Telecommunications and Computer Engineering, TCSET 2022, pp. 134–138.

ОПТИМІЗАЦІЯ МЕТОДІВ ТРИВИМІРНОГО РОЗПІЗНАВАННЯ ОБ'ЄКТІВ У РОБОТОТЕХНІЧНИХ СИСТЕМАХ З ВИКОРИСТАННЯМ НЕЙРОННИХ МЕРЕЖ

М. А. Чернюк, Д. О. Багаєв, М. О. Юріна, к.т.н. проф. М. М. Колендовська,
ХНУРЕ, м. Харків

Реконструкція тривимірних моделей — ключовий етап у роботизованому технічному зорі. Вона дозволяє будувати карту середовища, визначати форму об'єктів і їхню просторову позицію. Точність цих моделей безпосередньо впливає на успіх автономної навігації та маніпуляцій.

Попри значний прогрес, проблема балансу між точністю реконструкції та обчислювальною швидкодією залишається актуальною, особливо для мобільних і вбудованих роботів.

Structure from Motion (SfM). Ґрунтується на аналізі послідовності двовимірних зображень для відновлення просторових координат. Має високу точність при добрій текстурі поверхонь, але чутлива до зміни освітлення та не підходить для реального часу.

Multi-View Stereo (MVS). Використовує кілька поглядів сцени для отримання щільної хмари точок. Дає детальні результати, але вимагає значних обчислень. Підходить для офлайн-реконструкції у лабораторних умовах.

Depth Estimation. Застосовує стереокамери або глибинні сенсори (RGB-D). Має високу швидкість, однак точність обмежена роздільною здатністю сенсорів і якістю калібрування.

Neural Radiance Fields (NeRF). Використовує нейронні мережі для відтворення сцени як неперервного поля світлового випромінювання. Дає реалістичну геометрію й освітлення, проте поки що надто повільний для реального часу.

Для порівняння методів використано три критерії.

Середня похибка реконструкції (RMSE) — точність у визначенні просторових координат. Середній час обробки кадру (t_proc) — продуктивність системи. Стійкість до шумів (SNR-loss) — ефективність при неповних або спотворених даних. Тестування проводилося на наборах даних TUM RGB-D та ETH3D, а також у симуляційному середовищі Gazebo.

Отримані результати показують, що класичні методи (SfM, MVS) залишаються точними, але не відповідають вимогам реального часу. Гібридні підходи, які комбінують дані глибинних сенсорів із нейронним моделюванням, забезпечують оптимальний компроміс.

Таблиця 1.

Метод	RMSE (м)	t_proc (мс/кадр)	Стійкість до шумів	Придатність для реального часу
SfM	0.018	480	Середня	Низька
MVS	0.012	950	Висока	Низька
Depth Estimation	0.025	40	Середня	Висока
NeRF	0.009	1500	Висока	Низька
Гібридний NeRF+Depth	0.014	110	Висока	Середня

Аналіз підтверджує тенденцію переходу від чисто геометричних до нейронно-аналітичних методів. Використання гібридних моделей дозволяє:

– адаптуватися до різних типів сенсорів;
– зменшити похибки при низькій текстурі;
– підвищити стабільність до динамічних змін сцени.

Основною проблемою залишається потреба в оптимізації обчислень і пам’яті. Використання спрощених нейронних архітектур і GPU-прискорення є ключем до впровадження таких систем у мобільних роботах.

Проведений аналіз доводить, що жоден метод реконструкції не є універсальним, однак комбіновані підходи на основі глибинних сенсорів і нейронних моделей (NeRF+Depth) дають найкраще співвідношення точності та швидкості. Це відкриває перспективи для створення адаптивних 3D-зорових систем, здатних працювати в реальному часі без втрати деталізації.

Для подальшої роботи розглянемо методи розпізнавання об’єктів у тривимірному просторі, що є ключовим елементом сучасних робототехнічних систем, зокрема автономних роботів, безпілотних літальних апаратів та промислових маніпуляторів.

Традиційні методи двовимірного розпізнавання мають суттєві обмеження при роботі з глибокими сценами та складними текстурами. Перехід до 3D-моделей відкриває можливість точнішої просторової інтерпретації, однак вимагає значних обчислювальних ресурсів та адаптованих архітектур нейронних мереж.

Серед популярних підходів до 3D-розпізнавання можна виділити три групи:

1. Voxel-based моделі (наприклад, VoxNet, 3D-CNN).
2. Point-based архітектури (PointNet, PointNet++).
3. Projection-based підходи (MVCNN, 3D ShapeNets).

Аналіз показує, що найкращі результати забезпечує комбінація згорткових і точкових мереж, яка дозволяє збалансувати точність і швидкість. У дослідженні розроблено гібридну архітектуру, що поєднує 2D-CNN для попереднього вилучення ознак з глибинних карт і PointNet для аналізу хмар точок. Експериментальне тестування проводилося на наборах даних ModelNet40 та ShapeNetCore. Розроблена модель показала середню точність 92,4%, перевищивши класичний PointNet на 13,7%.

Оптимізація гібридної архітектури дозволила:

1. Зменшити кількість параметрів на 28%,

2. Підвищити швидкість розпізнавання на 21%,
3. Підвищити стійкість моделі до шумів і часткової відсутності даних.

Результати свідчать, що поєднання CNN і PointNet забезпечує ефективну інтеграцію локальних і глобальних просторових ознак. Це відкриває перспективи використання таких систем у мобільній робототехніці, де важлива висока швидкість реакції.

Запропонований підхід демонструє ефективність використання гібридних нейронних архітектур для тривимірного розпізнавання об'єктів. Подальші дослідження можуть бути спрямовані на оптимізацію структури мережі для роботи на вбудованих системах і використання методів трансферного навчання для адаптації до нових середовищ.

ІНТЕГРАЦІЯ 3D-ВІЗУАЛІЗАЦІЇ У СИСТЕМИ КОМП'ЮТЕРНОГО ЗОРУ ДЛЯ АВТОНОМНИХ РОБОТІВ

М. А. Чернюк, Д. О. Багаєв, М. О. Юріна, к.т.н. проф. М. М. Колендовська,
ХНУРЕ, м. Харків

Сучасні робототехнічні системи значною мірою покладаються на зорові сенсори — стереокамери, лідари, оптичні далекоміри. Проте навіть найкращі сенсорні системи не гарантують високу точність, якщо не враховано фізичні властивості оптики: спотворення, заломлення світла, втрата контрасту.

Автономні роботи дедалі частіше застосовуються у виробництві, логістиці та побуті. Їхня ефективність безпосередньо залежить від здатності орієнтуватися у просторі. Класичні 2D-методи комп'ютерного зору не завжди забезпечують достатню точність при роботі в умовах змінного освітлення, перекриттів або складних поверхонь.

Інтеграція тривимірної візуалізації дозволяє створити просторово точну модель навколишнього середовища, на основі якої робот може приймати рішення у реальному часі. Це поєднує переваги симуляції, геометричного моделювання та штучного інтелекту.

Тривимірне моделювання дає змогу симулювати поведінку світлових променів, що проходять крізь оптичну систему, та аналізувати їхній вплив на формування зображення. Це відкриває нові можливості для калібрування роботизованих камер і створення синтетичних наборів даних для навчання нейронних мереж.

Сучасні системи автономної навігації використовують кілька ключових технологій побудови тривимірних сцен:

LIDAR-сканування, яке формує точні хмари точок, проте вимагає складної фільтрації шумів.

Стереозір, що дозволяє відновити глибину за різницею між двома зображеннями, але чутливий до освітлення.

Глибинні камери (наприклад, Intel RealSense), які забезпечують баланс між точністю та швидкістю.