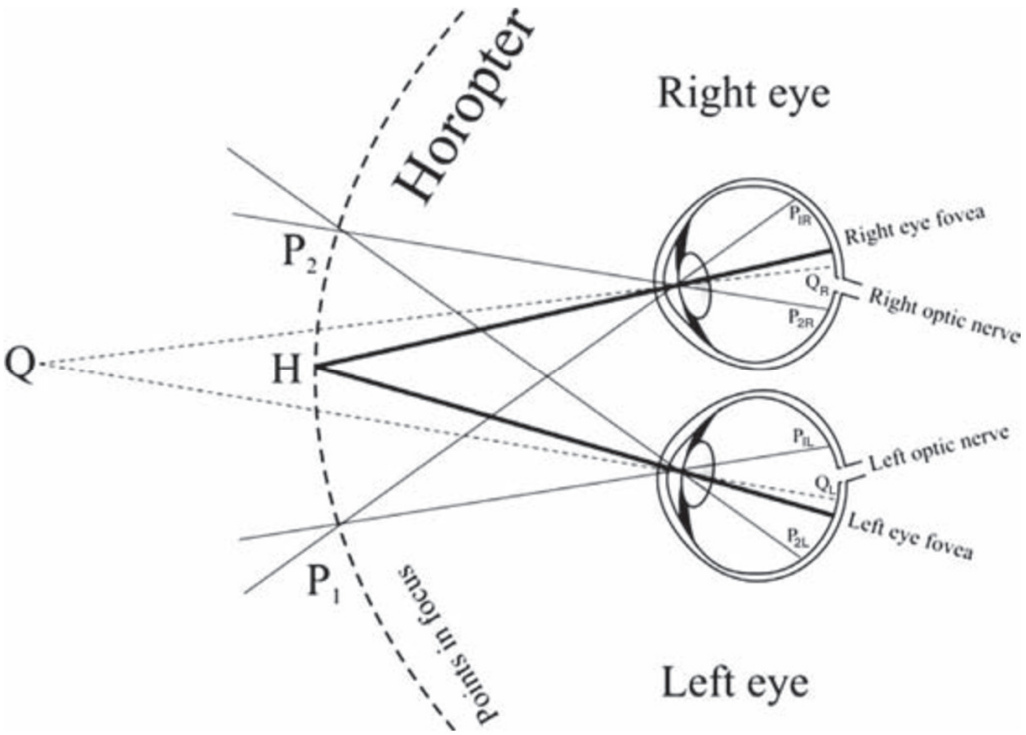


Рисунок 1.5 - Диспаритет на сітківці ока. Гороптер позначено пунктирною лінією. H — точка фіксації

На кожній сітківці також створюються зображення навколишніх 3D-точок. Ми позначаємо відстань цих зображень відносно відповідної фовеа. За цього припущення дві точки зображення на кожній із сітківок є відповідними, коли їхні відстані до відповідних фовеа однакові. На рисунку 1.5 ця умова виконується для точок P_1 і P_2, але не для Q. Тобто відстані P_{1R} і P_{1L} однакові. Це також справедливо для P_{2R} і P_{2L}, але не для Q_R і Q_L, які спрямовані в протилежні боки від фовеа. Однак ця остання властивість дозволяє ЛЗС зробити висновок, що Q знаходиться далі від гороптера. Гороптер позначає

ділянку простору перед спостерігачем, усі точки якої не бачаться подвоєними, оскільки ці точки утворюють зображення у парних ідентичних місцях сітківки, тобто у відповідних точках. Проводячи тепер зворотні міркування, тобто шукаючи 3D-точки такі, що їхні зображення на сітківці знаходяться на однаковій відстані від обох фовеа, ми знаходимо 3D-область, відому як гороптер. Зображення на сітківці всіх точок, крім тих, що належать гороптеру, називаються невідповідними. Відносна різниця у відстані від фовеа для кожної з цих невідповідних точок називається ретинальною диспаратністю (розбіжністю) [7, 13]. Тепер очевидно, що точки гороптера мають нульову ретинальну диспаратність. Ретинальна диспаратність використовується ЛЗС для оцінки відстані до 3D-об'єктів у світі [2].

Сигнали, індуковані на фовеа, передаються на вхід первинної зорової кори мозку, яку нейроанатоми позначають як зорову зону 1 (V1). Ця зона зорової кори є першим місцем у всій структурі, де окремі нейрони отримують бінокулярний вхід. Також було виявлено, що деякі нейрони у V1 реагують виключно на взаємне збудження від обох очей. Ці нейрони, звані детекторами диспаратності, чутливі до стереоскопічних стимулів [2, 13].

Крім того, взаємозв'язок між частотою спрацьовування цих нейронів-детекторів диспаратності, вимірюваною в імпульсах на секунду, та вхідною ретинальною диспаратністю називається функцією налаштування диспаратності. Вона має виражений максимум для нульової ретинальної диспаратності (тобто вона «налаштована» найкраще реагувати на нульову диспаратність), тобто для 3D-точок, що лежать на гороптері [2, 7].

Було проведено багато експериментів для кращого розуміння стереоскопічних процесів у ЛЗС. Явищем, вперше поміченим під час таких досліджень, був вплив зміни яскравості на процес зіставлення відповідних візуальних стимулів від кожного ока, тобто виявлення диспаратності. У найпростішому випадку це стосується виявлення, тобто кореляції відповідних країв зображення на кожній сітківці, тоді як кореляція відповідних текстурованих областей є складнішою. У 1979 році Марр і Поджо [10]

висунули теорію, що стереоскопічне зіставлення базується на кореляції місць на сітківці, де друга похідна сигналу яскравості перетинає нульове значення; це так звані нульові переходи [2].

Нульові переходи відповідають тим областям зображення, які демонструють найбільшу зміну сигналу, а не найбільше абсолютне значення самого сигналу. Подальші дослідження, проведені Мейх'ю та Фрізбі [10], показали, що стереоскопічна кореляція в ЛЗС залежить не лише від нульових переходів, а й від більш узагальненого механізму зіставлення, застосованого до спектральних компонентів двовимірного сигналу яскравості. Маллот та ін. [9] виявили можливість залучення вторинного механізму кореляції, коли сигнал яскравості змінюється дуже повільно. На основі цих результатів можна стверджувати, що ЛЗС воліє корелювати більш загальні ознаки, якщо вони доступні на зображенні. Це стосується кореляції на основі нульових переходів, а також кореляції на основі максимумів значень сигналу. Однак кореляція на основі зіставлення спектральних компонентів сигналу яскравості домінує, коли вони є найбільш виразними ознаками, знайденими на зображеннях. Коли немає ні значних нульових переходів, ні інших відмінностей сигналу, ЛЗС здатна оцінювати значення диспаратності на основі кореляції максимальних значень низькочастотних компонентів сигналу яскравості [2].

Якісно новий розвиток подій був повідомлений Юлешем [8] у 1960 році, коли він продемонстрував так званий стереограм з випадкових точок. Стереограм з випадкових точок складається зі стереопари зображень, у якій перше зображення пари створюється шляхом генерації поля випадкових точок. Друге зображення стереопари створюється шляхом копіювання першого зображення з подальшим вибором і зміщенням на невелику відстань конкретної області в межах копії. Рисунок 1.6 окреслює кроки цієї конструкції. При спостереженні двома очима стереограм з випадкових точок дозволяє сприймати глибину, як це видно на рисунку 1.6 у вигляді прямокутника, ближчого до спостерігача. Подальші дослідження на цю тему показали, що стереоскопічний ефект досягається, навіть якщо одне з випадкових зображень спотворене,

наприклад, додаванням випадкових точок або низькочастотною фільтрацією. З іншого боку, зміна полярності яскравості (тобто світлі та темні області змінюються місцями на одному зображенні стереограми) призводить до втрати стереоефекту [2].

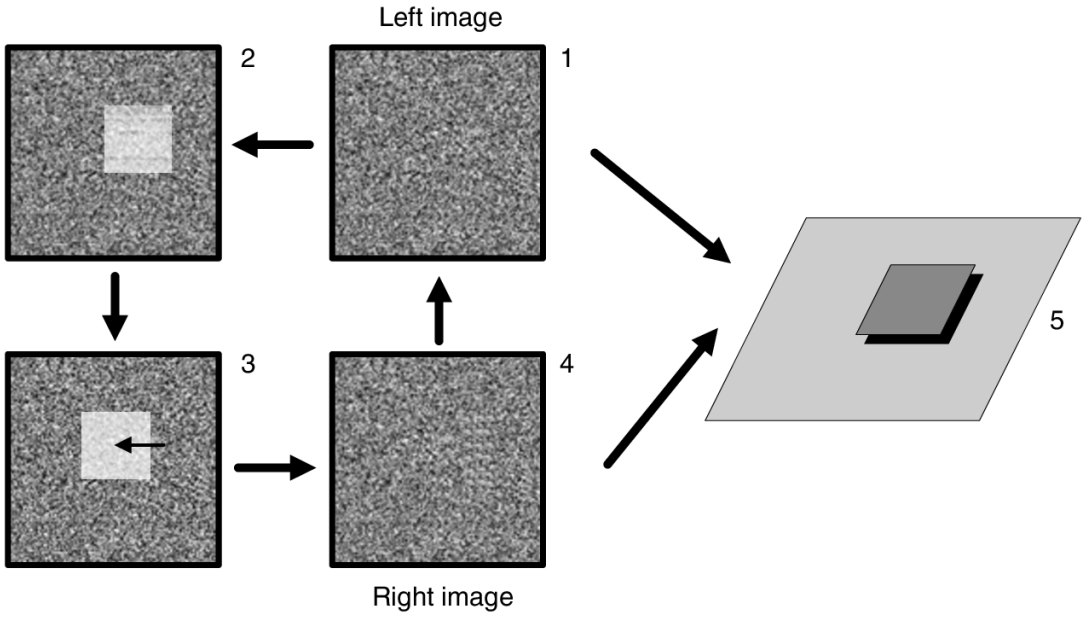


Рисунок 1.6 - Створення стереограми з випадковим розташуванням точок: (1) ліве зображення; (2) виділена ділянка на лівому зображенні; (3) зсув цієї ділянки; (4) праве зображення; (5) ефект глибини при спогляданні обома очима

Дослідження сприйняття глибини, що базується виключно на сприйнятті кольорів, показали, що інформація про колір також впливає на цей процес лише обмеженою мірою [2, 7].

Було виявлено, що процес стереокореляції також залежить від інших факторів, що призвело до теорії, яка передбачає, що ті порівнювані місця, які збігаються за розміром, формою, кольором і рухом, мають більше переваг під час стереозіставлення. Це також пояснювало б, чому ЛЗС потрібно більше часу для зіставлення стереограм з випадкових точок, які не мають таких ознак. Цю теорію також можна інтерпретувати в галузі обчислювальних методів стереозіставлення: якщо певний локальний оператор може зібрати достатньо інформації в заданій околиці пікселів, наприклад локальну частоту, орієнтацію

або фазу, то подальше зіставлення можна виконати надійніше і, можливо, швидше, ніж це було б можливо за відсутності такої інформації. Це досить евристичне правило можна підтвердити експериментом [2].

Ще одним відомим обмеженням стереозіставлення, прийнятим ЛЗС, є так зване «зіставлення найбільш пов'язаних зображень». Воно передбачає, що за наявності вибору зображення або підзображення вважається «зіставленим», якщо воно дає найбільшу кількість значущих збігів. В іншому випадку перевага надається зображенню, яке містить найбільшу кількість проекцій просторових точок. Завдяки цій стратегії ЛЗС надає перевагу тим зображенням або їхнім підзображенням, які є потенційно найбільш цікавими для спостерігача, оскільки вони знаходяться найближче до нього [2].

Ще одне обмеження, відкрите Юлешем, — це межа градієнта диспаратності [8]. Інші обмеження базуються на досвіді, отриманому з щоденних спостережень за навколишнім простором. Одне з них полягає в тому, що повсякденне середовище зазвичай помірно «щільне», оскільки ми повинні якось у ньому пересуватися. Подібне спостереження вказує на те, що навколишні об'єкти не є прозорими. З цих спостережень ми можемо вивести інші обмеження зіставлення, що базуються на: неперервності поверхні, фігурній гладкості, впорядкуванні точок зіставлення та унікальності точок зіставлення. Їхня функція та вплив на ЛЗС, хоча і вказані багатьма експериментами, ще не була повністю пояснена [2].

Інше явище, що відіграє важливу роль як у людському, так і в машинному стереозорі — це оклюзії, які пояснюються на рисунку 1.7.

Те, як часткові оклюзії спостережуваних об'єктів впливають на їхнє бінокулярне сприйняття, досліджував Леонардо да Вінчі [2]. Нещодавня робота Андерсона вказує на те, що явище оклюзії має значний вплив на процес стереоскопічного сприйняття [3]. Область, видима виключно для лівого ока, називається лівою видимою областю. Аналогічно для правого ока ми отримуємо праву видиму область. На рисунку 1.7 ці області позначені світло-сірим кольором. Область, доступна для спостереження обома очима одночасно,

може сприйматися у повному стереозорі. Натомість темна область ліворуч від об’єкта на рисунку 1.7 представляє область, повністю закриту для обох очей. Далеко за об’єктом знову є область, видима обома очима, тому об’єкт не перекриває весь простір за собою, а лише його частину. Також відомо і легко перевірити, що напівзакриті області, видимі правим оком, знаходяться близько до правого краю об’єкта, що перекриває. Аналогічно, напівзакриті області, видимі лівим оком, знаходяться поблизу лівого краю такого об’єкта, що перекриває. Ця ситуація зображена на рисунку 1.7 [2].

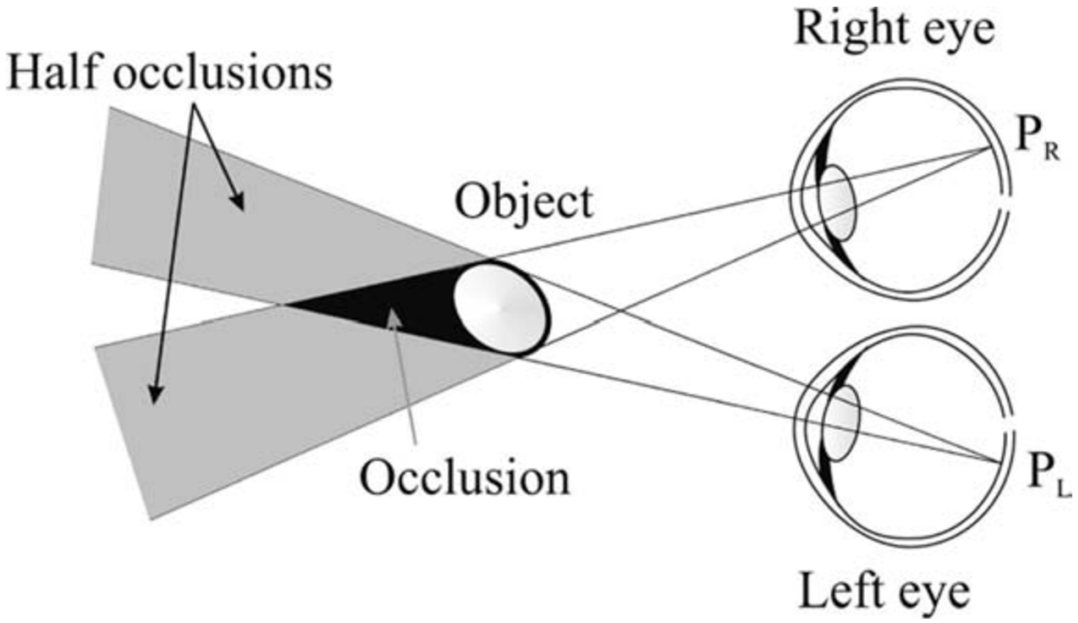


Рисунок 1.7 - Явище затінення (оклюзії). Часткові затінення мають світліший відтінок. Темна ділянка залишається повністю затіненою об’єктом

Ефект часткових оклюзій неминуче пов’язаний з розривом у гладкості (неперервності) сприйманої поверхні в глибині. Таким чином, завдяки наявності часткових оклюзій можна розрізнити розриви глибини від поверхонь, що поступово змінюються, які, своєю чергою, обмежені максимально допустимим градієнтом диспаратності. Ці та інші факти показують, що ЛЗС активно розкладає вертикальні та горизонтальні паралакси зображення на диспаратності та напівоклюзії [3]. Вони утворюють два взаємодоповнюючі

джерела візуальної інформації. Ретинальні диспаратності надають інформацію про відносну глибину спостережуваних поверхонь, видимих обома очима одночасно. З іншого боку, часткові оклюзії, які видимі кожним оком окремо, дають достатньо даних для сегментації спостережуваної сцени на цілісні об'єкти на межах об'єктів [2].

Цікаво згадати, що градієнт вертикальної диспаратності також можна використовувати для визначення відстані до спостережуваних об'єктів, як показали Мейх'ю та Лонге-Хіггінс [11] і обговорювали також Бреннер та ін. [4]. Однак останні психофізичні експерименти показали, що така інформація не використовується ЛЗС. Дійсно, вертикальні відмінності зображення не завжди є вертикальними паралаксами. Іноді вони спричинені напівоклюзіями. Ґрунтуючись на цих спостереженнях та психофізичних експериментах, Андерсен [3] припускає, що міжзорові відмінності у вертикальному положенні можуть впливати на стереоскопічну глибину, що сприймається ЛЗС, сигналізуючи про наявність контурів, що перекривають [2].

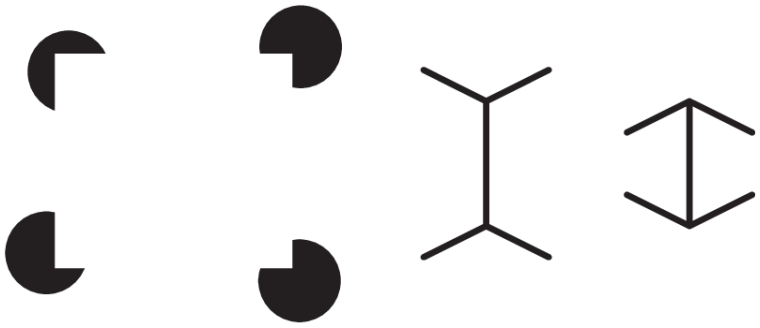
Сприйняття глибини за допомогою ЛЗС викликається не лише суто стереоскопічними механізмами, воно також підтримується явищами рухів голови та очей, а також паралаксом руху [2].

Багато психофізичних експериментів призводять до спостереження, що між різними зоровими сигналами, які впливають на ЛЗС, існує постійне суперництво. Тоді ЛЗС виявляє такі об'єкти, які виникають з максимумів щільності хороших збігів, коли вони одночасно узгоджуються з повсякденним досвідом [2].

Інформація про глибину, отримана ЛЗС, як і інші візуальні сигнали, такі як інформація про колір, краї, тіні та оклюзії, є лише інгредієнтами, зібраними мозком для формування висновків про світ. Як ці візуальні висновки потім інтегруються та інтерпретуються в єдине сприйняття — досі невідомо, хоча дослідниками було запропоновано гіпотези та моделі. Знання про функціонування зорової системи було отримано опосередковано за допомогою спостережень за двома різними наборами явищ, відомих з медицини та

психофізіології. Перша група явищ описана в тематичних дослідженнях, що фіксують захворювання зору та їх подальше лікування. Клінічно спостерігалося, що пацієнти, які були незрячими від народження, а потім отримали можливість сприймати візуальні відчуття, мали труднощі з навчанням сприйняття об’єктів та інтерпретацією сцен, хоча вони легко могли розрізняти базові ознаки [7, 13]. Опосередковано це дає нам певне розуміння концептуальних етапів та складності механізмів зору нашого мозку [2].

Візуальні ілюзії складають другу групу явищ, які допомагають нам зрозуміти, як зорові шляхи перетворюють зображення на сітківці у сприйняття об’єктів. Існує багато ілюзій, які «обманюють» нашу зорову систему, надаючи візуальні підказки, що не узгоджуються з фізикою 3D-світу, засвоєною нами з повсякденного досвіду [5, 6, 12]. Два прості приклади ілюзій, очевидно пов’язані зі сприйняттям глибини людиною, представлені на рисунку 1.8. Перший приклад (рисунок 1.8(a)) ілюструє роль оклюзії (перекриття об'єктів) у візуальному сприйнятті. Наш набутий досвід щодо прозорості об’єктів змушує нас сприймати ілюзорну фігуру, існування якої лише натякається (тобто стає очевидним) завдяки присутності контурів оклюзії, накладених на інші видимі об’єкти на зображенні [2].



(a) (b)

Рисунок 1.8 - Візуальні ілюзії: (a) штучний прямокутник чітко видно, хоча він не намальований безпосередньо; (б) дві вертикальні лінії мають однакову довжину (що можна перевірити за допомогою лінійки), хоча ліва з них сприймається як довша

Другий приклад (рис. 1.8(b)) показує дві лінії однакової довжини, які закінчуються стрілками на кожному кінці. Проте пари стрілок на відповідних кінцях ліній вказують у протилежних напрямках. Тим не менш, перша лінія створює враження довшої. Це явище можна пояснити повсякденним досвідом. Ліва конфігурація на рисунку 1.8(b) припускає, що центральна лінія знаходиться далі від спостерігача порівняно з правою конфігурацією. Це змушує нас вірити, що в 3D-світі ліва лінія має бути довшою [2].

Спільним спостереженням щодо таких ілюзій у 2D-зображеннях є те, що ми відчуваємо хибну інтерпретацію «плоских» візерунків, оскільки наша зорова система завжди намагається інтерпретувати дані зображення так, ніби це види реальних 3D-об’єктів [13].

Іншими словами, евристики, які ми розвинули для візуального сприйняття, ґрунтуються на припущенні, що ми спостерігаємо сцени, вбудовані в 3D-простір. Розуміння цих евристик може забезпечити потенційні засоби, за допомогою яких ми зможемо створювати алгоритми бінокулярного відновлення глибини, що працюють так само надійно, як механізми сприйняття глибини у зоровій системі людини [2].

2 ОСНОВИ ПРОЕКТИВНОЇ ГЕОМЕТРІЇ

У цьому розділі представлено основні концепції проективної геометрії та математична модель камери з точковим отвором (камери-обскури).

Ми знайомі з евклідовою геометрією, оскільки вона широко застосовується для опису навколишнього світу. Однак цього недостатньо, коли ми розглядаємо відображення з простору на зображення, оскільки деякі евклідові властивості не зберігаються під час процесу. Проективна геометрія вводиться для моделювання процесу візуалізації, що дозволяє використовувати більш загальний клас перетворень, ніж просто переміщення та обертання в евклідовій геометрії. Більш детальну інформація про концепції і властивості проективної геометрії, міститься, наприклад, в [15, 16]. У цьому розділі ми коротко розглянемо моделі камер та геометрію візуалізації.

2.1 Математичні моделі камер

У проективній геометрії точку зазвичай позначають в однорідній формі. Припустимо, $\vec{\mathbf{x}} = [u, v]^T$ що — двовимірна точка зображення, а $\vec{\mathbf{X}} = [x, y, z]^T$ — тривимірна точка простору, тоді відповідні однорідні форми виражаються як $\mathbf{x} = [\vec{\mathbf{x}}, 1]^T$ та $\mathbf{X} = [\vec{\mathbf{X}}, 1]^T$ [14].

Камери в комп'ютерному зорі базуються на моделі камери-обскури, яка проектує об'єкти в 3D-світі на об'єкти в 2D-зображенні. Коли використовуються однорідні координати, відображення з простору на зображення можна виразити лінійно наступним чином, незалежно від вибору систем координат [14]

$$\lambda \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \underbrace{\begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{44} \end{bmatrix}}_{\mathbf{P}} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}, \tag{2.1}$$

де λ — масштабний коефіцієнт, який зазвичай називають проективною глибиною; $\mathbf{P}$ — матриця 3 × 4, відома як матриця перспективної проекції або матриця камери.

Процес отримання зображень за допомогою камери-обскури показано на рис. 2.1. У найпростішому випадку, коли оптичний центр камери розміщений у початку світової (world) системи координат, а фокусна відстань дорівнює 1, матриця камери $\mathbf{P}$ спрощується до [14]

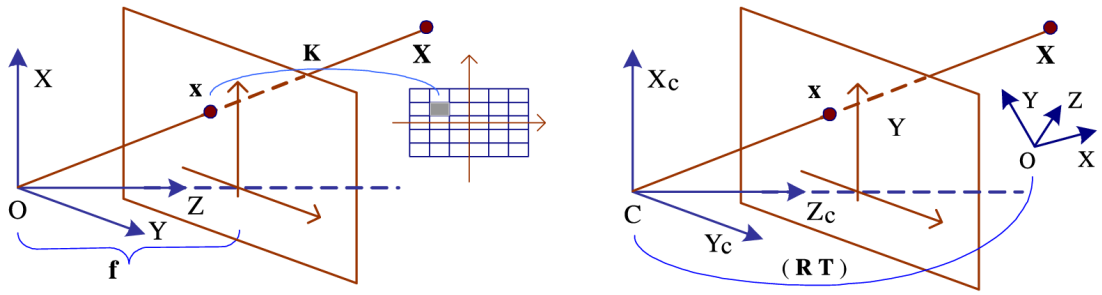
$$\mathbf{P}_0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} = [\mathbf{I}_3 \mid 0], \tag{2.2}$$

де $\mathbf{I}_3$ позначає одиничну матрицю 3 × 3. Матриця $\mathbf{P}_0$ називається канонічною формою матриці камери.

У загальному випадку, коли світова система відрізняється від системи камери, зв'язок між двома системами можна визначити за допомогою 3D-обертання $\mathbf{R}$ з подальшим перенесенням $\mathbf{T}$, що еквівалентно множенню матриці $\mathbf{P}_0$ праворуч на матрицю перетворення 4 × 4 [14]

$$\mathbf{A} = \begin{bmatrix} \mathbf{R} & \mathbf{T} \\ \mathbf{0}^T & 1 \end{bmatrix}, \quad \mathbf{R} = \begin{bmatrix} \mathbf{r}_1^T \\ \mathbf{r}_2^T \\ \mathbf{r}_3^T \end{bmatrix}, \quad \mathbf{T} = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}, \tag{2.3}$$

де $\mathbf{R}$ — матриця обертання 3 × 3, яка має 3 ступені свободи, а $\mathbf{R} \cdot \mathbf{R}^T = \mathbf{I}_3$; $\mathbf{r}_i^T$ позначає i-й рядок матриці обертання, $\mathbf{T}$ — це 3D-вектор переміщення. Матриця $\mathbf{A}$ описує положення камери, яке відоме як зовнішні параметри камери. Вона має 6 незалежних параметрів, 3 для обертання та інші 3 для переміщення.



(a) (b)

Рисунок 2.1 - Геометрія перспективної камери: (а) окремий випадок, коли оптичний центр камери збігається з початком координат світової системи; (б) загальний випадок з довільними світовою системою координат та системою координат камери

Для реальної камери координати зображення не відповідають фізичним координатам у площині зображення. Зв'язок між двома системами можна змоделювати за допомогою верхньої трикутної матриці 3 × 3, **K** яка залежить від власних параметрів камери, таких як розмір і форма пікселів і положення пристрою формування зображення. Зміна координат у площині зображення еквівалентна множенню матриці $\mathbf{P}_0$ ліворуч на матрицю **K** [14]

$$\mathbf{K} = \begin{bmatrix} f/s_u & (f/s_u)\cot\theta & u_0 \\ 0 & f/s_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} f_u & \varsigma & u_0 \\ 0 & f_v & v_0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad (2,4)$$

де f — фокусна відстань камери; s_u та s_v — ширина та висота піксельної області; $[u_0, v_0]^T$ — координати головної (principal) точки камери, тобто точки перетину оптичної осі та площини зображення (сенсору); θ — кут між двома осями зображення.

Матриця **K** відома як матриця калібрування камери. Ми визначаємо $\varsigma = (f/s_u)\cot\theta$ як параметр нахилу камери, а $\kappa = s_u/s_v$ як співвідношення

сторін камери. Для деяких точних промислових CCD-камер можна сміливо вважати одиничним співвідношення сторін та осі зображення ортогональними, тобто $\kappa \to 1,\ \theta \to \pi/2$. Таким чином, ми маємо $f_u = f_v$ та $\varsigma = 0$. Головна точка зазвичай знаходиться в центрі зображення, тобто $u_0 = v_0 = 0$. Тоді камера спрощується до одного внутрішнього параметра [14].

З урахуванням перетворення координат (2.3) та (2.4), загальна матриця перспективної проекції стає [14]

$$\mathbf{P} = \mathbf{K}\,\mathbf{P}_0\,\mathbf{D} = \mathbf{K}\,[\mathbf{R} \mid \mathbf{T}], \tag{2.5}$$

а процес зображення перспективної проекції можна записати як

$$\lambda\,\mathbf{x} = \mathbf{P}\,\mathbf{X} = \mathbf{K} \cdot [\mathbf{R} \mid \mathbf{T}] \cdot \mathbf{X}. \tag{2.6}$$

Загальна проекційна матриця має 12 параметрів, визначених до параметра довільного масштабу λ, таким чином вона має лише 11 ступенів свободи (6 для зовнішнього параметра $\mathbf{R}$ та $\mathbf{T}$, а решта 5 - для внутрішніх параметрів $\mathbf{K}$). З (2.6) ми бачимо, що кожна відповідність простору точці зображення забезпечує два обмеження на $\mathbf{P}$. Таким чином, проекційну матрицю можна відновити з мінімуму $5\frac{1}{2}$ таких відповідностей у загальному положенні [14].

Зазвичай ми позначаємо матрицю проекції камери у такому вигляді [14]

$$\mathbf{P} = [\mathbf{M}, \mathbf{p}_4] = [\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, \mathbf{p}_4] = [\mathbf{p}_1^T, \mathbf{p}_2^T, \mathbf{p}_3^T, \mathbf{p}_4^T]^T, \tag{2.7}$$

де $\mathbf{p}_i^T$ позначає i-й рядок $\mathbf{P}$, а матриця $\mathbf{M}$ складається з перших трьох стовпців $\mathbf{P}$. Матриця камери має такі геометричні властивості та значення.

Твердження 2.1. Проекційна матриця має ранг 3 та має одновимірний правий нуль-простір $\mathbf{C}$, який задовольняє $\mathbf{PC} = 0$. Можна перевірити, що $\mathbf{C}$ є

однорідним представленням центру камери. Якщо матриця $\mathbf{M}$ не є сингулярною (тобто існує обернена матриця $\mathbf{M}^{-1}$),

$$\mathbf{C} = \begin{bmatrix} -\mathbf{M}^{-1}\mathbf{p}_4 \\ 1 \end{bmatrix}.$$

В іншому випадку, якщо $\mathbf{M}$ є сингулярною, а $\mathbf{d}$ є правим нульовим вектором $\mathbf{M}$, маємо

$$\mathbf{C} = \begin{bmatrix} \mathbf{d} \\ 0 \end{bmatrix}.$$

що означає, що центр камери розташований у нескінченності. Таку ситуацію ми називаємо «камера в нескінченності» [14].

Твердження 2.2. Головна площина - це площина, що проходить через центр камери і є паралельною площині зображення; вона збігається з останнім рядком $\mathbf{p}^{(3)}$ матриці камери. Головна вісь - це лінія, що проходить через центр камери та перпендикулярна до головної площини. Вектор головної осі $v = \det(\mathbf{M}) \cdot \mathbf{m}^{(3)}$ спрямований у напрямку головного вектора до передньої частини камери. Перетин головної осі та площини зображення називається головною точкою, яку можна обчислити з $\mathbf{x}_0 = \mathbf{M}\,\mathbf{m}^{(3)}$, де $\mathbf{m}^{(3)}$ позначає третій рядок $\mathbf{M}$ [14].

Твердження 2.3. Перші три стовпці проекційної матриці відповідають точкам сходу в напрямках X, Y та Z світової системи координат відповідно. Останній стовпець матриці $\mathbf{P}$ є відображенням початку координат світової системи [14].

Доведення. У будь-якій світовій системі координат канонічні форми напрямків вздовж осей X, Y та Z можна записати як $\mathbf{X}_w = [1, 0, 0, 0]^T$,

$\mathbf{Y}_w = [0, 1, 0, 0]^T$ та $\mathbf{Z}_w = [0, 0, 1, 0]^T$ відповідно, а початок координат дорівнює $\mathbf{O}_w = [0, 0, 0, 1]^T$. Тоді з перспективної проекції (2.6) маємо

$$\begin{cases} \mathbf{v}_x \sim \mathbf{P}\mathbf{X}_w = \mathbf{p}_1, \\ \mathbf{v}_y \sim \mathbf{P}\mathbf{Y}_w = \mathbf{p}_2, \\ \mathbf{v}_z \sim \mathbf{P}\mathbf{Z}_w = \mathbf{p}_3, \\ \mathbf{v}_o \sim \mathbf{P}\mathbf{O}_w = \mathbf{p}_4, \end{cases} \tag{2.8}$$

де ~ позначає рівність з точністю до параметра масштабу [14].

2.2 Геометрія зображення з одним ракурсом

Рівняння (2.6) описує відображення точки з тривимірного простору в простір зображення під час перспективної проекції. І навпаки, для точки зображення $\mathbf{x}$ її зворотна проекція утворює промінь $\mathbf{l} = \{\mathbf{X}|\mathbf{P}\mathbf{X} \sim \mathbf{x}\}$, складений з набору точок простору, які проектуються на $\mathbf{x}$ матрицею камери $\mathbf{P}$. У подальшому тексті ми представимо деякі результати щодо прямої та зворотної проекції інших геометричних об’єктів, таких як прямі, площини та конічні перерізи [14].

2.2.1 Дія проекційної матриці на прямі

Лінію $\mathbf{L}$ в просторі можна визначити двома точками на цій прямій, як показано на рис. 2.2. Таким чином, [14]

$$\mathbf{L} = \{\mathbf{X}(\mu) = \mathbf{X}_1 + \mu\mathbf{X}\}, \tag{2.9}$$

де $\mathbf{X}_1$ та $\mathbf{X}_2$ – дві просторові точки на $\mathbf{L}$. Оптичний центр та $\mathbf{L}$ визначають площину, яка перетинає площину зображення по прямій $\mathbf{l}$. Згідно з матрицею проекції $\mathbf{P}$ маємо [14]

$$\mathbf{l} = \{\mathbf{P}\,\mathbf{X}(\mu) = \mathbf{P}\,\mathbf{X}_1 + \mu\,\mathbf{P}\,\mathbf{X}_2\} = \{\mathbf{x}(\mu) = \mathbf{x}_1 + \mu\,\mathbf{x}_2\}. \qquad (2.10)$$

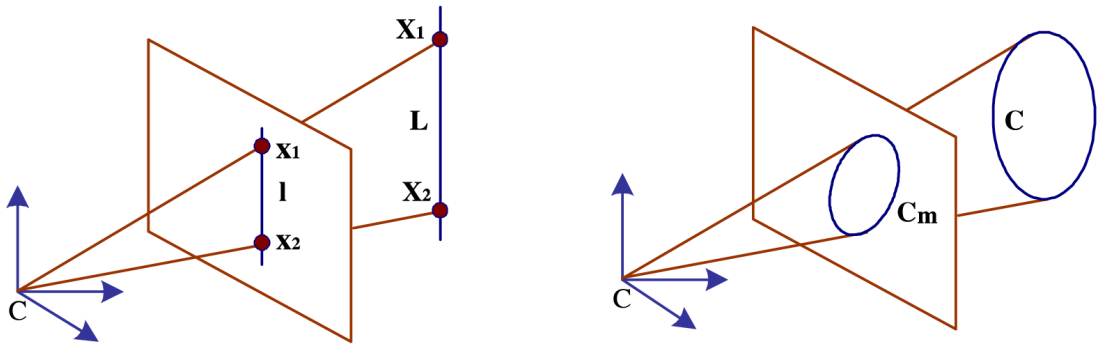


Рисунок 2.2 - Дія матриці проекції на прямі та конічні перерізи: (а) пряма проекція та зворотна проекція прямої; (б) пряма проекція та зворотна проекція конічного перерізу

Відображення прямої $\mathbf{L}$ - це пряма, що сполучає відображення $\mathbf{X}_1$ та $\mathbf{X}_2$. Зворотна проекція лінії зображення — це просторова площина, що проходить через оптичний центр $\mathbf{C}$ та лінію $\mathbf{l}$, яка визначається як $\Pi = \mathbf{P}^T\mathbf{l}$.

2.2.2 Дія проекційної матриці на площинах

Рівняння проекції (2.6) не залежить від вибору світової системи координат та системи координат зображення. Припустимо, що xy-координатна площина світової системи відліку задана на просторовій площині Π. Тоді координати точки на площині можна позначити як $\mathbf{X} = [x,\ y,\ 0,\ 1]^T$, зображення якої задається формулою [14]

$$\lambda\mathbf{x} = \mathbf{P}\mathbf{X} = \underbrace{[\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_4]}_{\mathbf{H}} \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \mathbf{H}\mathbf{x}_\pi, \qquad (2.11)$$

де $\mathbf{p}_i$ позначає i-й стовпчик $\mathbf{P}$; $\mathbf{x}_\pi = [x, y, 1]^T$ позначає двовимірну точку на просторовій площині. Тоді відображення стає гомографією просторової площини на площину зображення $\mathbf{H}$. (У комп'ютерному зорі гомографія - це матриця геометричного перетворення, яка відображає точки на одному плоскому зображенні у відповідні точки на іншому зображенні. Це показує, як виглядає сцена при спогляданні з різних ракурсів або точок зору.) Для точки у нескінченності $\mathbf{X}_\infty = [x, y, z, 0]^T$ її зображення називається точкою сходу $\mathbf{v}$, яка задається як [14]

$$\lambda \mathbf{v} = \mathbf{P}\mathbf{X}_\infty = \underbrace{[\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3]}_{\mathbf{H}_\infty} \cdot \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \mathbf{H}\overline{\mathbf{X}}_\infty, \tag{2.12}$$

де $\mathbf{H}_\infty = [\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3] = \mathbf{KR}$ називається гомографією нескінченної площини. $\mathbf{H}_\infty$ має велике значення в афінній реконструкції.

2.2.3 Дія проекційної матриці на конічні перерізи

Образ конічного перерізу $\mathbf{C}$ залишається конічним перерізом, як показано на рис. 2.2. Припустимо, що гомографія з конічної площини на площину зображення дорівнює $\mathbf{H}$, легко перевірити, що образ конічного перерізу $\mathbf{C}_m$ задається як [14]

$$\mathbf{C}_m = \mathbf{H}^{-T}\mathbf{C}\mathbf{H}^{-1}. \tag{2.13}$$

Для матриці камери $\mathbf{P}$ зворотна проекція $\mathbf{C}_m$ — це конус, визначений оптичним центром та зображеною конічною кривою [14]

$$\mathbf{Q} = \mathbf{P}^T\mathbf{C}_m\mathbf{P}. \tag{2.14}$$

Зокрема, абсолютний конічний переріз (абсолютна коніка) — це коніка на ідеальній площині $\Pi_\infty = [0, 0, 0, 1]^T$, яку можна виразити як $\Omega_\infty = \mathbf{I}_3$ або

$$[x, y, z]\Omega_\infty [x, y, z]^T = 0. \quad (2.15)$$

Образ абсолютної коніки можна отримати з (2.13) за допомогою нескінченної гомографії $\mathbf{H}_\infty$ [14]

$$\omega = \mathbf{H}_\infty^{-T} \Omega_\infty H_\infty^{-1} = \mathbf{K}^{-T}\mathbf{K}^{-1}. \quad (2.16)$$

Двосторонній образ (зображення) абсолютної коніки (абсолютного конічного перерізу) задається як $\omega* = \omega^{-1} = \mathbf{K}\mathbf{K}^T$. Як образ абсолютної коніки, так і двосторонній образ абсолютної коніки є невидимими уявними конічними перерізами на зображенні, які залежать лише від матриці калібрування камери $\mathbf{K}$. Якщо значення ω або $\omega*$ відоме, внутрішні параметри $\mathbf{K}$ можна відновити безпосередньо за допомогою розкладу Холецького [11]. Ми часто називаємо калібруванням камери обчислення образів абсолютної коніки або двосторонньої абсолютної коніки [14].

3 ВІДТВОРЕННЯ СТРУКТУРОВАНИХ СЦЕН НА ОСНОВІ ДВОХ ЗОБРАЖЕНЬ

3.1 Геометрія зображення з двох ракурсів

Застосування проективної геометрії в комп'ютерному зорі добре відоме в теорії відтворення форми на основі руху (structure from motion, SfM). Зокрема, коли припускаються лише два ракурси сценарію, проблему зазвичай називають стерео зором [14].

3.1.1 Епіполярна геометрія та фундаментальна матриця

Епіполярна геометрія — це внутрішня геометрія між двома ракурсами (проекціями), яка, по суті, є геометричною властивістю площин зображення та пучка площин, що проходять через базову лінію двох камер, як показано на рис. 3.1. Геометрія мотивована пошуком відповідностей між двома ракурсами. Дамо визначення ключових понять [14].

Епіполярна площина. Епіполярна площина – це площина, що проходить через два центри камер, позначена на рис. 3.1 як Π. Це пучок площин, віссю яких є базова лінія - пряма, що з’єднує центри двох камер [14].

Епіполь. Епіполь – це точка перетину базової лінії та площини зображення, що позначається як $\mathbf{e}$ та $\mathbf{e}'$ для лівого та правого епіполів. Очевидно, $\mathbf{e}$ – проекція центру правої камери на лівому зображенні, а $\mathbf{e}'$ – зображення лівого оптичного центру на правому зображенні. Іншими словами, положення правої камери відображається в точку $\mathbf{e}$ і навпаки [14].

Епіполярна лінія. Епіполярна лінія — це перетин епіполярної площини з площиною зображення, позначеною як $\mathbf{l}$ та $\mathbf{l}'$ на рис. 3.1. Очевидно, що проекція будь-якої точки на епіполярну площину лежить на епіполярних лініях. Усі епіполярні лінії перетинаються в епіполі [14].

Фундаментальна матриця. Фундаментальна матриця є алгебраїчним представленням епіполярної геометрії. Для будь-якої пари відповідних точок $\mathbf{x} \leftrightarrow \mathbf{x}'$ на двох зображеннях фундаментальна матриця $\mathbf{F}$ задовольняє умову

$$\mathbf{x}'^{T}\,\mathbf{F}\mathbf{x} = 0\,. \tag{3.1}$$

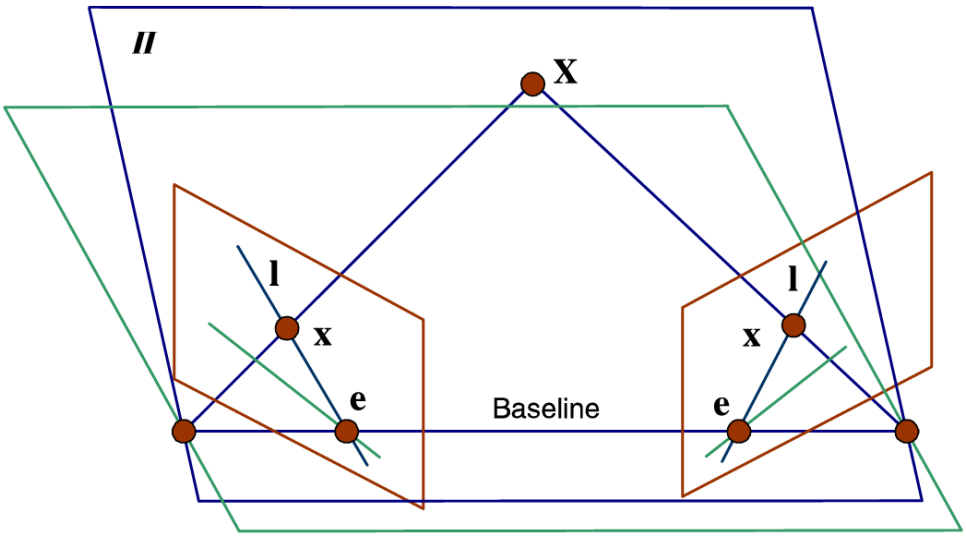


Рисунок 3.1 - Епіполярна геометрія між двома проекціями. Базова лінія, що з'єднує центри двох камер, перетинає кожне зображення в епіполі; будь-яка площина, що проходить через базову лінію, називається епіполярною площиною; епіполярна площина перетинає кожне зображення в епіполярній лінії. Усі епіполярні лінії проходять через епіполь

Фундаментальна матриця — це однорідна 3×3 матриця, визначена з точністю до масштабу. Однак вона має лише 7 ступенів свободи, оскільки ранг фундаментальної матриці $\text{rank}(\mathbf{F}) = 2$, що призводить до обмеження $\det(\mathbf{F}) = 0$ та усуває один ступінь свободи. У (3.1) кожна пара точок може задавати одне обмеження на $\mathbf{F}$, отже, фундаментальну матрицю можна вирахувати лише за 7 парами точок для нелінійного розв'язку або лінійно - за щонайменше 8 парами точок. Фундаментальна матриця тісно пов'язана з епіполярною геометрією [14].

Твердження 3.1. Нехай $\mathbf{F}$ — фундаментальна матриця двох зображень $\mathbf{I}$ і $\mathbf{I}'$, а $\mathbf{x} \leftrightarrow \mathbf{x}'$ - пара відповідних точок, $\mathbf{e}$ та $\mathbf{e}'$ — епіполі на двох зображеннях.

Тоді два епіполі задовольняють умовам $\mathbf{Fe}=0$, $\mathbf{F}^T\mathbf{e}'=0$. Епіполярна лінія, що відповідає $\mathbf{x}$ на іншому зображенні, дорівнює $\mathbf{l}'=\mathbf{Fx}$, а епіполярна лінія, що відповідає $\mathbf{x}'$, дорівнює $\mathbf{l}=\mathbf{F}^T\mathbf{x}'$ [14].

Твердження 3.2. Припустимо, що проекційні матриці двох ракурсів є $\mathbf{P}$ та $\mathbf{P}'$, тоді фундаментальна матриця задається як $\mathbf{F}=[\mathbf{e}']_\times\mathbf{P}'\mathbf{P}^+$, де $\mathbf{P}^+$- псевдообернена матриця до $\mathbf{P}$ (тобто $\mathbf{P}^+=(\mathbf{P}^T\mathbf{P})^{-1}\mathbf{P}^T$), $\mathbf{e}'$— епіполь, який задовольняє умову $\mathbf{e}'=\mathbf{P}'\mathbf{C}$ при $\mathbf{PC}=0$. Зокрема, для двох загальних камер $\mathbf{P}=\mathbf{K}[\mathbf{I}\,|\,\mathbf{0}]$, $\mathbf{P}'=\mathbf{K}'[\mathbf{R}\,|\,\mathbf{T}]$, маємо [14]

$$\mathbf{F}=\mathbf{K}'^{-T}[\mathbf{T}]_\times\mathbf{R}\mathbf{K}^{-1}=[\mathbf{K}'\mathbf{T}]_\times\mathbf{K}'\mathbf{R}\mathbf{K}^{-1}=\mathbf{K}'^{-T}\mathbf{R}\mathbf{K}^T[\mathbf{K}\mathbf{R}^T\mathbf{T}]_\times.$$

Тут $[\mathbf{v}]_\times=\begin{bmatrix}0 & -v_z & v_y\\ v_x & 0 & -v_x\\ -v_y & v_x & 0\end{bmatrix}$ - антисиметрична матриця вектора $\mathbf{v}=[v_x,v_y,v_z]^T$. Ця матрична форма безпосередньо використовується для обчислення векторного добутку шляхом множення матриць, тобто $\mathbf{v}\times\mathbf{x}=[\mathbf{v}]_\times\mathbf{x}$ [14].

Коли камери відкалібровані, а зображення нормалізовані як $\hat{\mathbf{x}}=\mathbf{K}^{-1}\mathbf{x}$, $\hat{\mathbf{x}}'=\mathbf{K}'^{-1}\mathbf{x}'$, ми маємо визначення, подібне до (3.1) [14]

$$\hat{\mathbf{x}}'^T\mathbf{E}\hat{\mathbf{x}}=0. \qquad (3.2)$$

Матриця $\mathbf{E}$ називається істотною (есенційною) матрицею. Очевидно, що істотна матриця та фундаментальна матриця пов'язані між собою як $\mathbf{E}=\mathbf{K}'^T\mathbf{F}\mathbf{K}$. Для нормалізованих зображень дві камери задаються як $\mathbf{P}=[\mathbf{I}\,|\,\mathbf{0}]$, $\mathbf{P}'=[\mathbf{R}\,|\,\mathbf{T}]$. Тоді з твердження 1.7 ми можемо одразу отримати [14]

$$\mathbf{E}=[\mathbf{T}]_\times\mathbf{R}=\mathbf{R}[\mathbf{R}^T\mathbf{T}]_\times. \qquad (3.3)$$

Істотна матриця - це однорідна матриця, визначена з точністю до загальної масштабної неоднозначності. Таким чином, $\mathbf{E}$ має лише п'ять ступенів свободи, оскільки $\mathbf{R}$ та $\mathbf{T}$ мають по три ступені свободи [14].

3.1.2 Тривимірна реконструкція

3D-реконструкція є однією з головних цілей комп'ютерного зору. Припустимо, що у нас є достатня кількість відповідностей зображень $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$, ми хочемо відновити відповідні 3D-точки $\mathbf{X}_i$ в просторі та матриці камери [14].

Як обговорювалося в попередньому підрозділі, фундаментальну матрицю можна відновити з точкових відповідностей. Враховуючи фундаментальну матрицю, можна визначити пару матриць проекції камери з точністю до проективного перетворення в P^3. Зокрема, пару канонічних камер можна вибрати як $\mathbf{P} = [\mathbf{I} \mid \mathbf{0}]$та $\mathbf{P}' = [[\mathbf{e}']_\times \mathbf{F} \mid \mathbf{e}']$, або, більш загально, $\mathbf{P}' = [[\mathbf{e}']_\times \mathbf{F} + \mathbf{e}'\mathbf{v}^T \mid \lambda \mathbf{e}']$, де $\mathbf{v}$ — довільний 3-вектор, а λ — ненульовий масштаб. Див. [16, 22] для доведення.

Після відновлення матриць проекцій, 3D-структуру будь-яких відповідностей зображень можна обчислити за допомогою тріангуляції [14]

$$\begin{cases} \lambda_i \mathbf{x}_i = \mathbf{P}\mathbf{X}_i, \\ \lambda'_i \mathbf{x}'_i = \mathbf{P}'\mathbf{X}_i. \end{cases} \tag{3.4}$$

Як показано на рис. 3.1, просторова точка $\mathbf{X}_i$ є перетином двох променів, що проектуються назад з двох точок зображення $\mathbf{x}_i$ та $\mathbf{x}'_i$. Це проективна реконструкція в , P^3 оскільки нічого не відомо про калібрування камер, а також про їхній відносний рух. Вибір матриць проекції камери не є унікальним, тому відновлена структура $\{\mathbf{X}_i\}$ відрізняється [14]. Будь-які дві реконструкції визначені з точністю до несингулярної матриці перетворення.

Припустимо, $(\mathbf{P}, \mathbf{P}', \{\mathbf{X}_i\})$ що - одна обчислена реконструкція, а $(\mathbf{P}_r, \mathbf{P}'_r, \{\mathbf{X}_{ri}\})$ - дійсна евклідова реконструкція, тобто $\{\mathbf{X}_{ri}\}$ - справжня структура в евклідовому просторі, яка зображена як $\{\mathbf{x}_i \leftrightarrow \mathbf{x}'_i\}$ двома дійсними камерами $\mathbf{P}_r$ та $\mathbf{P}'_r$. Тоді існує несингулярна матриця $\mathbf{H}$, яка задовольняє умову [14]

$$\mathbf{P}_r = \mathbf{P}\mathbf{H}^{-1} \qquad \mathbf{P}'_r = \mathbf{P}'\mathbf{H}^{-1}, \qquad \mathbf{X}_{ri} = \mathbf{H}\mathbf{X}_i . \tag{3.5}$$

Якщо $\mathbf{H}$ - проективна матриця перетворення, ми називаємо $(\mathbf{P}, \mathbf{P}', \{\mathbf{X}_i\})$ це проективною реконструкцією. Якщо $\mathbf{H}$ - афінна матриця перетворення, то реконструкція відбувається в афінній системі координат, ми називаємо її афінною реконструкцією. Якщо $\mathbf{H}$ - матриця перетворення подібності, ми називаємо її $(\mathbf{P}, \mathbf{P}', \{\mathbf{X}_i\})$ реконструкцією подібності або метричною реконструкцією. У більшості літератури її також називають евклідовою реконструкцією, оскільки евклідові властивості зберігаються в просторі подібності [14].

Зазвичай існує два способи отримання евклідової реконструкції: один – стратифікована реконструкція, інший – пряма метрична реконструкція. Стратифікований підхід починається з проективної реконструкції, яку можна відновити безпосередньо з фундаментальної матриці, потім розв'язок уточнюється до афінної та, нарешті, метричної реконструкції за допомогою додаткової інформації про сцену та камери. Суть стратифікації до афінної реконструкції полягає у визначенні площини на нескінченності за допомогою деяких обмежень сцени. Ключем до метричної реконструкції є ідентифікація абсолютної коніки [14, 15, 16].

Коли камери калібруються, ми можемо відновити матриці метричної проекції безпосередньо з основної матриці, і метричну реконструкцію можна отримати простим способом. Як показано в [16], у цьому випадку існує чотири можливі рішення, і лише одне рішення (істинна реконструкція) гарантує, що

реконструйовані точки лежать перед обома камерами. Хибні реконструкції можна виключити за допомогою точки вибірки [14].

3.2 Реконструкція структурованих сцен з двох зображень

Багато зусиль було витрачено на реконструкцію з двох ракурсів для розробки високоточних та повних методів моделювання складних сцен на основі зіставлення примітивних ознак. Однак багато методів не враховують геометричний контекст та обмеження, що виникають у сценах. У цьому розділі ми зосереджуємося на реконструкції структурованих сцен з некаліброваних зображень та представляємо практичний метод реконструкції, що включає інформацію про точки, лінії, площини та геометричні обмеження [14, 24].

У літературі є багато пов'язаних досліджень, які використовують обмеження сцени для реконструкції архітектури [18, 26]. Вернер та Ціссерман [25] запропонували генерувати площинну модель головних площин сцени шляхом прокрутки площин. Байяр та Ціссерман [17] використовували міжзображення гомографії для перевірки та оцінки моделей площин. Бартолі та Штурм [18] запропонували описувати геометричні зв'язки між точками та площинами за допомогою обмежень мультикомпланарності, а оцінювач максимальної правдоподібності, який включає обмеження та структури, використовувався у вигляді коригування зв'язків. Деякі комерційні системи, такі як Facade [20] та Photobuilder [19], можуть давати дуже реалістичні результати. Однак ці системи зазвичай вимагають багато взаємодії з людиною [14].

У цьому розділі представлено практичний гібридний метод реконструкції структурованих сцен з двох некаліброваних зображень. Метод базується на початковій оцінці головних гомографій за двовимірними точковими збігами. Потім зіставляються відрізки ліній між зображеннями та уточнюється гомографія [14].

3.2.1 Стратегія виявлення площин

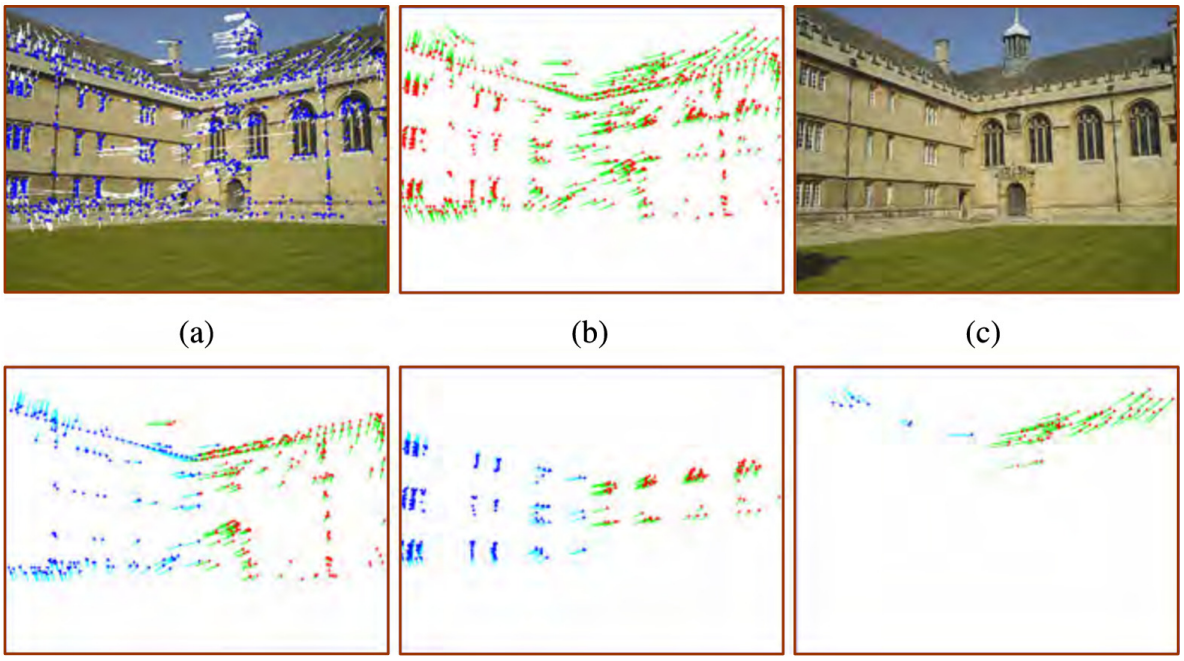
Стратегія базується на результаті початкового зіставлення точок. Спочатку ми оцінюємо головну площинну гомографію рекурсивним методом RANSAC. Потім знаходимо відповідні лінії, керуючись оціненою гомографією, та уточнюємо гомографію, включаючи як точкові, так і лінійні ознаки на площині [14].

3.2.1.1 Груба оцінка гомографії площини

Пошук відповідностей між зображеннями є складним завданням. Більшість доступних методів виявляються схильними до помилок для сцен штучних структур через неоднозначності, спричинені або великими однорідними областями текстури, або повторюваними візерунками на зображеннях. На рисунку 3.2 показано два зображення Оксфордського коледжу Вадхем. Було встановлено 1128 початкових збігів за допомогою методу кореляції та релаксації [27] з приблизно 13% викидів [14]. Для структурованого сценарію більшість збігів зазвичай лежать на кількох головних плоских поверхнях, і збіги на кожній поверхні пов'язані з плоскою гомографією. Для рекурсивної оцінки плоскої гомографії було використано механізм RANSAC [21].

Під час ітерації гіпотези про площини висуваються з випадково вибраної вибірки з чотирьох пар точкових збігів (якщо з початкових збігів можна обчислити прийнятну фундаментальну матрицю, то трьох пар збігів достатньо для генерації площинної гомографії [28]). Вибирається площина, яка найкраще підтримується тими парами точок, які не були призначені попереднім площинам. Слід зазначити, що цей метод може виявляти лише головні поверхні об'єкта з відносно великою кількістю початкових збігів. На рисунку 3.2 показано результат виявлення площин на основі початкових збігів. Допоміжні збіги для кожної площини показані різними кольорами з їх положеннями на першому зображенні, тоді як викиди та особливості, не пов'язані з виявленими площинами, виключаються. Всього 5 площин (які відповідають лівій та правій

стіні, правому даху та вікнам лівої та правої стін) виявляються автоматично [14].



(a) (b) (c)

(d) (e) (f)

Рисунок 3.2 - (a) та (c) два зображення Оксфордського коледжу Вадхем із загальною кількістю 1128 початкових збігів, показаних на першому зображенні з відмінностями від другого зображення; (b) виявлені правильні збіги на головних площинах, показані на першому зображенні з відмінностями; (d) опорні збіги лівої та правої стін; (e) опорні збіги лівого та правого вікон; (f) опорні збіги лівого та правого дахів

3.2.1.2 Зіставлення ліній на основі гомографії

Зіставлення ліній є складною проблемою, оскільки немає строгих геометричних обмежень, що дозволяють однозначно визначити суть, а топологічні зв'язки між відрізками ліній часто втрачаються під час сегментації. Оскільки більшість сегментів структурованого об'єкта лежать на виявлених головних поверхнях, зіставлення між цими відрізками ліній можна отримати за допомогою гомографій [14].

Автори [23] застосували детектор границь Кенні та алгоритм ортогональної регресії для підбору відрізків ліній на зображеннях. Припустимо, що загальна кількість витягнутих відрізків ліній на зображенні 1 та зображенні 2 дорівнює n_1 та n_2 відповідно. Для відрізка $\mathbf{l}'_j$ $(j = 1, ..., n_2)$ у другому вигляді ми можемо відобразити його в перший вигляд як $\hat{\mathbf{l}}_j \sim \mathbf{H}_k^T \mathbf{l}'_j$, де $\mathbf{H}_k$ — гомографія, індукована k-тою площиною об'єкта. Потім порівняти деякі співвідношення між $\hat{\mathbf{l}}_j$ та $.\mathbf{l}_i$ $(i = 1, ... n_1)$ згідно з наступними критеріями

$$\begin{cases} \text{angle}(\mathbf{l}_i, \hat{\mathbf{l}}_j) < \varepsilon_1, \\ \text{dist}^2(\mathbf{l}_i, \hat{\mathbf{l}}_j) < \varepsilon_2, \\ \text{overlap}(\mathbf{l}_i, \hat{\mathbf{l}}_j) > \varepsilon_3 \end{cases} \tag{3.6}$$

та застосувати стратегію «переможець отримує все» для вибору відповідних кандидатів, які відповідають площині. Тут $\text{angle}(\cdot,\cdot)$ позначає менший кут між двома прямими, $\text{dist}^2(\cdot,\cdot)$ позначає суму квадратів евклідової відстані між двома кінцевими точками першого відрізка до другої прямої, $\text{overlap}(\cdot,\cdot)$ позначає довжину перекриття двох відрізків [14].

Більшість отриманих вище варіантів співпадінь є правильними (тобто лежать на зображеннях k-тої площини в просторі). Однак деякі з них можуть лежати поза площиною (тут вони називаються неправильними співпадіннями). Для двох пар співпадаючих прямих $\mathbf{l}_i \leftrightarrow \mathbf{l}'_i$ і $\mathbf{l}_j \leftrightarrow \mathbf{l}'_j$ якщо вони відповідають двом компланарним прямим у просторі, як показано на рис. 3.3, то їхні перетини $\mathbf{x}_{ij}$ і $\mathbf{x}'_{ij}$ повинні задовольняти умову [14]

$$\mathbf{x}'_{ij} \sim \mathbf{H}_k \mathbf{x}_{ij} \tag{3.7}$$

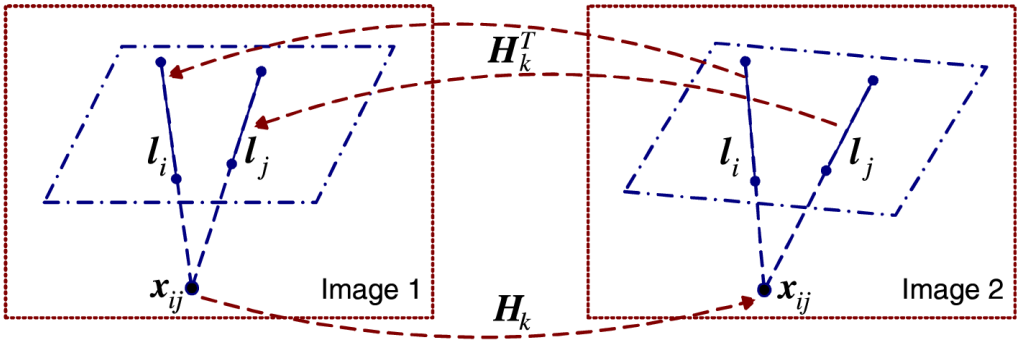


Рисунок 3.3 - Перетин двох компланарних прямих все ще лежить на площині, а його зображення у двох проекціях задовольняють обмеження гомографії

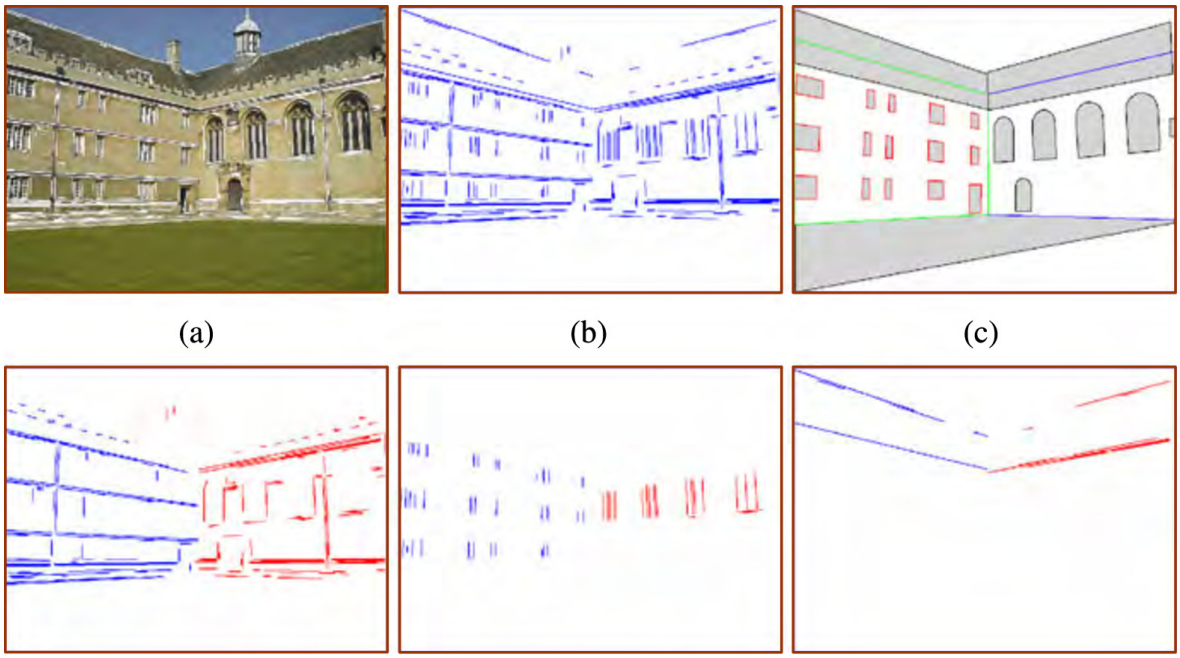
Для кожної пари отриманих кандидатів ми обчислюємо її точки перетину з рештою пар прямих. Якщо більше 50% перетинів задовольняють (1.41), то ця пара вважається правильною. В іншому випадку, цей збіг виключається зі списку. Аналогічно, ми можемо отримати всі правильні збіги прямих на кожній площині. На рисунку 3.4 показано результати збігу, що відповідають кожній виявленій просторовій площині [14].

3.2.1.3 Удосконалення гомографії

Після отримання правильних збігів характерних точок та відрізків ліній на виявлених площинах, гомографію можна повторно оцінити, включивши інформацію як про точки, так і про лінії, оскільки лінійні ознаки є стабільнішими та можуть бути виявлені точніше, ніж точкові ознаки [14]

$$\begin{cases} \mathbf{x}_i \sim \mathbf{H}_k \mathbf{x}_i & (i = 1, ..., m), \\ \mathbf{l}_j \sim \mathbf{H}_k^T \mathbf{l}_j & (j = 1, ..., n) \end{cases} \tag{3.8}$$

де m та n – кількість збігів точок та ліній на k-й площині в просторі. Наведена вище система рівнянь може забезпечити $2(m+n)$ лінійні обмеження на 8 об'єктів гомографії. Таким чином, $\mathbf{H}_k$ може бути обчислена методом найменших квадратів. Розв'язок може бути додатково оптимізований шляхом мінімізації наступної функції вартості [14]



(a) (b) (c)

(d) (e) (f)

Рисунок 3.4 - (a) початково виявлені 328 відрізків ліній на першому зображенні; (b) усі правильні збіги ліній на головних площинах; (c) виявлені контури 6 головних площин; (d) правильні збіги ліній, виявлені на лівій та правій стінах; (e) правильні збіги ліній, виявлені на лівому та правому вікнах; (f) правильні збіги ліній, виявлені на лівому та правому дахах

$$J(\mathbf{H}_k) = \frac{1}{m}\sum_{i=1}^{m}(d^2(\mathbf{x}_i' , \mathbf{H}_k\mathbf{x}_i) + d^2(\mathbf{x}_i, \mathbf{H}_k^{-1}\mathbf{x}_i)) + \frac{1}{2n}\sum_{j=1}^{n}(\mathrm{dist}^2(\mathbf{l}_j, \mathbf{H}_k^T\mathbf{l}'_j) + \mathrm{dist}^2(\mathbf{l}'_j, \mathbf{H}_k^{-T}\mathbf{l}_j)), \quad (3.9)$$

де перший член – це нормалізована геометрична помилка повторного проектування збігів точок у двосторонніх напрямках, другий член – це нормалізована помилка збігів прямих у двох проекціях та $d^2(\cdot,\cdot)$ позначає суму квадратів евклідової відстані між двома точками. Використовуючи уточнені гомографії, ми можемо знайти більше опорних збігів точок та прямих на виявлених головних поверхнях та знову уточнити гомографію [14].

Після отримання всіх гомографій плоских поверхонь ми можемо обчислити лінію перетину між кожною парою площин. Припустимо, що $\mathbf{H}_1$ і $\mathbf{H}_2$ - гомографії, індуковані двома просторовими площинами. Нехай $\mathbf{l} \leftrightarrow \mathbf{l}'$ - проекції перетину двох площин. Тоді з [14]

$$\mathbf{l} \sim \mathbf{H}_1^T \mathbf{l}', \qquad \mathbf{l} \sim \mathbf{H}_2^T \mathbf{l}' \tag{3.10}$$

ми маємо

$$(\mathbf{H}_2^{-1}\mathbf{H}_1)^T \mathbf{l} \sim \mathbf{l} \tag{3.11}$$

Таким чином, перетин у першому вигляді можна визначити з власного вектора, що відповідає дійсному власному значенню матриці $(\mathbf{H}_2^{-1}\mathbf{H}_1)^T$. Фактично, матриця $\mathbf{H}_2^{-1}\mathbf{H}_1$ є відображенням першого зображення на себе. Це площинна гомологія, яка має нерухому точку (вершину) та лінію нерухомих точок (вісь) [16]. Перетворення має два рівні та одне різне власні значення. Вісь, яка є лінією перетину двох площин, є точкою з'єднання власних векторів, що відповідають виродженим власним значенням. Третій власний вектор відповідає вершині, яка є епіполем у другому вигляді [14].

Згідно з перетинами будь-яких двох суміжних площин та збігами прямих на кожній площині, легко отримати контур кожної виявленої площини. Приклад наведено на рисунку 3.4. Гомографія забезпечує однозначне відображення точок і прямих між двома проекціями. Для будь-яких точок або ліній на виявлених площинах на одному зображенні ми можемо негайно отримати їх відповідності на іншому проекції, навіть якщо ці відповідності можуть бути перекриті або лежати поза зображенням [14].

ВИСНОВКИ

Відтворення тривимірної структури сцени на основі зображень є однією з основних задач комп'ютерного зору. Особливо ефективний підхід до тривимірної (3D) реконструкції передбачає використання великої кількості зображень нерухомої сцени. Ця проблема, яку зазвичай називають багаторакурсною реконструкцією структури з руху (multiview SfM) є предметом численних досліджень у галузі комп'ютерного зору.

В роботі дано опис процесу отримання зображень за допомогою однієї або двох камер. Основною концепцією тут є проста модель камери-обскури, яка використовується для пояснення перетворення з тривимірного простору світу у двовимірну площину зображення, що здійснюється камерою. Далі вводяться так звані зовнішні та внутрішні параметри камери. Коли зображення сцени знімаються одночасно двома камерами, ці камери називають стереопарою і вони створюють стереопари зображень. Властивості камер, налаштованих таким чином, визначаються їхньою епіполярною геометрією, яка описує взаємозв'язок між точками простору, що спостерігаються в їхніх полях зору, та зображеннями, що падають на відповідні площини зйомки. Місця розташування кожної точки простору на площині зображення, як їх сприймає пара камер, називаються відповідними або збіжними точками. Відповідні точки на зображеннях стереопари пов'язані між собою за допомогою фундаментальної матриці. Якщо вона відома, то надає фундаментальну інформацію про епіполярну геометрію розташування стереопари. Однак пошук відповідних точок між зображеннями — це нетривіальне завдання. Існує багато факторів, які можуть ускладнити цей процес, таких як затуляння, обмежена роздільна здатність зображення та квантування, спотворення, шум та багато інших. З технічної точки зору, зіставлення вважається обмеженим: у порівняних зображеннях немає достатньої інформації, щоб гарантувати знаходження єдиного відповідника. Проте зіставлення можна спростити,

застосувавши набір правил, відомих як стереообмеження, найважливішим з яких є епіполярне обмеження, що передбачає, що відповідні точки завжди лежать на відповідних епіполярних лініях. Епіполярне обмеження звужує пошук відповідних точок з усього двовимірного простору до одновимірного простору епіполярних ліній.

ПЕРЕЛІК ПОСИЛАНЬ

1. Tan, L., & Jiang, J. (2018). Digital signal processing: fundamentals and applications. Academic press.

2. Embree, P. M. (1995). C Language Algorithms for Real-Time DSP. Prentice-Hall, Inc.

3. Rabiner, L. R., & Schafer, R. W. (1978). Digital Processing of Speech Signals. Prentice Hall.

4. Webster, J. G. (Ed.). (2009). Medical instrumentation: application and design. John Wiley & Sons.

5. Deller Jr, J. R., Proakis, J. G., & Hansen, J. H. (1993). Discrete time processing of speech signals. Prentice Hall PTR.

6. Li, Z. N., Drew, M. S., & Liu, J. (2004). Fundamentals of multimedia (pp. 253-265). Upper Saddle River (NJ): Pearson Prentice Hall.

7. Pan, D. (1995). A tutorial on MPEG/audio compression. IEEE multimedia, 2(2), 60-74.

8. Gonzalez, R. C. (2009). Digital image processing. Pearson education India.

9. Ahmed, N., & Natarajan, T. (1983). Discrete-Time Signals and Systems, Reston Pub. Company. Inc., Virginia.

10. Ambardar, A. (1995). Analog and digital signal processing. Boston, MA, USA: PWS.

11. Alkin, O. (1994). Digital signal processing: a laboratory approach using PC-DSP. Prentice Hall PTR.

12. Oppenheim, A. V., & Shafer, R. W. (1975). Discrete-Time Signal Processing. Prentice Hall.

13. Proakis, J. G. (2007). Digital signal processing: principles, algorithms, and applications, 4/e. Pearson Education India.

14. Roddy, D., & Coolen, J. (1997). Electronic Communications, 4/e. Prentice Hall.

15. Tomasi, W. (2009). Electronic Communications System: Fundamentals Through Advanced, 5/e. Pearson Education India.

16. Stearns, S. D., & Hush, D. R. (1990). Digital signal analysis. Englewood Cliffs.