

УДК 004.9

DOI: 10.20998/2411-0558.2025.02.05

Д. С. ЖИКУЛ, студент, ОНУ ім. І.І. Мечникова, Одеса,

В. В. МОРОЗ, канд. техн. наук, доц., ОНУ ім. І.І. Мечникова, Одеса

АНАЛІЗ АЛГОРИТМІВ ВІЗУАЛЬНОГО ВИЯВЛЕННЯ МІСЦЬ

Задача візуального виявлення місць (VPR – Visual Place Recognition) є важливою в комп'ютерному зорі та робототехніці. Останніми роками інтерес до цієї теми суттєво зріс, зокрема через її актуальність для задач автономної навігації та локалізації. Алгоритми VPR широко застосовуються в системах Visual SLAM для виявлення замикання петель. Однак ця задача ускладнюється через низку факторів: наявність візуально схожих місць, зміну точки спостереження, варіації освітлення, сезонні й структурні зміни середовища тощо. У сучасних дослідженнях пропонуються різні підходи до розв'язання цієї проблеми, які зазвичай класифікують за типом використовуваних дескрипторів. У цій роботі розглянуто та проаналізовано ці основні групи методів. Іл.: 4. Табл.: 3. Бібліогр.: 16 назв.

Ключові слова: візуальне виявлення місць, автономна навігація, місце, дескриптор.

Постановка проблеми. Нехай Q та DB – не пусті множини, що містять зображення певного середовища. Q – множина запитів, а DB – база даних. Необхідно для кожного зображення $I_j \in Q$ знайти (якщо це можливо) відповідне зображення $I_k \in DB$, яке було зроблено в тій самій локації.

Аналіз останніх досліджень. Основні дослідження в цій галузі пов'язані зі створенням нових дескрипторів для репрезентації зображень місцевості. Ранні методи VPR базувалися на класичних дескрипторах, таких як SIFT [1], SURF [2], ORB [3] та HOG [4]. Головним їхнім недоліком була обмежена інваріантність до сильних змін освітлення, точки спостереження та структури сцени – наслідок ручного конструювання їхніх ознак.

У 2014 році було опубліковано першу роботу, пов'язану з використанням нейронних мереж для цієї задачі [5], та в якій було продемонстровано їх ефективність.

У 2016 було запропоновано NetVLAD [6] – шар агрегації ознак, який можна використовувати з будь-якою згортковою нейронною мережею та в результаті тренувати мережу наскрізно, від вхідного зображення до кінцевого представлення місця.

Усі провідні дескриптори для VPR сьогодні є навчальними й поділяються на дві групи:

- глобальні: NetVLAD, CosPlace [7] та інші;
- локальні: зокрема SuperPoint [8], а також нові DISK [9] і DeDoDe [10], які базуються на CNN архітектурах, та LoFTR [11] – трансформерний метод прямої відповідності без окремого детектора ключових точок.

У 2023 році було продемонстровано [12], що гібридні методи, які поєднують глобальні та локальні дескриптори, значно покращують якість у складних середовищах.

Мета роботи. Пошук ефективного алгоритму VPR на основі аналізу існуючих підходів з локальними і глобальними дескрипторами та їх комбінацій.

Огляд підходів. Загальний підхід, який було застосовано поділяється на наступні кроки (рис. 1):

- 1) Репрезентація зображень на основі дескрипторів (feature extraction).
- 2) Порівняння цих дескрипторів та відповідно оцінка схожості кожної пари зображень (I_Q, I_{DB}) , $I_Q \in Q$, $I_{DB} \in DB$ (descriptor similarity).
- 3) Відбір необхідної кількості кращих кандидатів з DB для кожного $I_Q \in Q$ (matching).

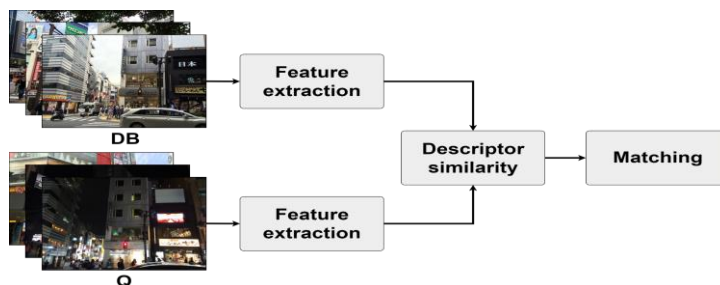


Рис. 1. Загальний підхід

В роботі розглянуті варіації трьох основних підходів:

- 1) На основі Глобальних дескрипторів. Кожному зображенню $I_Q \in Q$ ставимо у відповідність 1 вектор D розміру N . Критерій належності двох зображень одному місцю – косинус подібності відповідних їм векторів.
- 2) На основі Локальних дескрипторів. Кожному зображенню ставимо у відповідність n пар: ключова точка (x, y) та вектор d розміру M . Критерієм належності двох зображень до одного місця є кількість співставлених пар векторів. Далі проходить геометрична верифікація цих співставлень варіаціями алгоритму RANSAC [13]. Верифікація виконується з урахуванням відповідних ключових точок.
- 3) Гібридний підхід з використанням глобальних дескрипторів для формування початкового набору кандидатів (відбираються k найкращих кандидатів для кожного запиту), після чого здійснюється додаткове переранжування за допомогою локальних дескрипторів.

Було використано глобальні дескриптори CosPlace на основі ResNet50 і ResNet152 із розмірами дескриптора 512 і 2048 відповідно, локальні дескриптори DeDoDe та DISK, а також LoFTR – неявний локальний дескриптор (було використано лише для переранжування). Для співставлення локальних дескрипторів було використано MNN, MNN з Lowe ratio = 0.98 (SMNN), а також RANSAC та MAGSAC [14] для геометричної верифікації. Алгоритм геометричної верифікації визначався за найкращими показниками Recall@K.

Дані та методика експерименту. Перші два тести проводились на наборі даних Gardens Point (GP) [15]. У першому тесті (табл. 1) Q складалася з нічного проходу парком правою стороною, а DB – з денного проходу тією ж стороною (рис. 2). У другому тесті (табл. 2) Q залишалася незмінною, а DB містила денний прохід парком лівою стороною (рис. 3). Всі множини містять по 200 зображень та $I_j \in Q$ відповідає локації $I_k \in DB$, коли $|j - k| \leq 1$. Третій тест (табл. 3) проводився на наборі даних SPED [16]. У ньому множини Q та DB містять по 607 зображень різних

місць зі змінами у освітленні та сезони (рис. 4). Для SPED $I_j \in Q$ відповідає локації $I_k \in DB$, коли $j = k$.

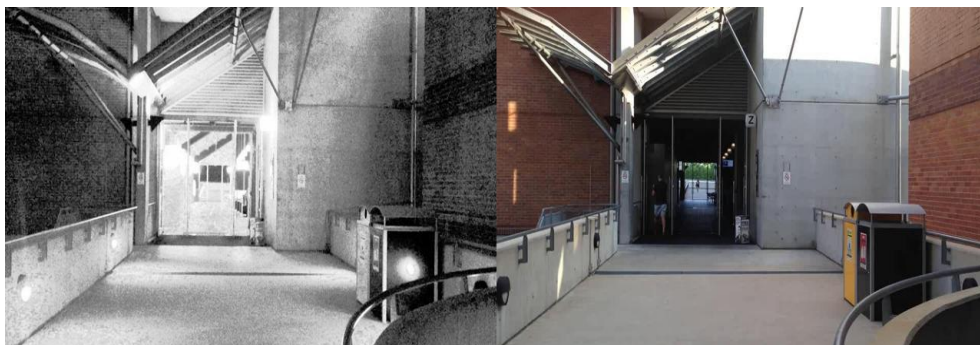


Рис. 2. Приклад правильної пари зображень з Q та DB (Gardens Point №1)

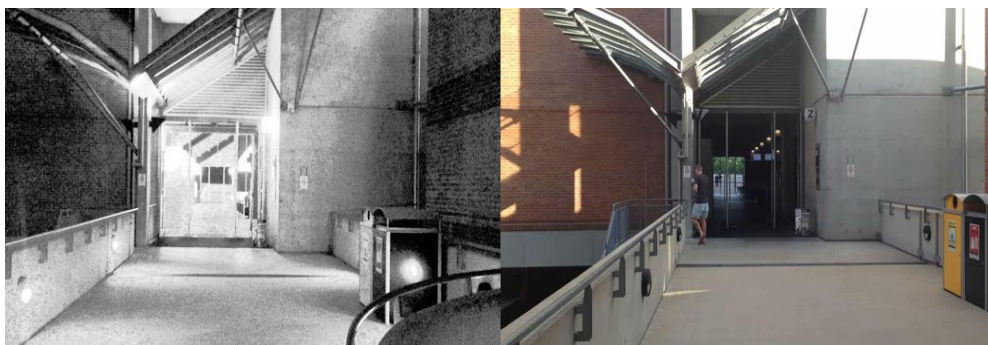


Рис. 3. Приклад правильної пари зображень з Q та DB (Gardens Point №2)



Рис. 4. Приклад правильної пари зображень з Q та DB (SPED)

Для оцінки результатів використовувалась метрика Recall@K для значень K: 1, 5, 10 та час (загальна сума, витрачена на розрахунок дескрипторів для Q та їх зіставлення з дескрипторами з DB).

Для переранжування у гібридному підході на наборі даних GP обиралось 10 найкращих кандидатів, а для SPED – 15.

Результати експериментів

У таблицях наведено результати оцінки різних комбінацій глобальних і локальних дескрипторів, а також методів зіставлення. Показники Recall@1, Recall@5 та Recall@10 відображають точність відповідних підходів, а останній стовпець показує загальний час обробки всіх запитів.

Таблиця 1

Результати оцінки підходів на наборі даних GP №1.

Глоб. дескриптор	Лок. дескриптор	Зіставлення	R@ 1	R@ 5	R@ 10	Час, с.
CosPlace (512)	-	-	79.5	98.5	100	1.15
CosPlace (2048)	-	-	75.5	97	98.5	1.98
CosPlace (512)	LoFTR+MAGSAC		95.5	94.5	99.5	100
CosPlace (2048)	LoFTR+MAGSAC		94	92.5	98.5	98.5
CosPlace (512)	DeDoDe	MNN+MAGSAC	97	100	100	49.2
CosPlace (2048)	DeDoDe	MNN+MAGSAC	95	98.5	98.5	45.4
-	DeDoDe	SMNN+RANSAC	91.5	99	99	363
CosPlace (512)	DISK	MNN+MAGSAC	94.5	100	100	50.4
CosPlace (2048)	DISK	MNN+MAGSAC	93.5	98.5	98.5	50.7
-	DISK	SMNN+MAGSAC	94.5	99/5	99/5	334.2

Таблиця 2

Результати оцінки підходів на наборі даних GP №2.

Глоб. дескриптор	Лок. дескриптор	Зіставлення	R@ 1	R@ 5	R@ 10	Час, с.
CosPlace (512)	-	-	60	94.5	98	1.16
CosPlace (2048)	-	-	57.5	93.5	97	1.98
CosPlace (512)	LoFTR+MAGSAC		74	74	95	98
CosPlace (2048)	LoFTR+MAGSAC		71	76.5	95.5	97
CosPlace (512)	DeDoDe	MNN+MAGSAC	77	97	98	51.3
CosPlace (2048)	DeDoDe	MNN+MAGSAC	76.5	96.5	97	52.1
-	DeDoDe	SMNN+RANSAC	71	95	96	372.6
CosPlace (512)	DISK	MNN+MAGSAC	71	94	98	50.7
CosPlace (2048)	DISK	MNN+MAGSAC	67.5	95.5	97	52
-	DISK	SMNN+MAGSAC	66	92.5	94.5	325.7

Таблиця 3

Результати оцінки підходів на наборі даних SPED.

Глоб. дескриптор	Лок. дескриптор	Зіставлення	R@ 1	R@ 5	R@ 10	Час, с.
CosPlace (512)	-	-	79.1	89.6	93.1	2.81
CosPlace (2048)	-	-	79.2	88.8	91.8	6.03
CosPlace (512)	LoFTR+MAGSAC		89.9	89	92.8	93.1
CosPlace (2048)	LoFTR+MAGSAC		89.9	87.8	91.3	91.8
CosPlace (512)	DeDoDe	MNN+MAGSAC	83.7	92.6	93.1	127.3
CosPlace (2048)	DeDoDe	MNN+MAGSAC	83.2	91.3	91.8	126.4
-	DeDoDe	SMNN+RANSAC	80.4	88.3	91.1	2954
CosPlace (512)	DISK	MNN+MAGSAC	85.7	91.9	93.1	156.9
CosPlace (2048)	DISK	MNN+MAGSAC	84.3	91.9	93.1	157.6
-	DISK	SMNN+MAGSAC	85.7	92.4	93.4	2368

Висновки. Отримані результати показують, що використання лише локальних дескрипторів займає значно більше часу порівняно з іншими підходами, за виключенням гібридного підходу з LoFTR. Глобальні дескриптори працюють значно швидше за інші підходи, демонструючи стабільно хороші, але не найкращі результати, незалежно від змін у середовищі.

При наявності лише варіації в освітленні (табл. 1) результати локальних і глобальних дескрипторів є досить близькими, проте локальні дескриптори мають помітну перевагу за значенням Recall@1. Найкраще на цьому тесті себе показали гібридні підходи.

Якщо до варіації освітлення додається зміна точки спостереження (табл. 2) можемо спостерігати значне погіршення результатів. Локальні дескриптори виявились кращими за глобальні, а гібридні підходи знову показали найкращий результат.

На наборі даних з присутніми змінами погоди та освітлення (табл. 3) результати локальних та глобальних дескрипторів виявились приблизно однаковими, а гібридні підходи виявились трохи кращими за інші.

Такі результати свідчать про те, що гібридні підходи здатні стабільно досягати кращих або не гірших результатів порівняно з їх окремими компонентами, незалежно від умов середовища. При цьому, завдяки зменшенню кількості кандидатів яких опрацьовують локальні дескриптори, вони також є значно швидшими за ці локальні дескриптори.

Окремо також можна відмітити LoFTR, бо підходи з ним досягли одних з найкращих результатів, але оскільки він є трансформером, то відповідно і час на розрахунок виявився найбільшим серед інших гібридних підходів.

Загалом, найвищих значень Recall@K досягли гібридні методи: CosPlace + LoFTR (забезпечив найкращу точність, але вимагає багато часу), пари CosPlace + DeDoDe та CosPlace + DISK показали дуже гарні за якістю результати, при цьому вони значно швидші за CosPlace + LoFTR та DeDoDe, DISK відповідно.

Список літератури:

1. Lowe D. G. Distinctive Image Features from Scale-Invariant Keypoints // *International Journal of Computer Vision*. – 2004. – Т. 60, № 2. – С. 91–110. – Режим доступу: <https://doi.org/10.1023/b:visi.0000029664.99615.94>.
2. Bay H., Tuytelaars T., Van Gool L. Speeded-Up Robust Features (SURF) // *Computer Vision and Image Understanding*. – 2008. – Т. 110, № 3. – С. 346–359. – Режим доступу: <https://doi.org/10.1016/j.cviu.2007.09.014>.
3. Rublee E., Rabaud V., Konolige K., Bradski G. ORB: An Efficient Alternative to SIFT or SURF // *2011 International Conference on Computer Vision*. – 2011. – С. 2564–2571. – Режим доступу: <https://doi.org/10.1109/iccv.2011.6126544>.
4. Dalal N., Triggs B. Histograms of Oriented Gradients for Human Detection // *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. – 2005. – С. 886–893. – Режим доступу: <https://doi.org/10.1109/cvpr.2005.177>.
5. Chen Z., Lam O., Jacobson A., Milford M. Convolutional Neural Network-Based Place Recognition [Електронний ресурс] // *arXiv.org*. – 2014. – Режим доступу: <https://doi.org/10.48550/arXiv.1411.1509>.
6. Arandjelovic R., Gronat P., Torii A., Pajdla T., Sivic J. NetVLAD: CNN Architecture for Weakly Supervised Place Recognition // *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. – 2016. – С. 5297–5307. – Режим доступу: <https://doi.org/10.1109/cvpr.2016.572>.
7. Berton G., Masone C., Caputo B. Rethinking Visual Geo-localization for Large-Scale Applications // *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. – 2022. – С. 4868–4878. – Режим доступу: <https://doi.org/10.1109/cvpr52688.2022.00483>.
8. DeTone D., Malisiewicz T., Rabinovich A. SuperPoint: Self-Supervised Interest Point Detection and Description [Електронний ресурс] // *arXiv.org*. – 2017. – Режим доступу: <https://doi.org/10.48550/arXiv.1712.07629>.
9. Tyszkiewicz M. J., Fua P., Trulls E. DISK: Learning Local Features with Policy Gradient [Електронний ресурс] // *arXiv.org*. – 2020. – Режим доступу: <https://doi.org/10.48550/arXiv.2006.13566>.
10. Edstedt J., Meyer T., Armeni I., Pollefeys M., Cohen A. DeDoDe: Detect, Don't Describe – Describe, Don't Detect for Local Feature Matching [Електронний ресурс] // *arXiv.org*. – 2023. – Режим доступу: <https://doi.org/10.48550/arXiv.2308.08479>.

11. Sun J., Jiang Y., Chen Y., Yuan L. LoFTR: Detector-Free Local Feature Matching with Transformers // *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. – 2021. – С. 8918–8927. – Режим доступу: <https://doi.org/10.1109/cvpr46437.2021.00881>.
12. Barbarani G., Dai Q., Pooch A., Thomas B., Zamir A. R. Are Local Features All You Need for Cross-Domain Visual Place Recognition? // *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Vancouver, BC, Canada, 17–24 June 2023. – 2023. – Режим доступу: <https://doi.org/10.1109/cvprw59228.2023.00655>
13. Fischler M. A., Bolles R. C. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography // *Communications of the ACM*. – 1981. – Т. 24, № 6. – С. 381–395. – Режим доступу: <https://doi.org/10.1145/358669.358692>.
14. Barath D., Noskova J., Ivashechkin M., Matas J. MAGSAC++, a Fast, Reliable and Accurate Robust Estimator [Електронний ресурс] // *arXiv.org*. – 2019. – Режим доступу: <https://doi.org/10.48550/arXiv.1912.05909>.
15. Sünderhauf N., Shirazi S., Dayoub F., Upcroft B., Milford M. On the Performance of ConvNet Features for Place Recognition // *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. – 2015. – С. 4297–4304. – Режим доступу: <https://doi.org/10.1109/IROS.2015.7354142>.
16. Chen Z., Jacobson A., Milford M., Lam O. Learning Context Flexible Attention Model for Long-Term Visual Place Recognition // *IEEE Robotics and Automation Letters*. – 2018. – Т. 3, № 4. – С. 4015–4022. – Режим доступу: <https://doi.org/10.1109/lra.2018.2859916>.

References:

1. Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60, 91-110.
2. Bay, H., Tuytelaars, T., & Van Gool, L. (2006). Surf: Speeded up robust features. In *Computer Vision–ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006. Proceedings, Part I 9* (pp. 404-417). Springer Berlin Heidelberg.
3. Rublee, E., Rabaud, V., Konolige, K., & Bradski, G. (2011, November). ORB: An efficient alternative to SIFT or SURF. In *2011 International conference on computer vision* (pp. 2564-2571). Ieee.
4. Dalal, N., & Triggs, B. (2005, June). Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)* (Vol. 1, pp. 886-893). Ieee.
5. Chen, Z., Lam, O., Jacobson, A., & Milford, M. (2014). Convolutional neural network-based place recognition. *arXiv preprint arXiv:1411.1509*.
6. Arandjelovic, R., Gronat, P., Torii, A., Pajdla, T., & Sivic, J. (2016). NetVLAD: CNN architecture for weakly supervised place recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 5297-5307).
7. Berton, G., Masone, C., & Caputo, B. (2022). Rethinking visual geo-localization for large-scale applications. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 4878-4888).
8. DeTone, D., Malisiewicz, T., & Rabinovich, A. (2018). Superpoint: Self-supervised interest point detection and description. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 224-236).

9. Tyszkiewicz, M., Fua, P., & Trulls, E. (2020). DISK: Learning local features with policy gradient. *Advances in Neural Information Processing Systems*, 33, 14254-14265.
10. Edstedt, J., Bökman, G., Wadenbäck, M., & Felsberg, M. (2024, March). DeDoDe: Detect, don't describe—Describe, don't detect for local feature matching. In *2024 International Conference on 3D Vision (3DV)* (pp. 148-157). IEEE.
11. Sun, J., Shen, Z., Wang, Y., Bao, H., & Zhou, X. (2021). LoFTR: Detector-free local feature matching with transformers. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 8922-8931).
12. Barbarani, G., Mostafa, M., Bayramov, H., Trivigno, G., Berton, G., Masone, C., & Caputo, B. (2023). Are local features all you need for cross-domain visual place recognition?. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 6155-6165).
13. Fischler, M. A., & Bolles, R. C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6), 381-395.
14. Barath, D., Nuskova, J., Ivashechkin, M., & Matas, J. (2020). MAGSAC++, a fast, reliable and accurate robust estimator. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 1304-1312).
15. Sünderhauf, N., Shirazi, S., Dayoub, F., Upcroft, B., & Milford, M. (2015, September). On the performance of convnet features for place recognition. In *2015 IEEE/RSJ international conference on intelligent robots and systems (IROS)* (pp. 4297-4304). IEEE.
16. Chen, Z., Liu, L., Sa, I., Ge, Z., & Chli, M. (2018). Learning context flexible attention model for long-term visual place recognition. *IEEE Robotics and Automation Letters*, 3(4), 4015-4022.

Статтю представив д-р ф.-м. наук, професор О.Д. Кічмаренко.

Надійшла (received) 14.05.2025

Zhykul Danylo, student
Odesa I.I.Mechnikov National University
Str. Krymska, 82, Odesa, Ukraine, 65043
Tel.: (050) 930-58-52, email: danylo.zhykul@stud.onu.edu.ua
ORCID ID:0009-0003-8023-2839

Moroz Volodymyr, PhD, Associate Professor
Odesa I.I.Mechnikov National University
Str. Dvorianska, 2, Odesa, Ukraine, 65082
Tel.: (093) 273-45-55, email: v.moroz@onu.edu.ua
ORCID ID:[0000-0002-3240-4590](https://orcid.org/0000-0002-3240-4590)

УДК 004.9

Аналіз алгоритмів візуального виявлення місць / Жикул Д.С., Мороз В.В. // Вісник НТУ "ХПІ". Серія: Інформатика та моделювання. – Харків: НТУ "ХПІ". – 2025. – № 2 (14). – С. 77 – 86.

Задача візуального виявлення місць (VPR – Visual Place Recognition) є важливою в комп'ютерному зорі та робототехніці. Останніми роками інтерес до цієї теми суттєво зріс, зокрема через її актуальність для задач автономної навігації та локалізації. Алгоритми VPR широко застосовуються в системах Visual SLAM для виявлення замикання петель. Однак ця задача ускладнюється через низку факторів: наявність візуально схожих місць, зміну точки спостереження, варіації освітлення, сезонні й структурні зміни середовища тощо. У сучасних дослідженнях пропонуються різні підходи до розв'язання цієї проблеми, які зазвичай класифікують за типом використовуваних дескрипторів. У цій роботі розглянуто та проаналізовано ці основні групи методів. Іл.: 4. Табл.: 3. Бібліогр.: 16 назв.

Ключові слова: візуальне виявлення місць, автономна навігація, місце, дескриптор.

УДК 004.9

Analysis of Visual Place Recognition Algorithms / Zhykul D.S., Moroz V.V. // Herald of the National Technical University "KhPI". Series of "Informatics and Modeling". – Kharkov: NTU "KhPI". – 2025. – № 2 (14). – P. 77 – 86.

The task of Visual Place Recognition (VPR) plays a crucial role in computer vision and robotics. Recently, interest in VPR has grown significantly, driven largely by its importance for autonomous navigation and localization. VPR algorithms are widely employed within Visual SLAM systems to detect loop closures. However, the problem is made challenging by factors such as visually similar places, changes in viewpoint, lighting variations, and seasonal or structural environmental alterations. Contemporary research offers a variety of solutions to these challenges, which are typically classified according to the type of descriptors they employ. In this work, we review and analyze these main groups of methods. Figs.: 4. Tabl.: 3. Refs.: 16 titles.

Keywords: visual place recognition, autonomous navigation, place, descriptor.