ISSN 2519-2310 (Online)

КОМП'ЮТЕРНІ НАУКИ ТА КІБЕРБЕЗПЕКА, 2024, № 1 (25)

COMPUTER SCIENCE AND CYBERSECURITY, 2024, Iss. 1 (25)

DOI: https://doi.org/10.26565/2519-2310-2024-1-06
УДК 004.056.5

ВІДНОВЛЕННЯ ТРИВИМІРНИХ СЦЕН НА ОСНОВІ ДАНИХ ВІДЕО ПОТОКІВ

Денис Грульов[1], магістрант, e-mail: xa11800855@student.karazin.ua,
ORCID: https://orcid.org/0009-0005-8506-770X
Анастасія Морозова[1], доцент, доктор філософії, e-mail: a.morozova@karazin.ua,
ORCID: https://orcid.org/0000-0003-2143-7992
Петро Доля[1], доцент, доктор філософії, e-mail: pdolya@karazin.ua,
ORCID: https://orcid.org/0009-0002-4062-4443
Лілія Бєлова[1], старший викладач, e-mail: l.belova@karazin.ua,
ORCID: https://orcid.org/0009-0007-0805-4547
[1]*Харківський національний університет імені В.Н. Каразіна, майдан Свободи, 4, Харків, 61022, Україна*

Рукопис надійшов 23 березня 2024 р. Отримано після рецензування 29 квітня 2024 р.
Прийнято 30 травня 2024 р.

Анотація: Дана робота присвячена застосуванню сучасних алгоритмів відновлення тривимірних сцен з зображень для отримання просторової інформації із відео. У роботі розглядається розмаїття сучасних методів, підходів та алгоритмів в області аналізу відео потоку. Приділено увагу послідовності розвитку підходів до вирішення задачі. У процесі дослідження області та результатів, пов'язаних з тривимірною реконструкцією на основі зображень та відео потоків, був винайдений алгоритм, що дозволяє будувати щільні мапи глибини, використовуючи інформацію з усіх кадрів відео. Ідея полягає у тому, щоб використовувати готові, загальноприйняті та перевірені рішення для вирішення двох задач: COLMAP - для візуальної одометрії, та RAFT - для обчислення оптичного потоку. Запропонований алгоритм показує досить точні результати, та відновлює мапу глибини в деталях на довільних статичних сценах.

Ключові слова: *відео потік, 3D-реконструкція, машинного навчання, одометрія, нейронна мережа, комп'ютерний зір, мапа глибини, оптичний потік*

Як цитувати: Грульов Д., Морозова А., Доля П., Бєлова Л.. Відновлення тривимірних сцен на основі даних відео потоків. *Комп'ютерні науки та кібербезпека.* 2024; № 1(25): С. 66–75. https://doi.org/ 10.26565/2519-2310-2024-1-06

In cites: Hrulov D., Morozova A., Dolia P., Bielova L. (2024). Reconstruction of three-dimensional scenes based on video flow data. *Computer Science and Cybersecurity.* 1(25): 66–75. https://doi.org/10.26565/2519-2310-2024-1-06 (in Ukrainian)

1. Вступ

Задача відновлення просторових даних з відео відноситься до загальної задачі відновлення структури із руху (Structure-from-Motion, SfM) [1]. На даний момент існують

алгоритми, що якісно відновлюють просторову інформацію з відео потоку. Відео потік заздалегідь містить більше інформації, ніж окремі фото, зокрема послідовність, в яких ці фото розташовані, і самі послідовні фото є зображеннями, що різняться дуже мало, що дозволяє, наприклад, точно знаходити щільні відповідності між точками та відстежувати траєкторії об'єктів. Проте, більшість алгоритмів, що аналізують просторову структуру на основі відео, спираються на попарний аналіз зображень. За допомогою попарної обробки кадрів, алгоритми можуть визначати відповідності між об'єктами на кадрах. Ці параметри можуть бути використані для подальшої реконструкції тривимірної моделі сцени. Недоліком цього підходу є те, що об'єм обчислень може бути неприйнятно великим для ряду застосувань.

Методи тривимірної реконструкції загалом можна розділити на класичні (традиційні) та підходи машинного навчання (ML). Класичні методи 3D-реконструкції зазвичай спираються на геометричні принципи, такі як тріангуляція, епіполярна геометрія та калібрування камери. Ці методи часто передбачають явне моделювання геометрії та фізики процесу створення зображень і можуть вимагати ручних або напівручних кроків для виділення ознак, зіставлення та реконструкції. З іншого боку, підходи машинного навчання (ML) для 3D-реконструкції використовують алгоритми, навчені на великих наборах даних, щоб вивчати шаблони та зв'язки між даними.

У багатьох випадках, в останніх дослідженнях класичні та засновані на машинному навчанні методи тривимірної реконструкції використовуються разом, щоб доповнити сильні та слабкі сторони першого та другого підходів. Поєднання класичного підходу та підходу на основі машинного навчання може використовувати переваги обох підходів, що призводить до підвищення точності та надійності завдань 3D-реконструкції.

Задача відновлення просторової інформації з відео потоку є актуальною задачею комп'ютерного зору, а саме його підрозділу – 3D реконструкції. Вилучення просторової інформації з відео потоку є дуже важливим для таких застосувань, як робототехніка, самокерування автомобілів, доповнена та віртуальна реальність, тощо.

Мета цієї статті є дослідити можливості застосування сучасних алгоритмів відновлення тривимірних сцен з двох зображень (Two-View Structure from Motion) для відновлення просторової інформації із відео потоку.

У статті розглядаються алгоритми та підходи, що допомагають відновити просторову інформацію з кадрів відео для відновлення тривимірних сцен лише із двох зображень. А також розробка алгоритму відновлення тривимірної інформації про сцену з усіх кадрів відео.

2. Огляд існуючих рішень

За останні два десятиліття сфера 3D-реконструкції з відеокадрів значно розвинулась завдяки прогресу як у традиційних методах вилучення ознак, так і в сучасних методах глибокого навчання.

Останні розробки демонструють потенціал глибокого навчання для вирішення складних сценаріїв реконструкції та продовжують розширювати межі можливого в цій галузі. У 2020 році розроблений алгоритм, що реконструює мапу глибини для кожного кадру відео [2]. Алгоритм показує якісні та стабільні результати застосовано до відео, що знімають довільні сцени, проте, у сенсі реконструювання динамічних сцен, спеціалізований до оцінки руху людей. Не зважаючи на те, що перевагою алгоритму висувається виключна ступінь узгодженості реконструкції між кадрами, алгоритм використовує попарну обробку кадрів, що вибираються за описаним принципом, як базовий етап навчання моделі.

В останні роки розробка застосувань візуальної одометрії на основі засобів, таких як ORB-SLAM [3] і DSO (пряма розріджена одометрія), продемонструвала здатність забезпечити оцінку руху камери в реальному часі з високою точністю. Ці методи використовують ефективні

методи вилучення функцій, зіставлення та оптимізації для досягнення продуктивності в реальному часі на сучасному обладнанні. ORB-SLAM – це алгоритм та програмне забезпечення, що активно доповнюється, розробляється та покращується. На даний момент розроблена вже третя 15 версія цього – алгоритму ORB-SLAM3 [4].

Сучасні методи глибокого навчання також дозволили побудувати алгоритми візуальної одометрії, що дають точні та надійні результати та на даний момент наближаються до готовності використання у реальних умовах. Одним із новітніх алгоритмів, що заслуговує уваги, є Deep Patch Odometry [5]. В основі даного алгоритму лежить розбиття кадрів відео на клаптики (patches) за допомогою однієї нейронної мережі та відстежування їх руху за допомогою іншої, рекурентної нейронної мережі.

Повертаючись до задачі відновлення тривимірних сцен з двох зображень, в 2021 була оприлюднена нейромережева модель RAFT-Stereo, що вирішує задачу, та на момент 2022 року посідає друге місце у рейтингу RVC Stereo [6]. Алгоритм є модифікацією нейромережі для знаходження оптичного потоку RAFT (Recurrent All-Pairs Field Transforms). У 2022 році був також винайдений алгоритм CREStereo, що показує ще кращий результат та може давати ще точніші реконструкції [7].

Проте, треба зазначити, що алгоритми, які базуються на машинному навчанні, сильно залежать від вибірки, на якій були навчені. Більш того, такі підходи часто мають погану узагальнювальну здібність між вибірками, тому на практиці потребують навчання під конкретні умови, для повного охоплення яких, потрібні надто великі вибірки. Тому адаптація алгоритму під конкретні умови, буде потребувати дуже серйозного об’єму обчислень без жодної гарантії на прийнятний результат в умовах, які можуть лише незначно відрізнятися від тих, до яких він був адаптований.

Хоча монокулярна реконструкція глибини відео досягла значного прогресу та застосовувалася до різних програм, таких як робототехніка, доповнена реальність та автономне водіння, вона залишається активною областю досліджень із постійними викликами та досягненнями. Тому задача відновлення мап глибини з двох зображень є актуальною задачею.

3. Класичний підхід до задачі відновлення просторової інформації з 2-вимірних зображень

У даній роботі розглядається задача аналізу просторової інформації у контексті знаходження мап глибини зображень. Мапа глибини – це зображення, в якому кожному пікселю відповідає значення, що відображає його відстань від камери або іншого датчика.

Структура з руху (Structure from Motion, SfM) — це техніка, яка використовується в комп’ютерному зорі для реконструкції 3Dструктури сцени або об’єкта з набору 2Dзображень або відеокадрів, знятих з різних точок зору. Основна ідея SfM полягає у використанні візуальної відповідності між об’єктами на двох зображеннях для оцінки 3D-положень цих об’єктів, а також позиції камери.

Рис. 3.1. Мапи глибини
Fig. 3.1. Depth maps

Існує декілька підходів до реалізації алгоритму Structure-from-Motion.

Класичний підхід до SfM передбачає ітеративне поєднання результати обробки двох кадрів. Для уточнення відновлених поз камер та розрідженої хмари точок використовується налаштування пучка (bundle adjustment), що мінімізує загальну помилку повторної проекції поміж усіма ракурсами (Рис. 3.2.)

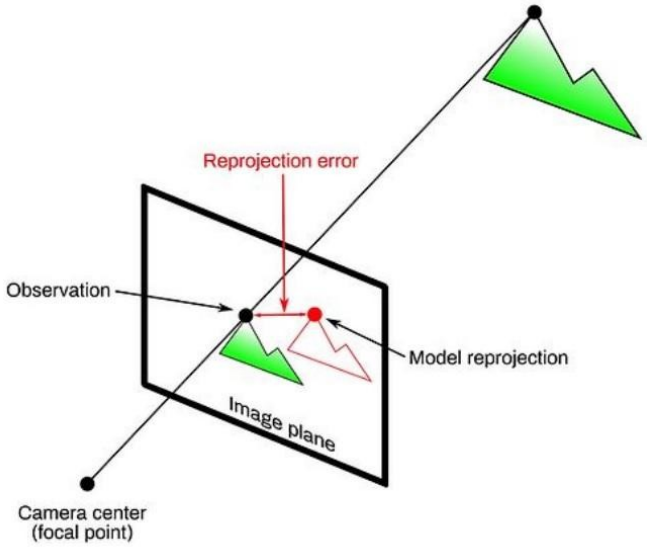

Рис. 3.2. Класичний підхід до SfM
Fig. 3.2. A classical approach to SfM

Розповсюдженою технікою аналізу відео потоку, є оптичний потік, що в тому числі дозволяє слідкувати за рухом об’єктів. Це поле векторів руху, яке описує, як об’єкти на зображенні рухаються відносно спостерігача.

Також, оптичний потік також можна використовувати безпосередньо для встановлення візуальних відповідностей між точками на двох зображеннях.

4. Комбінація сучасних засобів комп’ютерного зору для відновлення мап глибини з кадрів відео

Не дивлячись на те, що задача відновлення мап глибини з двох зображень, є складною, та дослідження якої ведеться багатьма науковцями у даний момент, реконструкція мап глибини з відео, як уже зазначалося, має досить задовільні рішення, якщо не брати до уваги швидкість роботи алгоритму. У даній роботі пропонується алгоритм відновлення мап глибини, що використовує традиційні геометричні методи для відновлення орієнтації та позиції камери у просторі для кожного кадру, та обчислення оптичного потоку за допомогою глибокого навчання.

Для реалізації класичного алгоритму Structure-from-Motion, існує достатньо готового програмного забезпечення. Алгоритм, що вирішує задачу Structure-From-Motion, за визначенням також вирішує задачу одометрії – задачі знаходження орієнтації та позиції камери. Вибір засобу для вирішення задачі одометрії не є принциповим, якщо швидкість виконання не є важливою характеристикою.

Основна проблема, що є у існуючих класичних засобів вирішення задачі Structure-from-Motion – знаходження щільних мап глибини. Мапи глибини, хоча і можуть бути отримані за допомогою таких програмних засобів як COLMAP, мають значні області невизначеності,

оскільки реконструкція базуються на знаходженні деякої кількості обраних пікселів на зображеннях та їх зіставленні як проекцій на площину зображення для знаходження відповідностей між ними. В результаті отримується хмара точок у просторі, кожна точка якої проектується на деяку кількість зображень. Таким чином, мапи глибини фактично будуються на базі проекцій хмари відтворених точок на різні кадри відео, що мають різні ракурси. Приклад мапи глибини, отриманої за допомогою програмних засобів як COLMAP, наведено на рис. 4.1.

Рис. 4.1. Мапа глибини за допомогою COLMAP
Fig. 4.1. Depth map using COLMAP

При застосуванні вищеописаного підходу, попіксельна мапа глибини не отримується. Більш щільні мапи глибини будуються вже на базі первинної реконструкції, проте, і ці мапи глибини на практиці мають області невизначеності. Для того, щоб відновити щільні мапи глибини, пропонується спочатку отримати позиції камер. Для того, щоб відновити мапу глибини для деякого кадру, пропонується використовувати інформацію про позиції камери інших кадрів, та поле оптичного потоку між кадром, для якого будується мапа глибини, та відповідними кадрами відео.

Першим етапом є відновлення положення камер. Для відновлення положення камер, обрано COLMAP, так як це добре перевірене ПЗ, що дуже тонко конфігурується та має зручний консольний інтерфейс та дозволяє зчитувати результати реконструкції, зокрема позиції камер для кожного кадру відео, та параметри калібровки камери.

Наступним етапом є переведення отриманих позицій камер у спільну систему координат. Дані про положення камери, згідно з традиційною схемою, отримуються у вигляді матриць повороту R та векторів переміщення t. Для того щоб перетворити вектор $(X, Y, Z)^T$ зі світової системи координат до локальної системи координат камери, треба застосувати вектор переміщення та матрицю повороту камери:

$$\begin{pmatrix} X' \\ Y' \\ Z' \end{pmatrix} = R \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} + t \tag{1}$$

Відповідно, для того, щоб знайти глобальні координати вектора, треба виразити його з рівняння:

$$\begin{pmatrix} \mathrm{X} \\ \mathrm{Y} \\ \mathrm{Z} \end{pmatrix} = \mathrm{R}^{-1} \left(\begin{pmatrix} \mathrm{X}' \\ \mathrm{Y}' \\ \mathrm{Z}' \end{pmatrix} - \mathrm{t} \right) \tag{2}$$

Враховуючи, що $R^{-1} = R^T$, бо матриця повороту завжди ортогональна, отримуємо зручний для обчислення перехід до єдиної системи координат.

Тепер, розглянемо вектор між фокусним центром камери та пікселем (u, v), що є точкою, що лежить на площині зображення. Z-координата будь якої точки площини зображення дорівнює фокальній відстані f, а центровані координати пікселя – $(u - c_x, v - c_y)$. Таким чином, координати точки на площині зображення у системі координат камери дорівнюють $(f, u - c_x, v - c_y)$. Оскільки оптичний центр камери знаходиться у центрі системи координат камери, то і шуканий вектор матиме вигляд

$$a = (f, u - c_x, v - c_y) \tag{3}$$

Тепер, якщо позиція камери відома, перетворенням вектору a до світової системи координат безпосередньо отримується напрям променя, на якому лежить точка, що відображається у піксель (u, v). Сам промінь отримується з обмеження, що у рівнянні

$$\begin{gathered} l = as + b \\ b = t, \text{при } s = 0 \end{gathered} \tag{4}$$

За допомогою оптичного потоку, для кожної точки першого зображення, можна знайти відповідну точку на іншому. Оптичний потік відображає відповідності краще всього, коли зображення, що порівнюються, відрізняються не сильно. Тому, для кожного кадру, є сенс підбирати деяку кількість найближчих кадрів. Для кожного кадру, що зіставляються, обчислимо оптичний потік, для кожного пікселя вихідного кадру (u, v), підберемо відповідний піксель (u', v') за допомогою обчислених полів оптичного потоку.

Для того, щоб отримати відповідну точку на іншому зображенні за допомогою оптичного потоку, треба просто змістити її на значення оптичного потоку між першим та другим зображенням:

$$(u', v') = (u, v) + flow_{u,v} \tag{5}$$

Тоді вектор між центром координат камери іншого кадру та відповідною точкою на іншому кадрі дорівнює

$$a' = \left(f, u' - c_x, v' - c_y\right) \tag{6}$$

Рис. 4.2. Відповідності між точками зображень
Fig. 4.2. Matches between image points

В такий спосіб, з наявної інформації про відповідності між точкою на кадрі, для якого будується мапа глибини, та точками на кожному іншому кадрі, отримуються промені у просторі, на перетині яких має безпосередньо лежати точка, що проектується у відповідні пікселі на двох кадрах.

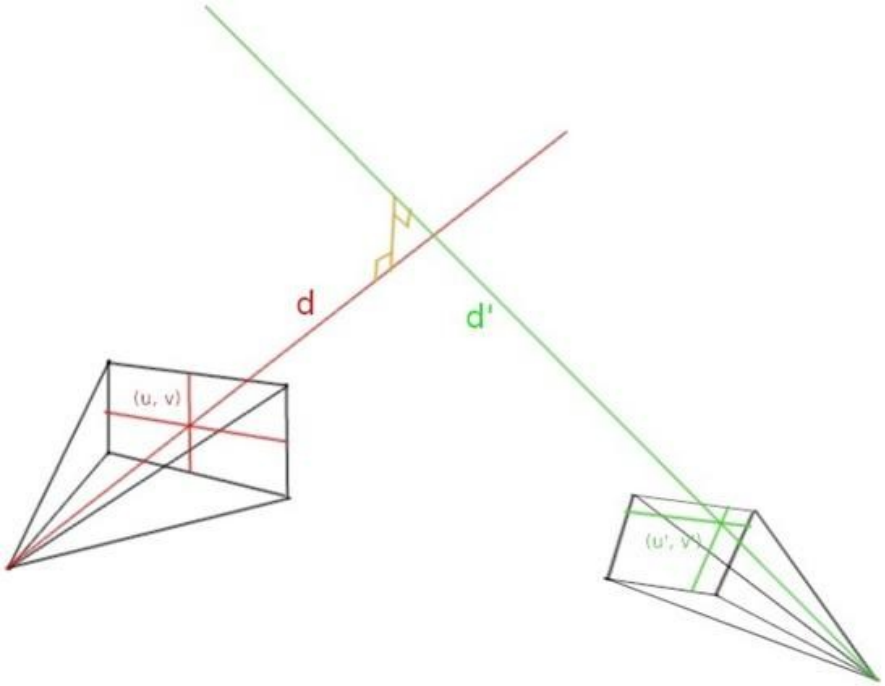

Рис. 4.3. Відстань до точки, що проектується на два кадри
Fig. 4.3. The distance to the point projected on two frames

На практиці, промені, що розглядаються, не будуть точно перетинатися, бо для знаходження відповідностей для одометрії використовується розріджене співставлення точок, за допомогою алгоритмів SIFT, а для знаходження щільних відповідностей використовується оптичний потік, отриманий за допомогою алгоритмів глибокого навчання. Тому, за допомогою такого методу точку перетину можна знайти лише з деякою точністю. Треба зауважити, що точку перетину при побудові мапи глибини знаходити не обов'язково, а лише треба знайти відстань d від початку променя у центрі координат камери до найближчої точки до прямої l' на прямій l.

Позначимо відповідний промінь як функцію від s:

$$l(s): l = as + b, s \geq 0 \tag{7}$$

Якщо вектор a нормалізований, то s дорівнює відстані від початку променя до $l(s)$. Таким чином, $d = s$.

Розглянемо прямі, що відповідають двом відповідним точках на двох кадрах:

$$\begin{aligned} l(d) &= ad + b \\ l'(d') &= a'd' + b' \end{aligned} \tag{8}$$

Тепер, задачу можна переформулювати як знаходження таких s, s', за яких відстань від $l(s)$ до $l'(s')$ мінімальна (рис 4.3).

Щоб вирішити цю задачу, треба прийняти до уваги той факт, що вектор між точками двох прямих у просторі, що має найменшу можливу довжину, перпендикулярний до обох прямих:

$$\begin{aligned} &< l(d) - l'(d'), a >= 0 \\ &< l(d) - l'(d'), a' >= 0 \end{aligned} \tag{9}$$

Дана система лінійна відносно d та d'. Безпосередньо після вирішення системи з двома рівняннями та невідомими, отримуємо значення d, що дорівнює відстані від фокусного центру камери до об'єкта у просторі, що відображається на обраний піксель.

Проводячи дані обчислення для кожного пікселя кадру, для якого будується мапа глибини стільки разів, скільки є кадрів, на яких було знайдено відповідний піксель до даного, отримуємо "гіпотези", які можна певним чином фільтрувати та усереднювати для отримання остаточного передбачення відстані до об'єкта. В імплементації цього алгоритма було вирішено брати до уваги тільки ті відповідності, для яких

$$\frac{\|l(d) - l'(d')\|}{d} \leq 0.01 \tag{10}$$

Також, для базової фільтрації хибних відповідностей, що генерує оптичний потік, алгоритм обчислює також оптичний потік між кадром, що є середнім за номером між вихідним та одним із кадрів на якому шукаються відповідності. Це означає, що крім оптичного потоку

між вихідним кадром I_j та кадром I_{j+k}, також обчислюється оптичний потік між $I_{j+\frac{k}{2}}$ та I_{j+k}.

Позначимо положення пікселів з кадру I_j на кадрі I_{j+k} як $I_j \to I_{j+k}$. Так, як оптичний потік між I_j та $I_{j+\frac{k}{2}}$ вже обчислений за побудовою алгоритму, то для кожного пікселя також можна знайти відповідність $(I_j \to I_{j+\frac{k}{2}}) \to I_{j+k}$.

Якщо оптичний потік знаходиться алгоритмом ідеально, то $I_j \to I_{j+k}$ повинне представляти ту ж саму відповідність що і $(I_j \to I_{j+\frac{k}{2}}) \to I_{j+k}$. Але на практиці, навіть такий точний алгоритм як RAFT, може давати хибні результати у ряді випадків. Для фільтрації хибних відповідностей у даній роботі пропонується перевіряти вищезазначену умову (10).

Проте, як і у випадку знаходження "приблизного" перетину між променями, треба задати деякий поріг допустимого відхилення. Загалом, це дає просте правило, що відсіює значну кількість хибних відповідностей.

Для усереднення результатів використовувалися емпіричні ваги, які виникли з ідеї про те, що при малих кутах між прямими, важливу роль мають помилки обчислення та неточності роботи алгоритмів одометрії та знаходження оптичного потоку, тому достовірність таких результатів знижується, і при нульовому куті – це взагалі вироджений випадок.

Простими для обчислення ваговими коефіцієнтами, що дозволяють "заглушити" шум від недостовірних результатів, є $sin^2(\alpha)$, де α – кут між α та α'. Обчислити, як відомо, ці коефіцієнти можна як

$$sin^2(\alpha) = 1 - cos^2(\alpha) = 1 - < a, a' >^2 \tag{11}$$

5. Висновки

У роботі були проаналізовані можливості застосування сучасних алгоритмів відновлення тривимірних сцен з зображень для відновлення просторової інформації із відео потоку. Були розглянуті методи, підходи та алгоритми, а також сучасні тренди в області реконструкції тривимірної інформації з відео. Приділено увагу послідовності розвитку підходів до вирішення задачі – можна підсумувати, що алгоритми відновлення тривимірної інформації починали розвиватися з суто геометричних підходів і в останні роки все більше використовують глибоке навчання. У процесі дослідження області та результатів, пов'язаних з тривимірною реконструкцією на основі зображень та відео отоків, був винайдений алгоритм, що дозволяє будувати щільні мапи глибини, використовуючи інформацію з усіх кадрів відео.

Конфлікт інтересів

Автори повідомляють про відсутність конфлікту інтересів.

References

1. Deep Two-View Structure-from-Motion Revisited. URL: https://openaccess.thecvf.com/content/CVPR2021/papers/Wang_Deep_Two-View_Structure-From-Motion_Revisited_CVPR_2021_paper.pdf
2. Xuan Luo, Jia-Bin Huang, Richard Szeliski, Kevin Matzen, and Johannes Kopf. 2020. Consistent video depth estimation. *ACM Trans. Graph*. 39, 4, Article 71 (August 2020), https://doi.org/10.1145/3386569.3392377

3. T. Caselitz, B. Steder, M. Ruhnke and W. Burgard, "Monocular camera localization in 3D LiDAR maps," 2016 *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2016, pp. 1926-1931, DOI: https://doi.org/10.1109/IROS.2016.7759304
4. C. Campos, R. Elvira, J. J. G. Rodríguez, J. M. M. Montiel and J. D. Tardós, "ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual–Inertial, and Multimap SLAM," in *IEEE Transactions on Robotics*, vol. 37, no. 6, pp. 1874-1890, Dec. 2021, https://doi.org/10.1109/TRO.2021.3075644
5. Teed Zachary, Lipson Lahav, Deng Jia, "Deep Patch Visual Odometry". arXiv e-print, 2022, DOI: https://doi.org/10.48550/arXiv.2208.04726
6. Lahav Lipson, Zachary Teed, Jia Deng, "RAFT-Stereo: Multilevel Recurrent Field Transforms for Stereo Matching", arXiv e-print, 2021, https://doi.org/10.48550/arXiv.2109.07547
7. Jiankun Li, Peisen Wang, Pengfei Xiong, Tao Cai, Ziwei Yan, Lei Yang, Jiangyu Liu, Haoqiang Fan, Shuaicheng Liu, "Practical Stereo Matching via Cascaded Recurrent Network with Adaptive Correlation", arXiv e-print, 2022, https://doi.org/10.48550/arXiv.2203.11483
8. Richard Hartley, Andrew Zisserman, "Multiple View Geometry in Computer Vision", 2nd Edition, Cambridge University Press, 2003.
9. Weirong Chen, Suryansh Kumar, Fisher Yu, "Uncertainty-Driven Dense Two-View Structure from Motion", arXiv e-print , 2023, https://doi.org/10.48550/arXiv.2302.00523
10. Denys Hrulov (2024) *Analysis of Three-dimensional Scenes based on Video flow data* (master diploma) V. N. Karazin Kharkiv National University

RECONSTRUCTION OF THREE-DIMENSIONAL SCENES BASED ON VIDEO FLOW DATA

Denys Hrulov[1], Master's student; e-mail: xa11800855@student.karazin.ua;
ORCID: https://orcid.org/0009-0005-8506-770X
Anastasiia Morozova[1], Assistant Professor, PhD; e-mail: a.morozova@karazin.ua;
ORCID: https://orcid.org/0000-0003-2143-7992
Petro Dolia[1], Assistant Professor, PhD; e-mail: pdolya@karazin.ua;
ORCID: https://orcid.org/0009-0002-4062-4443
Liliia Bielova[1], Senior Lecturer; e-mail: l.belova@karazin.ua;
ORCID: https://orcid.org/0009-0007-0805-4547
[1] *V. N. Karazin Kharkiv National University, Ukraine*
Manuscript was received March 23, 2024; Received after review April 29, 2024; Accepted May 30, 2024

Abstract. This work is dedicated to the application of modern algorithms for reconstructing spatial scenes from images to restore spatial information from video. The work is looking at a variety of modern methods, approaches, algorithms and trends in the field. The attention was paid to the sequence of development of approaches to the completion of the task. While researching the field and results related to three-dimensional reconstruction based on images and video streams, an algorithm was invented that allows constructing dense depth maps using information from all video frames. The idea is to use ready-made, commonly accepted, and tested solutions to solve two problems: COLMAP for visual odometry, and RAFT for computing optical flow. The algorithm shows quite accurate results and reconstructs the depth map in detail on arbitrary static scenes.

Keywords: *video flow, 3D reconstruction, machine learning, odometry, neural network, computer vision, depth map, optical flow*

Conflicts of Interest: the authors declare no conflict of interest.