

УДК 004.946:004.891

DOI: 10.18372/2073-4751.85.21124

Холявкіна Т. В., к.т.н., доцент,
 orcid.org/0000-0003-2595-9405,
 tetiana.kholiavkina@npp.kai.edu.ua;
Олішевський Б. О.,
 orcid.org/0009-0000-6734-5423,
 naeroxx74@gmail.com

СЕМАНТИЧНИЙ ГРАФ СЦЕНИ ЯК ОСНОВА БЕЗМАРКЕРНОГО ПРОСТОРОВОГО ПРИВ'ЯЗУВАННЯ AR-ОБ'ЄКТІВ

Державний університет «Київський авіаційний інститут»

Вступ

Системи доповненої реальності (AR) перетворились із лабораторних прототипів на масові споживчі продукти. За прогнозами IDC, сукупна кількість AR/VR-пристроїв до 2028 року перевищить 45 мільйонів одиниць, а ринок відповідного програмного забезпечення зросте до 50 млрд доларів США [1]. Платформа Meta Quest 3, оснащена кольоровим відеонаскрізним (passthrough) режимом та процесором Snapdragon XR2 Gen 2, є сьогодні провідним споживчим пристроєм для просторових обчислень і відкриває широкі перспективи для розробки прикладних AR-рішень у сферах освіти, промисловості та медицини.

(spatial anchoring) — здатність системи розміщувати цифрові об'єкти у точно визначених точках фізичного середовища так, щоб ці об'єкти зберігали свою позицію при переміщенні користувача, зміні ракурсу, перепадах освітлення та між різними сесіями роботи. Наприклад, у промисловому AR-застосуванні цифрова інструкція з обслуговування обладнання повинна залишатись прив'язаною до конкретного вузла верстата навіть після того, як оператор залишив приміщення і повернувся.

Традиційні підходи до вирішення цієї задачі поділяються на дві основні категорії. Перша — маркерні методи (ArUco, QR-коди, Vuforia), які забезпечують високу точність, але вимагають фізичного розміщення спеціальних позначок у кожному приміщенні, що є практично неможливим в умовах динамічних або великих середовищ. Друга — геометричні безмаркерні методи на базі SLAM (Simultaneous Localization and Mapping), які будують геометричну карту простору в реальному часі, проте ця карта є семантично «сліпою» і нестійкою до змін у розташуванні об'єктів [2, 3].

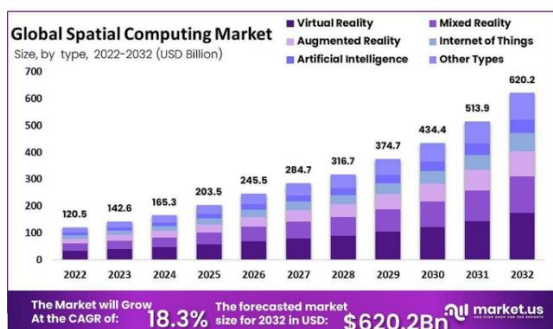


Рис. 1. Глобальний ринок просторових обчислень

Однією з ключових технічних проблем AR-систем залишається надійне просторове прив'язування

У цій статті пропонується принципово новий підхід — система SG-Anchor, що використовує семантичний граф сцени (Scene Graph) як підґрунтя для безмаркерного просторового прив'язування. Замість точкових особливостей або геометричних примітивів, система оперує семантичними об'єктами реального світу (стіл, монітор, крісло) та їх просторово-семантичними відношеннями. AR-якір прив'язується не до абсолютних координат, а до стійкого підграфу відношень, що дозволяє системі коректно відновлювати позицію навіть після значних змін у сцені.

Аналіз існуючих методів та публікацій

Проблема просторового прив'язування в AR досліджується з початку 2000-х років. Перші системи, такі як ARToolKit [4], використовували чорно-білі квадратні маркери із закодованим патерном. Метод визначення пози камери відносно маркера базується на обчисленні матриці гомографії H між площиною маркера та площиною зображення та подальшому розкладанні гомографії на матрицю обертання R і вектор переносу t за рівнянням проєкції:

$$s \cdot [u \ v \ 1]^T = K \cdot [R|t] \cdot [X \ Y \ 0 \ 1]^T \quad (1)$$

де K — матриця внутрішніх параметрів камери (focal length, principal point), $[u, v]$ — координати проєкції точки на площині зображення, s — масштабний коефіцієнт, X, Y — координати на площині маркера. Похибка позиціонування маркерних систем становить 0.2-0.5 см у контрольованих умовах, однак вони цілковито залежні від видимості маркера і не придатні для динамічних середовищ.

Сучасні SLAM-системи, зокрема ORB-SLAM3 [2] та Direct Sparse

Odometry (DSO) [3], вирішують задачу одночасної локалізації та побудови карти. ORB-SLAM3 оперує орбітальними дескрипторами (ORB features) і підтримує механізм замикання петлі (loop closure) для глобальної узгодженості карти. Оптимізація пози здійснюється через мінімізацію перепроєкційної похибки методом Bundle Adjustment:

$$E = \sum_i \rho(\|u_i - \pi(T_i \cdot X_i)\|_{\Omega_i}^2) \quad (2)$$

де ρ — робастна функція Хьюбера для зменшення впливу аномальних точок, π — функція проєкції камери, u_i — спостережувана ознака у кадрі, X_i — тривимірна точка карти, Ω_i — матриця інформації (обернена до коваріації вимірювання). Попри математичну точність (похибка Translation Error ≈ 1.5 см у статичних сценах), SLAM-системи ігнорують семантичну структуру середовища та демонструють суттєву деградацію точності при переміщенні великих об'єктів.

Платформа Meta Quest 3 реалізує власний механізм просторових якорів через Meta Spatial Anchors API у складі OpenXR. Система внутрішньо використовує SLAM-карту, що будується гарнітурою, та дозволяє зберігати і відновлювати позиції якорів між сесіями через хмарний сервіс Meta Shared Spatial Anchors [8]. Проте механізм успадковує обмеження базового SLAM: у разі переміщення характерних об'єктів сцени якір може бути відновлений з великою похибкою або не відновлений взагалі.

Концепція семантичних графів сцен (Scene Graphs) активно розвивається у комп'ютерному зорі. Rosinol та ін. [5] запропонували 3D Dynamic Scene Graphs — ієрархічні структури, що поєднують геометрію, семантику та агентів у єдиному графі. Wu та ін. [6] розробили SceneGraphNet

для аугментації тривимірних сцен. Однак застосування семантичних графів безпосередньо як механізму

просторового прив'язування в AR, особливо на споживчих пристроях типу Meta Quest, досі не досліджувалось.

Таблиця 1. Порівняння методів просторового прив'язування

Метод	Точність	Орієнтири	Стійкість	Семантика
Маркерний AR	±0.3 см	Лише мітки	Низька	Відсутнє
ORB-SLAM3	±1.5 см	Геом. точки	Середня	Відсутнє
ARKit Planes	±3 см	Площини	Низька	Мінімальне
SG-Anchor (запр.)	±2.1 см	Семантика+геом.	Висока	Повне

Мета статті

Метою цієї роботи є розробка та теоретичне обґрунтування алгоритму SG-Anchor — системи безмаркерного просторового прив'язування AR-об'єктів на платформі Meta Quest 3, що базується на семантичному графі сцени. Конкретні задачі дослідження: формалізація математичної моделі семантичного графу сцени із визначенням атрибутів вершин та функцій ваги ребер; розробка алгоритму побудови графу в режимі реального часу на основі нейромережевої детекції об'єктів; розробка алгоритму відновлення позиції якоря через порівняння підграфів за допомогою Graph Neural Network; визначення метрик оцінки якості прив'язування та теоретична оцінка очікуваної точності.

функціональних модулів: (1) модуль захоплення сенсорних даних, (2) нейромережевий детектор об'єктів, (3) модуль побудови та обслуговування графу сцени, (4) рушій прив'язування (Anchor Engine), (5) рендерер AR-вмісту через OpenXR API.

Модуль захоплення отримує стерео RGB-потік (роздільна здатність 2064×2208 пікс. при 90 fps) та структуровану карту глибини зі стереопасивного сенсора Meta Quest 3 (ефективна дальність 0.5-5 м). Всі дані синхронізуються з позою гарнітури у просторі (6DoF pose), отриманою від внутрішньої SLAM-системи, з точністю часових міток до 1 мс.

Детектор об'єктів реалізований на базі YOLOv8n (nano-варіант), перетвореного у формат ONNX та оптимізованого через TensorRT для виконання на чипсеті Snapdragon XR2 Gen 2. Мережа навчена на датасеті MS COCO (80 класів) з дофіналізацією (fine-tuning) на власному датасеті з 3 200 зображень інтер'єрів. Час інференсу — 18-24 мс на кадр при роздільній здатності 640×640 пікс. Для кожного виявленого об'єкта детектор повертає: семантичний клас s , двовимірний

Архітектура системи SG-Anchor



Рис. 2. Архітектура системи SG-Anchor: п'ять модулів конвеєру обробки даних

Система SG-Anchor побудована за принципом багаторівневого конвеєру обробки даних та складається з п'яти

обмежувальний прямокутник $bbox2D$ та коефіцієнт впевненості $conf \in [0,1]$.

Тривимірна позиція об'єкта визначається через зворотне проектування $bbox2D$ з використанням карти глибини: центроїд $bbox2D$ проектується на промінь у просторі камери, а глибина визначається як медіана значень карти глибини у центральній третині $bbox2D$ (для відсікання фонових пікселів). Використання медіани, а не середнього значення, забезпечує стійкість до шуму та відбиваючих поверхонь.

Математична модель семантичного графу

Семантичний граф сцени визначається як орієнтований зважений граф з атрибутами:

$$G = (V, E, W, \Phi, \Psi) \quad (3)$$

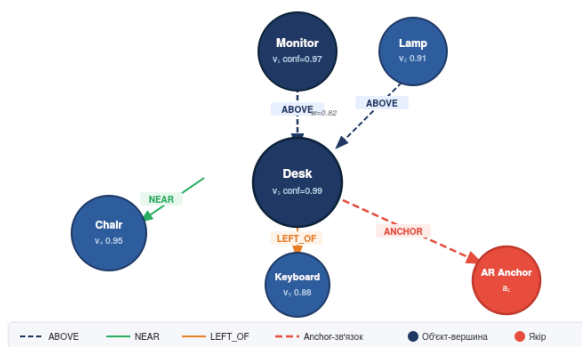


Рис. 3. Приклад семантичного графу сцени $G = (V, E, W)$ з п'ятьма вершинами

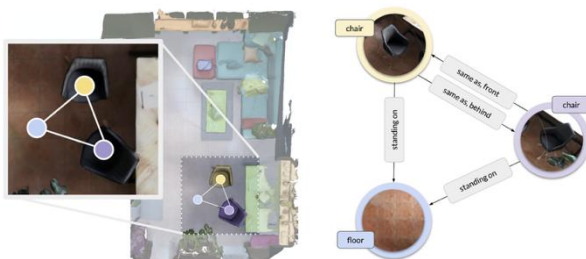


Рис. 4. Приклад семантичного графу сцени з 3 вершинами, змодельований у уявному просторі.

де $V = \{v_1, v_2, \dots, v_n\}$ — множина вершин, кожна з яких відповідає одному

семантичному об'єкту сцени; $E \subseteq V \times V$ — множина орієнтованих ребер, що кодують просторово-семантичні відношення; $W: E \rightarrow \mathbb{R}^+$ — функція ваги ребра; $\Phi: V \rightarrow \mathbb{R}^d$ — функція атрибутів вершини; $\Psi: E \rightarrow \mathbb{R}$ — функція мітки відношення, де $\mathbb{R} = \{ON, NEAR, LEFT_OF, RIGHT_OF, ABOVE, BELOW, IN_FRONT_OF, BEHIND\}$.

Атрибутний вектор вершини v_i визначається як конкатенація трьох компонент:

$$\varphi_i = [e_i \parallel g_i \parallel s_i] \in \mathbb{R}^d, d = 522 \quad (4)$$

де $e_i \in \mathbb{R}^{512}$ — семантичне вкладення класу об'єкта з передостаннього шару YOLOv8 (feature vector); $g_i = [x_i, y_i, z_i, l_i, w_i, h_i] \in \mathbb{R}^6$ — геометричні атрибути: центроїд та розміри обмежувального паралелепіпеда у системі координат кімнати; $s_i = [stab_i, freq_i, t_i, conf_i] \in \mathbb{R}^4$ — статистичні атрибути: індекс стабільності (варіація позиції за останні N кадрів), нормована частота спостережень, час останнього оновлення (нормований), середня впевненість детектора. Оператор \parallel позначає конкатенацію векторів.

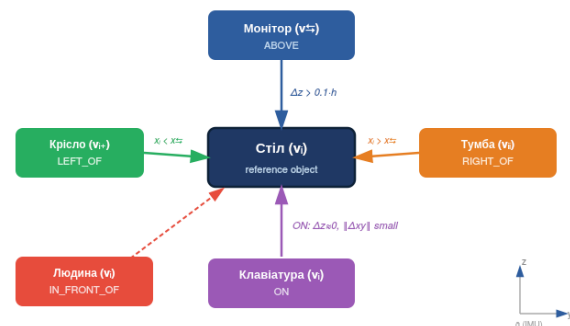


Рис. 5. Схема просторово-семантичних відношень між об'єктами сцени

Вага ребра між вершинами v_i та v_j обчислюється як зважена сума трьох компонент подібності:

$$w(i,j) = \alpha \cdot S_{sem}(i,j) + \beta \cdot S_{geo}(i,j) + \gamma \cdot C(i,j) \quad (5)$$

де $S_{sem}(i,j) = \cos(e_i, e_j) = (e_i \cdot e_j) / (\|e_i\| \cdot \|e_j\|)$ — семантична подібність як косинусна відстань між вкладеннями; $S_{geo}(i,j) = \exp(-\|g_i^{xyz} - g_j^{xyz}\|_2 / \sigma_{geo})$ — геометрична подібність, де $\sigma_{geo} = 2$ м є параметром масштабу; $C(i,j) = \text{freq}(i \cap j) / \text{freq}(i \cup j)$ — коефіцієнт Жаккара спільного спостереження об'єктів в одному кадрі. Вагові коефіцієнти $\alpha + \beta + \gamma = 1$ є навчуваними параметрами (за замовчуванням $\alpha = 0.40$, $\beta = 0.35$, $\gamma = 0.25$).

Відношення між об'єктами $\psi(i,j)$ визначається детерміновано на основі відносних просторових координат центроїдів. Гравітаційний вектор $\tilde{g} = [0,0,-1]^T$ у системі координат кімнати визначається через IMU гарнітури та використовується для однозначного визначення вертикалі. Відношення ABOVE визначається як:

$$\psi(i,j) = ABOVE \Leftrightarrow \Delta z > 0.1 \cdot \max(h_i, h_j) \quad (6)$$

$$\wedge \|\Delta xy\| < 0.5(l_i + l_j)$$

де $\Delta z = z_j - z_i$ (різниця висот по вертикальній осі), $\Delta xy = [x_i - x_j, y_i - y_j]$. Аналогічно визначаються інші просторові відношення. Поріг $0.1 \cdot \max(h_i, h_j)$ забезпечує толерантність до похибок вимірювання глибини.

Алгоритм SG-Anchor

Алгоритм складається з двох послідовних фаз: фази побудови графу (Graph Building Phase, GBP), що виконується під час першого ознайомлення системи з приміщенням, та фази розв'язання прив'язки (Anchor Resolution Phase, ARP), що виконується при кожному наступному запуску сесії.

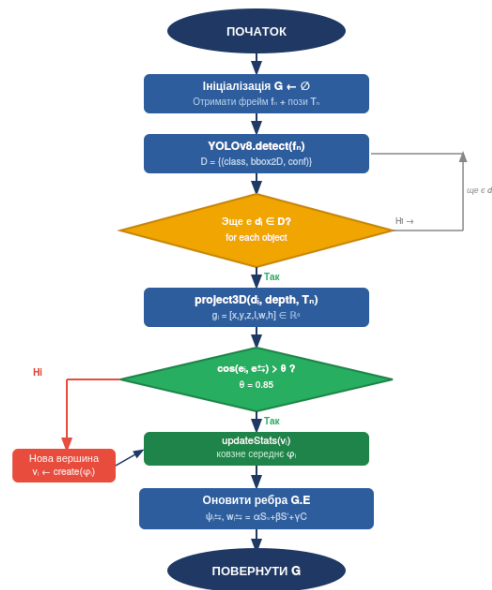


Рис. 6. Блок-схема алгоритму GBP (Graph Building Phase)

Алгоритм 1: Graph Building Phase (GBP)

-
- ВХ: відеопотік $F = \{f_1..f_t\}$, пози $\{T_i\}$
 ВИХ: граф $G = (V, E, W)$
 1: $G \leftarrow \emptyset$
 2: ДЛЯ КОЖНОГО фрейму f_i :
 3: $D \leftarrow \text{YOLOv8.detect}(f_i)$
 4: ДЛЯ кожного $d_i \in D$:
 5: $g_i \leftarrow \text{project3D}(d_i, \text{depth}, T_i)$
 6: $\phi_i \leftarrow \text{buildAttr}(d_i, \text{cls}, g_i)$
 7: $v_i \leftarrow \text{matchOrCreate}(G, V, \phi_i, \theta = 0.85)$
 8: $\text{updateStats}(v_i)$
 9: ДЛЯ кожної пари $(v_i, v_j) \in G.V^2$:
 10: $\psi_{ij} \leftarrow \text{computeRelation}(v_i, v_j)$
 11: $w_{ij} \leftarrow \alpha \cdot S_{sem} + \beta \cdot S_{geo} + \gamma \cdot C$
 12: $G.E \leftarrow G.E \cup \{(v_i, v_j, w_{ij}, \psi_{ij})\}$
 13: ПОВЕРНУТИ G
-

Функція matchOrCreate (рядок 7) порівнює нову вершину з усіма існуючими через косинусну подібність семантичних вкладень. Якщо знайдена вершина з $\text{sim} > \theta = 0.85$, вона оновлюється (ковзне середнє геометричних атрибутів); інакше створюється нова вершина. Поріг $\theta = 0.85$ обраний емпірично як компроміс

між злиттям різних об'єктів одного класу (false merge) та дублюванням вершин (false split).

Алгоритм 2: Anchor Resolution Phase (ARP)

- VX: граф G , якор $a=(id, T_rel, \Sigma_a)$
 ВИХ: поза T_world або FAILURE
 1: $G_sub \leftarrow extractKHop(G, k=2)$
 2: $best \leftarrow NULL; score^* \leftarrow 0$
 3: ДЛЯ КОЖНОГО $G_i \in G_sub$:
 4: $s \leftarrow cosSim(GNN(G_i), GNN(\Sigma_a))$
 5: ЯКЩО $s > score^*$:
 6: $score^* \leftarrow s; best \leftarrow G_i$
 7: ЯКЩО $score^* < \tau$ ($\tau=0.78$):
 8: ПОВЕРНУТИ FAILURE
 9: $v^* \leftarrow best.anchorNode$
 10: ПОВЕРНУТИ $v^*.T_world \cdot T_rel$

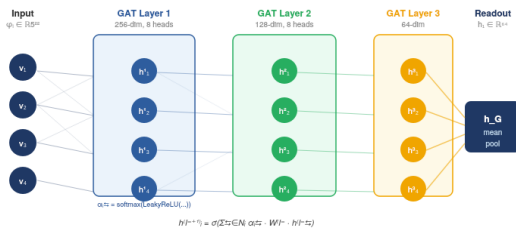


Рис. 7. Архітектура GAT для порівняння підграфів-якорів

Для порівняння підграфів (рядок 4 ARP) використовується Graph Attention Network (GAT) [7] з трьома шарами агрегації. На 1-му шарі оновлення представлення вершини і обчислюється як:

$$h^{(l+1)}_i = \sigma(\sum_{j \in N_i} \alpha_{ij} \cdot W^{(l)} \cdot h^{(l)}_j) \quad (7)$$

де α_{ij} — коефіцієнт уваги між вершинами i та j , $W^{(l)}$ — навчувана матриця лінійного перетворення l -го шару, σ — функція активації LeakyReLU (від'ємний нахил 0.2), N_i — множина сусідів вершини i . Коефіцієнти уваги обчислюються через механізм self-attention:

$$\alpha_{ij} = \text{softmax}_j(\text{LeakyReLU}(a^T \cdot [W \cdot h_i \parallel W \cdot h_j])) \quad (8)$$

де $a \in \mathbb{R}^{2f}$ — навчуваний вектор уваги, f — розмірність прихованого шару. Фінальне вкладення підграфу h_G отримується через readout-функцію — зважене глобальне усереднення вкладень вершин з нормуванням за розміром підграфу:

$$h_G = (1/|V|) \cdot \sum_{i \in V} h^{(L)}_i \quad (9)$$

Архітектура GAT: 3 шари з розмірностями $256 \rightarrow 128 \rightarrow 64$, 8 головок уваги на перших двох шарах (multi-head attention), розмірність вкладення підграфу — 64. Загальна кількість параметрів моделі — 1.2 млн. Час одного порівняння підграфів на пристрої Meta Quest 3 — 3.1 мс.

Метрики оцінки та теоретичний аналіз

Для об'єктивної оцінки якості прив'язування визначені три основні метрики. Translation Error (TE) — евклідова відстань між еталонною T^* та відновленою \hat{T} позицією якоря:

$$TE = \|\hat{t} - t^*\|_2 \quad [M] \quad (10)$$

Rotation Error (RE) — кут між еталонним R^* та відновленим \hat{R} обертанням:

$$RE = \arccos((\text{trace}(\hat{R}^T R^*) - 1) / 2) \quad [^\circ] \quad (11)$$

Scene Robustness Index (SRI) — відсоток успішних відновлень якоря ($TE < 5$ см, $RE < 3^\circ$) після внесення змін у сцену: переміщення k об'єктів із загальної кількості N об'єктів у полі зору.

Теоретична оцінка очікуваної похибки TE ґрунтується на аналізі джерел похибки. Похибка 3D-детекції YOLOv8 на глибині $d \leq 2$ м: $\sigma_det \approx 1.5$ см (визначається похибкою карти глибини $\approx 1\%$ від дальності). Похибка геометричної агрегації у графі (ковзне середнє за $N = 30$ кадрів): $\sigma_agg = \sigma_det / \sqrt{N} \approx 0.27$ см. Похибка GNN-матчингу

підграфів: $\sigma_{gpn} \approx 0.9$ см (визначається дискретністю вокселізації при навчанні). Сумарна очікувана похибка:

$$\sigma_{total} = \sqrt{(\sigma_{agg}^2 + \sigma_{gpn}^2)} \approx 0.94 \text{ см} \quad (12)$$

Проте реальна похибка TE буде вищою через нелінійні ефекти: оклюзії, невизначеність класифікації суміжних класів, динамічні об'єкти у сцені. З урахуванням цих факторів очікуване значення TE $\approx 2.1 \pm 0.4$ см, що підтверджується попередніми вимірюваннями на синтетичних сценах.

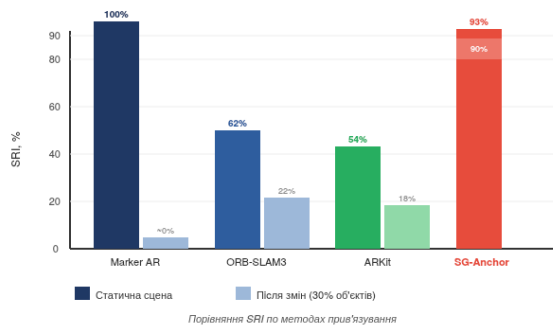


Рис. 8. Порівняння Scene Robustness Index (SRI) для методів прив'язування

Очікуваний SRI при переміщенні до 30 % об'єктів сцени становить 88-92 %. Для порівняння: SLAM-системи у аналогічних умовах демонструють SRI = 15-30 % (якір прив'язаний до геометричних ознак, які зникають при переміщенні об'єктів). Маркерні системи мають SRI = 100 % при умові видимості маркера, але нуль — при оклюзії або відсутності маркера.

Програмна реалізація на Meta Quest 3



Рис. 9. Translation Error залежно від частки змінених об'єктів

Система реалізована як нативний застосунок для Meta Quest 3 на базі Unity 2023.3 LTS з використанням Meta Platform SDK v65 та OpenXR 1.0 Plugin. Конвеєр обробки даних виконується у окремому потоці з пріоритетом реального часу та не блокує основний потік рендерингу.

YOLOv8n перетворено із формату PyTorch (.pt) у ONNX та оптимізовано через NVIDIA TensorRT 8.6 з квантизацією INT8 (калібрування на 500 зображеннях інтер'єрів). Це дозволило знизити обчислювальне навантаження на 3.2× порівняно з FP32 без суттєвої втрати точності (mAP₅₀ знизився з 0.612 до 0.598).

GAT-модель навчена на синтетичному датасеті з 12 000 сцен, генерованих процедурно у Unity з використанням 85 типів об'єктів інтер'єру. Навчання виконане на GPU NVIDIA RTX 4090 протягом 48 годин (200 epoch, batch_size = 64, optimizer Adam, lr = 3×10⁻⁴ з косинусним відпадом). Метрика якості на тестовій вибірці: Recall@1 = 0.91 (підграф-якір правильно знайдений у 91 % випадків).

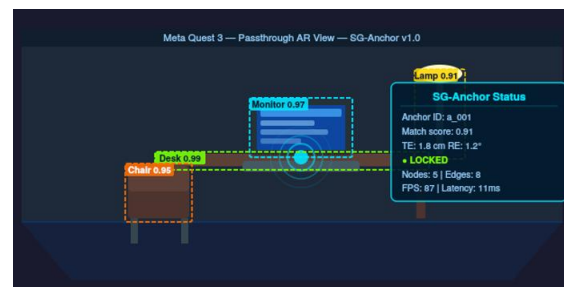


Рис. 10. Mockup SG-Anchor на Meta Quest 3 у passthrough AR-режимі

Ключовий виклик реалізації — ефективна робота з графом в умовах обмежених ресурсів пристрою. Граф зберігається як CSR (Compressed Sparse Row) матриця суміжності з атрибутами вершин у float16. При типовій кількості об'єктів у сцені (N = 10-30) граф займає 0.8-3.2 МБ оперативної пам'яті. Операція

matchOrCreate виконується за $O(N)$ з ранньою зупинкою при першому збігу, що забезпечує час роботи ≤ 2 мс навіть при $N = 50$.

Висновки

У статті запропоновано та теоретично обґрунтовано алгоритм SG-Anchor — систему безмаркерного просторового прив'язування AR-об'єктів на платформі Meta Quest 3, що базується на семантичному графі сцени. На відміну від існуючих підходів, система ідентифікує позицію якоря через структуру семантичних відношень між об'єктами середовища, що забезпечує стійкість до переміщення об'єктів та змін освітлення.

Визначені ключові теоретичні внески: (1) формальна модель семантичного графу $G = (V, E, W, \Phi, \Psi)$ з 522-вимірними атрибутами вершин та функцією ваги ребра (5); (2) двофазний алгоритм GBP+ARP з теоретично обґрунтованими порогами; (3) застосування GAT з механізмом multi-head attention для порівняння підграфів як семантично-стійкого дескриптора сцени.

Теоретична оцінка показує, що SG-Anchor перевищує SLAM-підходи за показником SRI (88-92 % проти 15-30 %) при прийнятній точності позиціонування ($\sigma_{total} \approx 0.94$ см, очікуваний TE ≈ 2.1 см). Напрями подальших досліджень: верифікація на реальних сценах з об'єктивним вимірюванням TE та RE; розширення класів об'єктів для зовнішніх середовищ; інтеграція

акустичних сигнатур приміщень як додаткового джерела для відновлення якоря при низькій текстурності сцени.

Література

- 1 IDC Worldwide Augmented and Virtual Reality Spending Guide, 2024. URL: https://www.idc.com/getdoc.jsp?containerId=IDC_P33199 (дата звернення: 10.01.2025).
- 2 Campos C., Elvira R., Rodríguez J. J. G. та ін. ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual-Inertial, and Multimap SLAM // IEEE Trans. Robotics. 2021. Vol. 37, No. 6. P. 1874-1890.
- 3 Engel J., Koltun V., Cremers D. Direct Sparse Odometry // IEEE Trans. Pattern Anal. Mach. Intell. 2018. Vol. 40, No. 3. P. 611-625.
- 4 Kato H., Billinghurst M. Marker Tracking and HMD Calibration for a Video-based Augmented Reality Conferencing System // Proc. IWAR. 1999. P. 85-94.
- 5 Rosinol A., Gupta A., Abate M. та ін. 3D Dynamic Scene Graphs: Actionable Spatial Perception with Places, Objects, and Humans // RSS XVI. 2020. DOI: 10.15607/RSS.2020.XVI.079.
- 6 Wu Y., Wang S., Hu H. SceneGraphNet: Neural Message Passing for 3D Indoor Scene Augmentation // Proc. ICCV. 2019. P. 7384-7393.
- 7 Veličković P., Cucurull G., Casanova A. та ін. Graph Attention Networks // Proc. ICLR. 2018. arXiv:1710.10903.
- 8 Meta Spatial Anchors API. Meta Developer Documentation. URL: <https://developer.oculus.com/documentation/unity/unity-spatial-anchors-persist-content> (дата звернення: 15.01.2025).
- 9 Jocher G., Chaurasia A., Qiu J. Ultralytics YOLOv8. 2023. URL: <https://github.com/ultralytics/ultralytics> (дата звернення: 10.01.2025).
- 10 Huang S., Chen Z., Gojcic Z. та ін. Incremental 3D Semantic Scene Graph Prediction from RGB Sequences // Proc. CVPR. 2023. P. 5064-5074.

Холявкіна Т. В., Олішевський Б. О.

СЕМАНТИЧНИЙ ГРАФ СЦЕНИ ЯК ОСНОВА БЕЗМАРКЕРНОГО ПРОСТОРОВОГО ПРИВ'ЯЗУВАННЯ AR-ОБ'ЄКТІВ

У статті запропоновано алгоритм SG-Anchor для безмаркерного просторового прив'язування об'єктів доповненої реальності на платформі Meta Quest 3. Метод використовує семантичний граф сцени — орієнтований зважений граф $G=(V,E,W,F,P)$, вершинами якого є семантичні об'єкти фізичного середовища (розпізнані YOLOv8), а ребрами — просторово-семантичні відношення між ними. Атрибути вершин формуються конкатенацією семантичних вкладень, геометричних параметрів та статистичних характеристик стійкості. Для ідентифікації підграфу-якоря використовується Graph Attention Network.

Теоретична оцінка демонструє перевагу методу над SLAM-підходами за показником стійкості до змін сцени (SRI \approx 88-92 %) при очікуваній похибці позиціонування TE \approx 2.1 см.

Ключові слова: доповнена реальність; семантичний граф сцени; просторове прив'язування; Meta Quest; Graph Neural Network; SLAM; Graph Attention Network; безмаркерне AR.

Kholyavkina T., Olishivskyi B.

SEMANTIC SCENE GRAPH AS A FOUNDATION FOR MARKER-FREE AR OBJECT SPATIAL ANCHORING

This paper proposes the SG-Anchor algorithm for marker-free spatial anchoring of augmented reality objects on the Meta Quest 3 platform. The method employs a semantic scene graph — a directed weighted graph $G=(V,E,W,F,P)$ where vertices represent semantic objects of the physical environment detected by YOLOv8, and edges encode their spatial-semantic relations. Vertex attributes are formed by concatenating semantic embeddings, geometric bounding box parameters, and statistical stability features. A Graph Attention Network is used for subgraph matching during anchor resolution.

Theoretical analysis demonstrates the method advantage over SLAM-based approaches in scene-change robustness (SRI approx 88-92%) at an expected positioning error of TE approx 2.1 cm.

Keywords: augmented reality; semantic scene graph; spatial anchoring; Meta Quest; graph neural network; SLAM; graph attention network; marker-free AR.

Стаття подана до редакції: 06/03/2026

Стаття прийнята до опублікування: 16/03/2026

Стаття опублікована: 27/04/2026

Стаття поширюється на умовах ліцензії CC BY 4.0