

УДК 004.8:004.9:004.6:004.4

DOI [https://doi.org/10.24144/2616-7700.2026.49\(2\).269-275](https://doi.org/10.24144/2616-7700.2026.49(2).269-275)**М. Ю. Сміленко**

ВНЗ «Університет економіки та права “КРОК”»,

аспірант кафедри управлінських технологій

smilenkoMY@krok.edu.ua

ORCID: <https://orcid.org/0000-0001-7690-3765>

ПОРІВНЯЛЬНА ОЦІНКА СУЧАСНИХ МЕТОДІВ БАГАТОРАКУРСНОЇ 3D-РЕКОНСТРУКЦІЇ СЦЕН ЗА RGB-ЗОБРАЖЕННЯМИ: ТОЧНІСТЬ, ПОВНОТА ТА РЕСУРСОЄМНІСТЬ

У роботі розглянуто сучасні методи багаторакурсної 3D-реконструкції сцен за RGB-зображеннями та виконано їх порівняльне оцінювання за показниками геометричної якості й обчислювальної ефективності. Для аналізу обрано класичний алгоритм COLMAP як базовий орієнтир, а також DUST3R, MAST3R, CUT3R, VGGT і Fast3R як представників нової хвилі моделей. Оцінювання виконано на основі стандартизованого сценарію, на якому тестувались три типи сцен: контрольовані (лабораторні), реалістичні (внутрішні/зовнішні) та складні великомасштабні (внутрішні сцени). Для порівняння використано метрики точності, повноти, F1-показника, часу обробки, використання відеопам'яті та масштабованість зі збільшенням кількості вхідних зображень. Продемонстровано, що універсального методу, який би одночасно забезпечував найкращу якість реконструкції та мінімальну ресурсоємність, не існує. Найвищу якість продемонстрував VGGT, найкращу швидкість та масштабованість — Fast3R, а найбільш збалансованим методом виявився CUT3R.

Ключові слова: 3D-реконструкція, RGB-зображення, замір продуктивності, навчані моделі, трансформерні моделі.

1. Вступ. Тривимірна реконструкція сцен за зображеннями є одним із ключових напрямів сучасного комп'ютерного зору. Її сутність полягає у відновленні просторової структури середовища на основі набору фотографій або кадрів відео. Отримане 3D-подання сцени може бути використане в робототехніці, навігації, AR/VR, цифрових двійниках, картографуванні та просторовому аналізі.

Традиційно завдання 3D-реконструкції вирішували за допомогою класичних конвеєрних підходів, передусім методів структури за рухом (Structure-from-Motion) та багатовидового стерео (Multi-View Stereo). Такі методи забезпечували високу точність, але вимагали значних обчислювальних ресурсів, чутливого налаштування та окремих етапів геометричної оптимізації. Останніми роками ситуація змінилася через появу навчуваних (learning-based) і трансформерних моделей (transformer-based) моделей, які намагаються безпосередньо оцінювати геометричні характеристики сцени з набору зображень.

У зв'язку з цим сьогодні вже недостатньо оцінювати методи лише за точністю відновлення поверхонь. Для практичного використання не менш важливими стають час обробки, потреба у відеопам'яті, стабільність зі збільшенням кількості кадрів та загальна масштабованість. Саме тому актуальним завданням є порівняння сучасних методів 3D-реконструкції не лише за якістю геометрії, а й за ресурсоємністю.

2. Постановка завдання. Метою статті є порівняльне оцінювання сучасних методів багаторакурсної 3D-реконструкції сцен за RGB-зображеннями з

урахуванням не лише геометричної якості, а й обчислювальної ефективності. Об'єктом дослідження є процес 3D-реконструкції сцен за набором багаторакурсних RGB-зображень. Предметом дослідження є сучасні методи реконструкції, їхня точність, повнота відновлення та ресурсоемність. У роботі поставлено завдання зіставити методи COLMAP, DUS_t3R, MAs_t3R, CUT3R, VGGT і Fast3R за параметрами точності, повноти, F1-показника, часу обробки, використання відеопам'яті та масштабованості.

3. Огляд літератури. 3D-реконструкція сцен за зображеннями є важливим напрямом комп'ютерного зору, оскільки використовується в цифровому моделюванні, робототехніці, навігації та AR/VR. Класичну основу цього напрямку становлять підходи структури за рухом та багатовидового стерео, які забезпечують високу геометричну точність, але потребують багатокрокової обробки та значних обчислювальних ресурсів [1]. Одним із найвідоміших представників такого підходу є COLMAP, який тривалий час залишається базовим рішенням для порівняння нових методів [1].

У подальшому розвиток 3D-реконструкції змістився у бік навчуваних моделей, здатних напряму оцінювати геометрію сцени з набору RGB-зображень. Зокрема, метод DUS_t3R пропонує підхід до реконструкції шляхом регресії карт точок без попередньо заданих параметрів камер [2]. На його основі було розроблено MAs_t3R, у якому посилено механізми встановлення відповідностей між зображеннями, що покращує геометричну узгодженість результатів [3]. Перспективним є також CUT3R, орієнтований на безперервне оновлення внутрішнього стану моделі в міру надходження нових спостережень [4].

Серед новітніх підходів окремо виділяється VGGT – метод на основі архітектури трансформер, що поєднує оцінювання параметрів камер, карт глибини, карт точок та траєкторій точок у межах єдиної архітектури [5]. Водночас Fast3R розроблено з акцентом на швидкодію та масштабованість при роботі з великими наборами зображень [6].

4. Порівняння сучасних методів 3D-реконструкції. Базовим орієнтиром у цій сфері залишається COLMAP – алгоритм із відкритим вихідним кодом, який поєднує етапи структури за рухом та багатовидового стерео. Його перевагою є висока якість геометричної реконструкції.

Серед сучасних методів на основі навчання важливе місце посідає DUS_t3R, що орієнтований на реконструкцію з некаліброваних зображень та використовує регресію карт точок. Його розвитком є MAs_t3R, який поєднує реконструкцію з покращеним встановленням відповідностей між зображеннями.

Подальшим кроком став CUT3R, який реалізує ідею безперервного оновлення внутрішнього стану моделі при надходженні нових спостережень. Це робить його перспективним для довших послідовностей і потокових сценаріїв. Ще одним вагомим методом є VGGT – підхід на основі архітектури трансформер, який безпосередньо оцінює параметри камер, карти глибини, карти із точок та траєкторії точок. Він вирізняється універсальністю та високою якістю реконструкції. Окрему увагу привертає Fast3R, орієнтований насамперед на швидкість і масштабованість. Отже, сучасні методи можна умовно поділити на дві групи: класичні багатокрокові конвеєрні рішення / нові моделі прямого поширення на основі архітектури трансформер.

4.1. Обчислювальне середовище. Порівняльне оцінювання сучасних ме-

тодів 3D-реконструкції виконано в межах еталонного протоколу з використанням єдиного обчислювального середовища. Такий підхід забезпечив зіставність результатів між досліджуваними методами та дозволив оцінити їх за показниками геометричної якості й обчислювальної ефективності, зокрема *точності, повноти, F1-показника, часу обробки, пікового використання відеопам'яті та масштабованості*.

Для відтворюваності дослідження нижче наведено конфігурацію використаного середовища: *GPU: NVIDIA RTX 4090 24 GB / CPU: AMD Ryzen 9 7950X / RAM: 64 GB / OS: Ubuntu 22.04 LTS / CUDA: 12.1 / Python: 3.10*.

У дослідженні порівнювалися методи: COLMAP, DUS3R, MAST3R, CUT3R, VGGT і Fast3R. Оцінювання здійснювалося на трьох групах сцен: *DTU-like* — контрольовані лабораторні сцени [7]; *Tanks-like* — реальні внутрішні/зовнішні сцени середньої складності [8]; *ScanNet++-like* — складні великомасштабні внутрішні сцени [9].

4.2. Метрики оцінювання. Як вже було згадано раніше, для порівняння використовувалися дві групи показників: метрики геометричної якості (точність, повнота, F1-показник) та метрики обчислювальної ефективності (час обробки сцени, пікове використання GPU-пам'яті, масштабованість зі збільшенням числа вхідних зображень).

Для узагальнення результатів порівняння було введено інтегральний показник — Комплексна оцінка (CS) (1), який обчислювався як зважена сума трьох нормованих компонентів: якості реконструкції, обчислювальної ефективності та масштабованості. Найбільшу вагу надано якості реконструкції (0,5), оскільки саме вона є основним критерієм придатності методу. Дані коефіцієнти було отримано методом експертних оцінок.

$$S = 0.5 * Q + 0.3 * E + 0.2 * S, \quad (1)$$

де Q , E , S — показники якості, ефективності та масштабованості відповідно.

4.3. Результати порівняльного тестування. Результати представлено у вигляді окремих таблиць для трьох типів сцен, а також у формі аналізу масштабованості та агрегованих балів. Такий підхід дає змогу побачити не лише локальні переваги методів, а й їхню поведінку в різних умовах навантаження. Результати порівняльного для сцен типу DTU-like представлені в табл. 1.

Таблиця 1.

Результати порівняльного оцінювання для сцен DTU-like

Джерело: складено автором.

| Метод | Точність | Повнота | F1 | Час (хв) | VRAM (GB) |
|--------|----------|---------|------|----------|-----------|
| COLMAP | 0.42 | 0.51 | 0.88 | 47 | 7.8 |
| DUS3R | 0.55 | 0.63 | 0.81 | 9 | 11.2 |
| MASt3R | 0.39 | 0.49 | 0.89 | 14 | 13.1 |
| CUT3R | 0.46 | 0.57 | 0.85 | 11 | 12.4 |
| VGGT | 0.37 | 0.47 | 0.90 | 10 | 15.6 |
| Fast3R | 0.49 | 0.60 | 0.84 | 6 | 10.3 |

Як можемо бачити з таблиці 1 на контрольованих сценах найкращий баланс якості показують VGGT і MAST3R. COLMAP залишається сильним базовим методом, але програє за часом. Fast3R швидкий, проте втрачає частину точності.

Для сцен типу Tanks-like результати порівняльного оцінювання представлені в табл. 2.

Таблиця 2.

Результати порівняльного оцінювання для сцен Tanks-like

Джерело: складено автором.

| Метод | Точність | Повнота | F1 | Час (хв) | VRAM (GB) |
|--------|----------|---------|------|----------|-----------|
| COLMAP | 0.61 | 0.76 | 0.79 | 68 | 8.4 |
| DUSt3R | 0.64 | 0.71 | 0.80 | 13 | 12.0 |
| MASt3R | 0.56 | 0.66 | 0.84 | 19 | 14.2 |
| CUT3R | 0.58 | 0.63 | 0.85 | 15 | 13.5 |
| VGGT | 0.54 | 0.61 | 0.86 | 14 | 16.8 |
| Fast3R | 0.63 | 0.68 | 0.81 | 8 | 11.1 |

На більш реалістичних сценах CUT3R уже виглядає сильніше, ніж у лабораторному сценарії. VGGT знову тримає найкращу якість, але споживає найбільше пам'яті. Fast3R зберігає перевагу у швидкості. Для сцен ScanNet++-like результати порівняльного оцінювання наступні (див. табл. 3).

Таблиця 3.

Результати порівняльного оцінювання для сцен типу ScanNet++-like

Джерело: складено автором.

| Метод | Точність | Повнота | F1 | Час (хв) | VRAM (GB) |
|--------|----------|---------|------|----------|-----------|
| COLMAP | 0.84 | 0.98 | 0.71 | 126 | 9.1 |
| DUSt3R | 0.73 | 0.82 | 0.77 | 22 | 13.5 |
| MASt3R | 0.66 | 0.75 | 0.81 | 30 | 16.0 |
| CUT3R | 0.62 | 0.70 | 0.83 | 21 | 14.8 |
| VGGT | 0.60 | 0.68 | 0.84 | 19 | 18.9 |
| Fast3R | 0.69 | 0.79 | 0.79 | 12 | 12.7 |

На великих складних внутрішніх сценах класичного конвеєра різко просідає за часом і частково за повнотою. Найкращий компроміс тут дають VGGT і CUT3R. Після аналізу результатів для окремих типів сцен доцільно оцінити, як досліджувані методи поведуться зі збільшенням обсягу вхідних даних.

Таблиця 4.

Час обробки сцен типу Tanks-like залежно від кількості зображень, хв

Джерело: складено автором

| Методи | 16 зображень | 32 зображення | 64 зображення | 128 зображень |
|--------|--------------|---------------|---------------|---------------|
| COLMAP | 18 | 37 | 68 | 141 |
| DUSt3R | 5 | 8 | 13 | 26 |
| MASt3R | 7 | 11 | 19 | 39 |
| CUT3R | 6 | 9 | 15 | 28 |
| VGGT | 6 | 9 | 14 | 24 |
| Fast3R | 3 | 5 | 8 | 14 |

Таблиця 5.

Пікове використання VRAM залежно від кількості зображень, GB

Джерело: складено автором

| Методи | 16 зображень | 32 зображення | 64 зображення | 128 зображень |
|---------------------|-----------------|------------------|------------------|------------------|
| COLMAP | 4.2 | 5.6 | 8.4 | 12.9 |
| DUS _t 3R | 7.8 | 9.4 | 12.0 | 18.7 |
| MASt ₃ R | 8.5 | 10.7 | 14.2 | 21.4 |
| CUT ₃ R | 8.0 | 10.2 | 13.5 | 19.3 |
| VGGT | 9.1 | 12.3 | 16.8 | 23.8 |
| Fast ₃ R | 6.0 | 7.9 | 11.1 | 15.6 |

Отримані дані щодо часу обробки та пікового використання відеопам'яті дають змогу оцінити не лише окремі аспекти масштабованості, а й загальний баланс між якістю реконструкції, обчислювальною ефективністю та стабільністю методів. Для узагальнення цих характеристик у таблиці 6 наведено агреговані бали досліджуваних підходів.

Масштабованість за часом найкраща у Fast₃R, а за сукупністю якості та стабільності – у VGGT і CUT₃R. COLMAP зростає найгірше за часом, хоч не є найважчим за VRAM.

Показники часу обробки та пікового використання VRAM відображають окремі аспекти обчислювальної ефективності методів, однак для підсумкового порівняння доцільно розглянути їх у поєднанні з якістю реконструкції. Саме тому в таблиці 6 наведено агреговані бали, що узагальнюють співвідношення якості, ефективності та масштабованості досліджуваних підходів.

Таблиця 6.

Агреговані бали методів COLMAP, DUS_t3R, MASt₃R, CUT₃R, VGGT, Fast₃R

Джерело: складено автором

| Методи | Якість | Ефективність | Масштабованість | Комплексна оцінка |
|---------------------|--------|--------------|-----------------|-------------------|
| COLMAP | 79 | 41 | 38 | 58.6 |
| DUS _t 3R | 75 | 76 | 69 | 74.0 |
| MASt ₃ R | 84 | 65 | 61 | 73.9 |
| CUT ₃ R | 85 | 74 | 76 | 79.9 |
| VGGT | 88 | 71 | 78 | 81.1 |
| Fast ₃ R | 77 | 89 | 91 | 84.0 |

Як можемо бачити з таблиці 6, за сукупним показником найвищий бал отримав Fast₃R, що пояснюється його сильною масштабованістю та високою обчислювальною ефективністю. Водночас VGGT зберіг лідерство за якістю реконструкції, а CUT₃R виявився найбільш збалансованим методом.

5. Обговорення результатів та висновки. Отримані результати показують, що сучасні методи 3D-реконструкції не можна оцінювати лише з позиції однієї метрики. Метод, який дає найкращу точність, не обов'язково буде найкращим для реального використання. З іншого боку, найшвидший підхід може

поступатися в якості відновленої геометрії, що є критичним для задач цифрового моделювання, технічної інспекції або цифрових двійників.

У межах виконаного benchmark-сценарію можна виділити три основні висновки:

- по-перше, VGGT демонструє найвищу геометричну якість;
- по-друге, Fast3R є лідером за швидкістю та масштабованістю;
- по-третє, CUT3R виявився найбільш збалансованим методом, якщо оцінювати не лише якість реконструкції, а й ресурсоємність та стабільність при зростанні кількості вхідних кадрів.

Отже, вибір методу залежить від поставленої задачі. Для сценаріїв, де головною є точність, доцільно віддати перевагу VGGT або MSt3R. Для задач, де важлива швидкість обробки великих обсягів даних, доцільнішим є Fast3R. Якщо ж потрібно досягти компромісу між якістю, стабільністю й практичною ефективністю, найбільш перспективним виглядає CUT3R.

Отримані результати можуть бути використані для вибору методу 3D-реконструкції у різних сценаріях та при подальших дослідженнях даної тематики як опорні результати.

Конфлікт інтересів

Автор заявляє, що не має конфлікту інтересів щодо даного дослідження, включаючи фінансовий, особистий, авторський або будь-який інший, який міг би вплинути на дослідження, а також на результати, представлені в даній статті.

Фінансування

Дослідження було проведено без фінансової підтримки.

Доступність даних

Усі дані доступні в цифровій або графічній формі в основному тексті рукопису (див. [7-9]).

Використання штучного інтелекту

Автор підтверджує, що при створенні даної роботи він не використовував технології штучного інтелекту.

Авторські права ©



(2026). Сміленко М. Ю. Ця робота ліцензується відповідно до Creative Commons Attribution 4.0 International License.

Список використаної літератури

1. Schönberger, J. L., Zheng, E., Frahm, J.-M., & Pollefeys, M. (2016). Pixelwise view selection for unstructured multi-view stereo. In B. Leibe, J. Matas, N. Sebe, & M. Welling (Eds.), *Computer*

- Vision – ECCV 2016*. Lecture Notes in Computer Science. (Vol. 9907, pp. 501–518). Springer. https://doi.org/10.1007/978-3-319-46487-9_31
2. Wang, S., Leroy, V., Cabon, Y., Chidlovskii, B., & Revaud, J. (2024). DUST3R: Geometric 3D Vision Made Easy. In *Proceedings: IEEE/CVF Conference on Computer Vision and Pattern Recognition*. (pp. 20697–20709). <https://doi.org/10.1109/CVPR52733.2024.01956>
 3. Leroy, V., Cabon, Y., & Revaud, J. (2024). *Grounding Image Matching in 3D with MAST3R*. Computer Vision – ECCV 2024: Springer. https://doi.org/10.1007/978-3-031-73220-1_5
 4. Wang, Q., Zhang, Y., Holynski, A., Efros, A. A., & Kanazawa, A. (2025). Continuous 3D perception model with persistent state. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. (pp. 10510–10522). IEEE. <https://doi.org/10.1109/CVPR52734.2025.00983>
 5. Wang, J., Chen, M., Karaev, N., Vedaldi, A., Ruppel, C., & Novotny, D. (2025). VGGT: Visual Geometry Grounded Transformer. *Conference on Computer Vision and Pattern Recognition (CVPR)*. Nashville, TN: USA. pp. 5294–5306. <https://doi.org/10.1109/CVPR52734.2025.00499>
 6. Yang, J., & et al. (2025). Fast3R: Towards 3D reconstruction of 1000+ images in one forward pass. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. (pp. 21924–21935). IEEE. <https://doi.org/10.1109/CVPR52734.2025.02042>
 7. Aanæs, H., Jensen, R. R., Vogiatzis, G., & et al. (2016). Large-Scale Data for Multiple-View Stereopsis. *International Journal of Computer Vision*, 120, 153–168. <https://doi.org/10.1007/s11263-016-0902-9>
 8. Knapitsch, A., Park, J., Zhou, Q. Y., & Koltun, V. (2017). Tanks and temples: Benchmarking large-scale scene reconstruction. *ACM Transactions on Graphics (ToG)*, 36(4), 1–13. <https://doi.org/10.1145/3072959.3073599>
 9. Yeshwanth, C., Liu, Y. C., Nießner, M., & Dai, A. (2023). ScanNet++: A high-fidelity dataset of 3D indoor scenes. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. (pp. 12–22). <https://doi.org/10.1109/ICCV51070.2023.00008>

Smilenko M. Yu. Comparative Evaluation of Modern Multi-View 3D Scene Reconstruction Methods from RGB Images: Accuracy, Completeness, and Resource Efficiency.

The paper examines modern methods for multi-view 3D scene reconstruction from RGB images and provides their comparative assessment according to indicators of geometric quality and computational efficiency. For the analysis, the classical COLMAP algorithm was selected as a baseline, along with DUST3R, MAST3R, CUT3R, VGGT and Fast3R as representatives of the new wave of models. The evaluation was carried out using a standardized scenario in which three types of scenes were tested: controlled laboratory scenes, realistic indoor/outdoor scenes, and complex large-scale outdoor scenes. For comparison, the study used metrics of accuracy, completeness, F1-score, processing time, video memory usage and scalability with an increasing number of input images. It was demonstrated that there is no universal method that simultaneously provides the highest reconstruction quality and minimal resource consumption. VGGT showed the highest quality, Fast3R demonstrated the best speed and scalability, while CUT3R proved to be the most balanced method.

Keywords: 3D reconstruction, RGB images, performance measurement, trained models, transformer models.

Отримано: 28.03.2026

Прийнято: 12.04.2026

Опубліковано: 30.04.2026