

ТЕХНОЛОГІЇ ФОРМУВАННЯ 3D-ВІДЕО: ПРИНЦИПИ СТЕРЕОСКОПІЇ, ЗАПИС ТА КОНВЕРТАЦІЯ 2D-КОНТЕНТУ

Вінницькій національний технічний університет

Анотація

Розглянуто основні принципи формування 3D-відео, що базуються на особливостях бінокулярного зору людини та явищі стереоскопії. Описано практичні способи отримання 3D-зображення, зокрема нативну стереозйомку, багаторакурсні системи, світлопольові підходи та використання активних сенсорів глибини. Узагальнено сучасні методи конвертації 2D-контенту у стереоскопічний формат, від класичних алгоритмів комп'ютерного зору до глибоких нейронних мереж і нейронного рендерингу.

Ключові слова: 3D-відео, стереоскопія, бінокулярний зір, диспаратність, стереозйомка, конвертація 2D у 3D, глибина сцени, нейронний рендеринг, VR.

Abstract

The basic principles of 3D video formation, based on the features of human binocular vision and the phenomenon of stereoscopy, are considered. Practical methods of obtaining 3D images are described, including native stereo capture, multi-view systems, light-field approaches, and active depth sensing. Modern methods of converting 2D content into stereoscopic format are summarized, ranging from classical computer vision algorithms to deep neural networks and neural rendering.

Keywords: 3D video, stereoscopy, binocular vision, disparity, stereo capture, 2D to 3D conversion, depth estimation, neural rendering, VR.

Вступ

Технології 3D-відео в сучасній практиці ґрунтуються на поданні різних зображень лівому та правому оку. Така пара зображень називається стереопарою. Саме різниця між двома ракурсами дає мозку можливість оцінювати відносну глибину об'єктів і формувати просторове сприйняття сцени. У технічному сенсі якість 3D-ефекту залежить від того, наскільки правильно система відтворює геометрію зйомки, мінімізує вертикальні розбіжності між лівим і правим кадром та утримує диспаратність у межах, комфортних для глядача.

Фізіологічною основою стереоскопії є бінокулярний зір людини. Кожне око спостерігає сцену з трохи іншої точки, тому на сітківці формуються неідентичні зображення. Мозок аналізує цю різницю і використовує її для оцінки просторового положення предметів [1]. У 3D-відео головним керованим параметром є паралакс, тобто горизонтальний зсув відповідних точок у лівому і правому зображеннях. Саме він визначає, чи буде об'єкт сприйматися перед екраном, на його площині або за нею.

Для коректного формування стереоефекту важливо, щоб відповідні точки лівого і правого кадру лежали майже на одних і тих самих рядках. Це означає, що вертикальний паралакс має бути практично нульовим. Якщо ця умова порушується через неточне калібрування камер або неправильну побудову рига, у глядача зростає навантаження на зір і знижується природність сприйняття. Тому в реальних системах перевагу зазвичай надають схемі з паралельними осями камер, а потрібну площину нульового паралаксу встановлюють уже під час постобробки шляхом горизонтального зсуву зображень.

Окремою проблемою стереоскопічного показу є конфлікт вергенції та акомодатії. У природному зоровому сприйнятті очі зводяться і фокусуються на одному й тому самому об'єкті. У випадку зі стереодисплеєм фокус ока залишається на площині екрана, а зведення очей змінюється відповідно до уявної глибини сцени. Через це надмірна диспаратність або її різкі зміни можуть викликати втому, дискомфорт і навіть головний біль. Саме тому під час створення 3D-контенту дуже важливо контролювати діапазон глибин і плавність їх зміни.

Сприйняття глибини формується не лише за рахунок бінокулярних, а й завдяки монокулярним підказкам. До них належать оклюзії, перспектива, освітлення, тіні, градієнти текстури та руховий паралакс. У якісному 3D-відео ці сигнали мають узгоджуватися між собою. Якщо стереопара побудована правильно, але освітлення або оклюзії суперечать створеній глибині, сцена виглядає штучно. Особливо важливе це для систем віртуальної реальності, де під час повороту голови користувач очікує правдоподібної зміни ракурсу та, по можливості, появи наближеного ефекту 6DoF.

Результати дослідження

Практичні методи отримання 3D-відео можна поділити на кілька основних груп. Найпоширенішим способом є нативна стереозйомка двома камерами, які синхронно записують лівий і правий відеопотоки. Такий підхід забезпечує прямий контроль над стереобазою і добре підходить для динамічних сцен, але вимагає точної механічної юстировки, калібрування та узгодження кольору, експозиції і геометрії між каналами.

Для крупних планів і складних сцен часто застосовують дзеркальні beam-splitter риги. У них одна камера працює через напівпрозоре дзеркало, а друга знімає відбиття. Це дає змогу зменшити відстань між оптичними центрами камер і знімати об'єкти зблизька без надмірного паралаксу. Водночас така схема ускладнює оптичне налаштування і може створювати додаткові проблеми через різні оптичні тракти, втрату світла та відмінності в яскравості або кольорі.

Ще один підхід пов'язаний із використанням багаторакурсних систем, тобто масивів камер. У таких системах багато камер одночасно захоплюють сцену з різних точок, що особливо важливо для 360°-відео і VR. Після цього відбувається складний етап стічінгу, калібрування та узгодження стереопари на сфері [2]. Перевагою є високе охоплення сцени та можливість формувати омнінаправне стерео, однак ці системи потребують значних обчислювальних ресурсів і точного контролю паралаксу між сусідніми камерами.

Окремий клас становлять світлопольові та плентоптичні системи, які фіксують не лише яскравість світла, а й напрям променів. Це дозволяє після зйомки змінювати фокус, оцінювати глибину та синтезувати нові ракурси. Такі системи перспективні для побудови більш гнучких 3D-представлень сцени, але вони мають компроміс між просторовою та кутовою роздільністю і створюють великі обсяги даних.

Важливе місце в сучасних системах займають активні сенсори глибини, зокрема ToF та LiDAR. Вони безпосередньо вимірюють відстань до об'єктів і формують карту глибини, яку потім поєднують із кольоровим зображенням. Такі RGB-D підходи добре підходять для AR, робототехніки, сканування простору та технічних застосувань. Їхньою перевагою є наявність метричної глибини, але якість може знижуватися на прозорих, блискучих або дуже темних поверхнях.

Окрім захоплення з нуля, важливим напрямом є конвертація звичайного 2D-контенту у стереоскопічний формат. Цей процес включає оцінювання глибини сцени, побудову карти диспаратності, синтез другого ракурсу, заповнення ділянок, які відкриваються після зміни точки спостереження, і подальше пакування матеріалу у потрібний формат [3]. На ранніх етапах для цього використовувалися класичні методи комп'ютерного зору, зокрема Structure-from-Motion, аналіз оптичного потоку та Depth image-based rendering (DIBR). Рендеринг на основі зображення глибини зображено на рис. 1. Такі підходи спираються на геометрію сцени і дають інтерпретовані результати, але часто погано працюють на слабо текстурованих ділянках, при складних оклюзіях або швидкому русі.

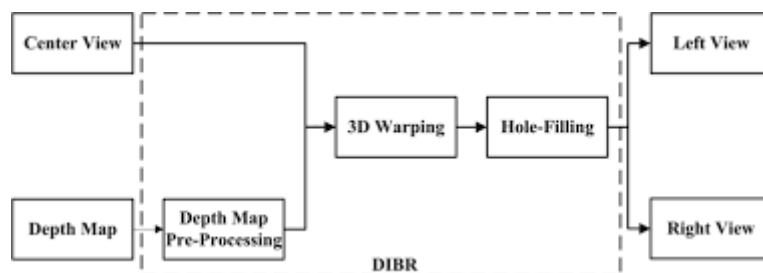


Рис. 1. Рендеринг на основі зображення глибини

Сучасний етап розвитку 2D у 3D конверсії пов'язаний із широким застосуванням глибинних неймереж [4]. Моделі монокулярної оцінки глибини дозволяють отримати карту простору навіть з одного зображення або відеопотоку. Далі ці дані використовуються для синтезу другого ока, створення диспаратності та підготовки стереопари. Нові foundation-моделі глибини і дифузійні підходи покращують деталізацію та узагальнюваність, однак залишаються проблеми часової стабільності. Якщо глибина змінюється від кадру до кадру непослідовно, у відео виникає неприємна пульсація простору.

Перспективним напрямом став нейронний рендеринг, зокрема підходи на основі NeRF і 3D Gaussian Splatting [5]. Вони дозволяють не просто будувати карту глибини, а відтворювати сцену як тривимірне

подання, з якого можна синтезувати нові ракурси. Це відкриває можливості для більш природного паралаксу, покращення VR-досвіду та підвищення якості конвертації, особливо у випадках, коли необхідно отримати не лише стереопару, а й зміну точки огляду.

Незважаючи на значний прогрес автоматичних методів, у професійному кіно та преміальному постпродакшені важливу роль і далі відіграє ручна корекція. До неї належать ротоскопінг, поділ сцени на шари глибини, точкове виправлення меж об'єктів, корекція stereo window і depth grading. Причина в тому, що саме на межах оклюзій, біля волосся, напівпрозорих поверхонь та тонких контурів автоматичні системи найчастіше помиляються. Без ручного контролю навіть технічно правильна конверсія може виглядати нестабільною або неприродною.

Для запису, передавання і зберігання 3D-відео використовуються різні формати. Найпростішими є способи frame-compatible пакування, коли ліве і праве зображення об'єднуються в один кадр. До них належать формати Side-by-Side і Top-Bottom. Вони забезпечують сумісність із традиційною інфраструктурою, але зменшують роздільність для кожного ока. Повнороздільні підходи, наприклад frame packing або MVC, дають кращу якість, однак вимагають більшої пропускну здатності та складніших засобів декодування. Для 360°-відео та VR застосовуються спеціалізовані формати омнінапрямного медіа, орієнтовані на потокове передавання та адаптацію до області, на яку дивиться користувач.

Якість 3D-відео визначається не лише чіткістю чи рівнем шуму. Дуже важливим є комфорт перегляду. Основними джерелами проблем виступають надмірна диспаратність, конфлікт вергенції та акомодатії, вертикальні геометричні невідповідності, крос-ток або ghosting, а також часові розсинхронізації між лівим і правим каналами. Якщо один із цих чинників виходить за межі норми, глядач швидко відчуває втому, а об'ємність сцени починає сприйматися як штучна.

Сфери застосування 3D-відео постійно розширюються. У кіноіндустрії такі технології використовуються для посилення ефекту присутності та видовищності. У VR і 360°-відео вони забезпечують більш переконливе занурення в середовище. В AR системах карти глибини потрібні для правильного перекривання віртуальних і реальних об'єктів. У медицині 3D-візуалізація допомагає хірургам краще оцінювати просторове розташування інструментів і тканин. У робототехніці та промислових системах RGB-D рішення полегшують навігацію, інспекцію та аналіз простору.

Висновки

Отже, технології формування 3D-відео об'єднують знання з фізіології зору, геометрії зйомки, комп'ютерного зору, машинного навчання та мультимедійних форматів. Основою стереоскопії є контрольована подача різних зображень кожному оку, але для якісного результату цього недостатньо. Необхідно також забезпечити правильну геометрію стереопари, узгодженість підказок глибини, комфортний рівень диспаратності та стабільність зображення в часі. Сучасні технології дедалі активніше поєднують нативну стереозйомку, сенсори глибини, автоматичну оцінку простору та нейронний рендеринг. Це робить 3D-відео більш доступним і гнучким, але водночас підвищує вимоги до якості обробки та контролю артефактів.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Lin C. J., Canny S. Effects of virtual target size, position, and parallax on vergence-accommodation conflict as estimated by actual gaze // Scientific Reports. – 2022. – Vol. 12. – Article 20100. – DOI: 10.1038/s41598-022-24450-9.
2. Hannuksela M. M., Wang Y.-K. An Overview of Omnidirectional Media Format (OMAF) // Proceedings of the IEEE. – 2021. – Vol. 109, No. 9. – P. 1590–1606. – DOI: 10.1109/JPROC.2021.3063544.
3. Hachaj T. Adaptable 2D to 3D Stereo Vision Image Conversion Based on a Deep Convolutional Neural Network and Fast Inpaint Algorithm // Entropy. – 2023. – Vol. 25, No. 8. – Article 1212. – DOI: 10.3390/e25081212.
4. Yang L., Kang B., Huang Z., Xu X., Feng J., Zhao H. Depth Anything: Unleashing the Power of Large-Scale Unlabeled Data // 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). – Seattle, WA, USA, 17–21 June 2024. – P. 10371–10381. – DOI: 10.1109/CVPR52733.2024.00987.
5. Kerbl B., Kopanas G., Leimkühler T., Drettakis G. 3D Gaussian Splatting for Real-Time Radiance Field Rendering // arXiv. – 2023. – DOI: 10.48550/ARXIV.2308.04079.

Дудукало Нікіта Сергійович — студент групи ЗПІ-226, факультет інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, м. Вінниця, e-mail: nikitadudukalo2@gmail.com.

Dudukalo Nikyta S. — student of the Faculty of Information Technology and Computer Engineering, Vinnytsia National Technical University, Vinnytsia, e-mail: nikitadudukalo2@gmail.com.