ISSN 1870-4069

A Geometric Strategy for Recognizing Images of Highly Similar Places within Urban Environments

Carlos A. Martínez-Miwa, Mario Castelán, L. Abril Torres-Méndez

Centro de Investigación y de Estudios Avanzados Campus Saltillo,
Robotics and Advanced Manufacturing Group,
Mexico

{carlos.miwa, mario.castelan, abril.torres}@cinvestav.edu.mx

Abstract. The recognition of previously visited places within urban environments is an essential skill for autonomous vehicles, as it may reduce localization errors during their navigation. The search for improvements in detection capabilities within regions where other sensors, such as lasers or GPS (Global Positioning System) do not perform accurately, has contributed to considerable advances in location recognition systems. Some state of the art approaches require a priori knowledge of the environment. However, this is not always useful due to constant changes in the outside world, variations of viewpoint, or the occurrence of similar images captured from different locations. In this work, we propose a methodology to carry out the visual recognition of places with highly similar characteristics, and prone to spatial variations, illumination changes and occlusions. Our recognition strategy is based on image retrieval by means of detector-descriptors pairs, from which the combination GFTT-SIFT (Good Features to Track - Scale Invariant Feature Transform) exhibits the best performance. For results refinement, we use an image similarity threshold based on geometric constraints. Compared to a high-level learning approach the proposed methodology has a greater precision and discrimination power to identify images of similar zones, besides differentiating those belonging to different sites.

Keywords: Place recognition, machine learning, computer vision, feature detection.

1 Introduction

Visual recognition of previously visited places is a fundamental part of our daily lives. The study of how living beings recognize places, taking into account the movement from one place to another, has a long history in neuroscience [3, 11]. Several discoveries in this area have provided a physiological basis for the representation of spatial locations in our brain [22, 24].

As humans, when visiting a place for the first time, we seem to be more attentive to those details that we believe will best represent it, looking for them to be sufficiently distinctive to create a strong association. Hence, by revisiting that location in the future,

even when different conditions exist, selected features may be activated leading to an accurate detection [27].

These concepts find application in a wide variety of research fields. Such is the case of robotics. One fundamental goal of this area is to develop fully functional systems that can operate robustly in the real world. Mobile robots, particularly autonomous ones, must have a deep understanding of their surroundings so that they can be entrusted with highly complicated or risky responsibilities that humans should not take on, for example, preventing natural disasters, space and underwater exploration, or even search and rescue activities. Therefore, visual place recognition (VPR) becomes an extremely important process as it enables robots to reduce uncertainty and location errors during their exploration.

This paper proposes a methodology for the identification of previously visited sites under challenging conditions, mostly based on geometric constraints. In the context of urban environments, the term "challenging" refers to spatial variations, illumination, occlusions and the presence of similar elements with great frequency. We first designed a novel database to depict this sort of settings. We also tested a significant number of local feature detector-descriptor combinations aimed at selecting the one that performed the best for our dataset. Place recognition is determined from a geometrical nature concept that, in spite of being more commonly associated with topics such as visual odometry, generates highly favorable results in the search for previosly seen places. Our method achieves a reliable place recognition, without the requirement of prior training, surpassing the performance of a state of the art algorithm.

This article is organized as follows: relevant work is described in Section 2; the process of gathering the database for testing our method is explained in Section 3; proposed method is briefly depicted in Section 4; experimental results are presented in Section 5; and, finally, conclusions and future research directions are provided in Section 6.

2 Related Work

State of the art related with VPR includes research works such as the displacement of a robot along a previously learned route [14]. The information acquired by means of sensors, i.e. cameras, is first described and then compared with an internal representation, or map of the environment, in order to estimate the probability of data matching an image inside the map. Unfortunately, if a robot intends to act without previous knowledge of its surroundings, this procedure becomes extremely challenging, mainly due to three main factors:

1. Variability in the appearance of the same scene (changing illumination, occlusions, weather conditions).
2. The possibility that a scene viewpoint may not always be the same.
3. Images from different locations looking too similar, effect known as perceptual aliasing.

Other conventional approaches are those based on visual scene detection and description techniques. These can operate with local features (involving a

Fig. 1. The left column exhibits images from the same site at different times of the day. The top left image was captured at 19 h while the bottom left image at 13 h. The second column depicts two frames of locations that were far from each other, but visually similar that they may appear to belong to the same site and the same hour.

detection-description pair), such as scale-invariant feature transformations (SIFT) [13] and Speeded-Up Robust Features (SURF) [2], or, alternatively, resource to whole images without the need for a detection stage [26].

Since feature extraction does not involve a very complex and demanding process, it is not surprising to discover combinations of these methods [18, 20]. Nevertheless, a poor performance of this kind of descriptors has been reported upon varying circumstances, especially those related with illumination changes [8].

In [6] location or object recognition problems are addressed through the Bag of Words (BoW). This involves representing image features in terms of a numeric vector, that can be efficiently compared to other vectors encoding information about a series of words. These words are the names given to the image descriptors. While this approach performs well and is scalable to large amounts of data, its performance and functionality decline when regions with conditions other than those included in the training images are encountered.

Analogous to the BoW model, the "Bag of Relevant Regions" was presented in [15]. This novel method aimed at describing a scene in terms of relevant regions, extracted from a visual attention algorithm. Although this work outperforms well-known approaches such as the Fast Appearance Based-Mapping (FAB-MAP) [5], it outputs a great amount of false negatives. FAB-MAP applies probabilistic calculations, based on the local appearance of a site, for its identification. Perceptual aliasing is tackled not only by considering whether two scenarios are similar in terms of the visual words they have in common, but also that these are sufficiently distinctive.

As a result, if two sites seem similar but their words are frequently observed, FAB-MAP generates a low correspondence probability. FAB-MAP uses a BoW model, with SIFT or SURF features, for image description and computes the dissimilarity of each word during a training phase. Nevertheless, this training causes a computational cost increase.

Authors in [9] suggest a solution for VPR based on a BoW built on a local feature detector-descriptor combination for the purposes of simultaneous localization

Table 1. List of 22 detector-descriptor combinations used to identify previously visited locations at different times of the day. The overall success rate was $52.62 \pm 14.79\%$.

Detector-Descriptor	Success %	Detector-Descriptor	Success %
AVA-ORB	73.33	STAR-SIFT	46.66
AVA-SIFT	73.33	STAR-SURF	44.44
AVA-SURF	73.33	KAZE-KAZE	42.22
GFTT-BRISK	73.33	AKAZE-AKAZE	40
GFTT-ORB	73.33	BRISK-BRISK	40
GFTT-SIFT	73.33	FAST-BRISK	40
GFTT-SURF	73.33	FAST-ORB	40
ORB-ORB	51.11	FAST-SIFT	40
AVA-BRISK	46.66	FAST-SURF	40
STAR-BRISK	46.66	SIFT-SIFT	40
STAR-ORB	46.66	SURF-SURF	40

and mapping (SLAM). The selection of these algorithms aimed at reducing processing time, although no prior evaluation of detector-descriptor combinations was carried out. Moreover, although some of the databases used are spatially dynamic, they do not reflect changes in the hours of the day.

A variety of machine learning methods have also been resorted to. In [23], Histogram of Oriented Gradients (HOG) [7] fetaures and Local Binary Patterns (LBP) [21] are concatenated for visual localization. Then, given an image, a Support Vector Machine (SVM) model identifies the most similar one within a geo-referenced database. Other approaches rely on Convolutional Neural Networks (CNNs) as strong feature extractors for place recognition in changing environments.

Researchers in [4] and [28] performed an analysis of the robustness of different CNN layers against visual appearance and viewpoint modifications across a set of images. It was concluded that intermediate layers exhibit robustness to appearance alterations, while higher level layers perform better facing viewpoint shifts.

Notwithstanding, no mechanism is presented for an automatic selection of the best layer for the task at hand. A dependency on the training database is also evident. Thus, features that generate good results on one dataset, may have little impact against a different one. Further works related to deep learning have emerged recently [1, 10, 19, 31]. Nonetheless, overall the main disadvantage is the need for large amounts of training data and the consequent high computational costs.

3 Dataset Collection

For this work, we collected a new place recognition dataset. Our image collection focuses on depicting urban environments that could reflect highly challenging conditions for a computer vision system. Here, the term challenging alludes to settings that do not contain significant visual information or that are subject to dynamic factors.

The gathering of these images was inspired by [17], where the authors explored how participants recognized, through defiant conditions, different pleaces recorded

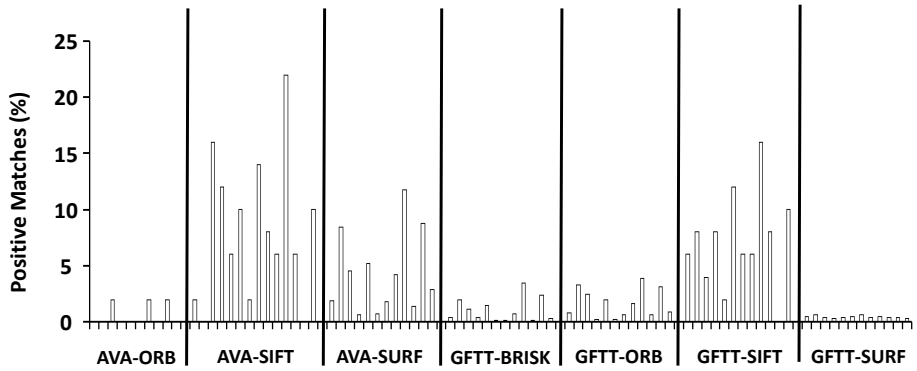

Fig. 2. Match percentages for the 7 best algorithms evaluated on the 12 most challenging images. Note that the strongest performing detector-descriptor pairs are AVA-SIFT and GFTT-SIFT.

along a video sequence describing the navigation of a car. The image collection was gathered at the city of Saltillo, Mexico, driving along a 0.8 km route, at a speed of 30 km/h, through a series of streets which could be identified as belonging to a typical urban neighborhood.

Three sequences compose the dataset, each of which was captured at different hours (07 h, 13 h and 19 h) of the day. For this procedure, a GoPro Hero4 camera mounted on a Chevrolet Cruze vehicle was used.

The data comprises a total of 447 images, 149 per time of the day. The first column of Figure 1 illustrates examples of images representing the same location at 19 h and 13 h. The second column of the figure presents frames from distinct scenarios that share very similar characteristics. The database is challenging, since many of the major image processing problems are addressed, i.e., illumination and spatial variations, occlusions, or the presence of frequent similar elements along navigation.

In addition, the fact that the images were captured under different environmental conditions can result in the detection of mismatches in several elements, for example building colors, plants or even the sky, leading to undesirable detections and confusion.

4 Evaluating Detector-Descriptor Pairs for Geometric-Based VPR

The first need to be fulfilled for geometric-based VPR is to count on a reliable detector-descriptor pair. For this reason, we conducted a thorough evaluation of 22 detector-descriptor couples for identifying whether a reference image was or not included in its related video sequence. Such techniques were designated because of their ease of implementation and access availability. Results are listed in Table 1.

A threshold was applied to every couple in order to determine if the found correspondences were sufficient to establish a positive match between two images. For setting this threshold, 80 % of the maximum number of matched points was selected.

From the analysis of the table, it is possible to appreciate how 7 out of 22 combinations reached the highest success rates, while the worst performance was attributed to other 7 pairs. For this reason, we focused on the 7 highest-rated.

Fig. 3. Example of GFFT-SIFT qualitative results. It is to note how, even at different times of the day and between relatively distant scenes, GFTT-SIFT is capable of detecting enough features so as to determine a positive match among both images.

These 7 best detector-descriptor pairs were then tested on the 12 worst performing images to point out the strongest performance.

AVA-SIFT (Aqua Visual Attention [16] - SIFT) and GFTT-SIFT (Good Features To Track [25] - SIFT) emerged as the most outstanding in terms of the number of matches found, as shown in Figure 2.

Although the number of correspondences is a good indicator for determining the similarity between two scenes, there is a possibility that this parameter carries some uncertainty. The latter refers to the fact that if, for a couple of images, a detector-descriptor generates a number of correspondences lower than the threshold set (80% of the maximum found), these could be enough to state that both scenes depict the same location.

Results of the 7 top detector-descriptor combinations were revisited for the 12 worst performing images, but this time in a qualitative fashion, to verify whether or not they constituted a good match.

In this way, it was possible to identify that, despite not exhibiting the best performance under the previous metric, GFTT-SIFT stood out from the rest. This pairing detected correspondences between images belonging to the same scenario, even if they suffered from a significant spatial offset as illustrated by Figure 3.

Once the detector-descriptor pair was selected, a new metric was defined to better discriminate among images that do and do not represent the same site. We chose this parameter to be based on the epipolar constraint.From this restriction, it is understood that there must exist a transformation $\mathbf{x} \rightarrow \mathbf{l'}$ of a point in one image with its respective epipolar line in a second one. The transformation from points to lines results in a correlation, expressed by the fundamental matrix $\mathbf{F}$ [12]:

$$\mathbf{l'} = \mathbf{Fx}. \tag{1}$$

The fundamental matrix, then, satisfies that for any couple of matching points $\mathbf{x}$ and $\mathbf{x'}$ in two images:

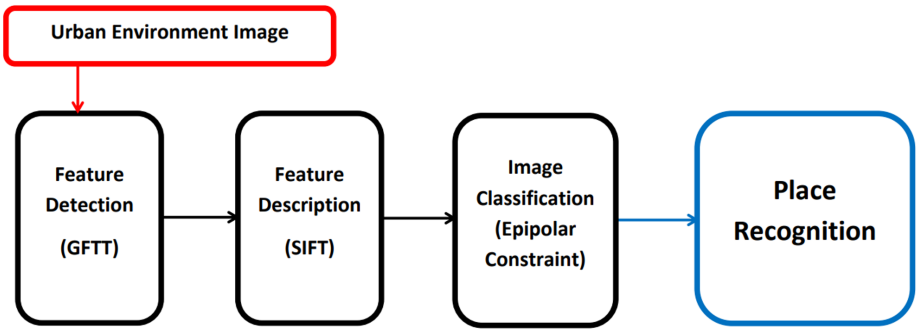

Fig. 4. Methodology proposed for place recognition in challenging environments.

$$\mathbf{x'^{T}Fx} = 0. \tag{2}$$

This condition is true because, if points **x** and **x'** are correspondent, **x'** lies on the epipolar line $\mathbf{l'} = \mathbf{Fx}$ related to point **x**. In other words, $\mathbf{x'^{T}l'} = \mathbf{x'^{T}Fx} = 0$. Besides, if image points satisfy $\mathbf{x'^{T}Fx} = 0$, rays defined by them are coplanar, a necessary criterion to establish correspondence between them.

Taking as a reference equation (2), the proposed metric is introduced: if for two images, the number of matches found by GFTT-SIFT is high enough to generate a fundamental matrix, they will be considered as positive correspondences, that is, coming from the same scenario; otherwise, they will be catalogued as belonging to different locations.

A diagram describing the proposed methodology is shown in Figure 4. As a first step, starting from a given scene to be recognized, the most important features are detected and described in order to locate the best match. For this purpose, a classic detector-descriptor combination such as GFTT-SIFT is adopted.

These correspondences undergo a geometric classification method based on the epipolar constraint. In this procedure, we managed to eliminate all images that lie below a defined threshold, and also managed to categorize the remaining images into four main groups: True and False Positives, and True and False Negatives.

5 Results

The process depicted in Figure 4 was evaluated on the database described in Section 3. The evaluation consisted of an image retrieval task, i.e., each of the 447 images (149 morning $\times$ 149 afternoon $\times$ 149 night) was compared to the rest, aiming at determining whether matches found in each pairing were enough to create a fundamental matrix (i.e., satisfy the epipolar constraint).

By means of these trials, a tool that allows visualization and comparison of the results, known as the similarity matrix, could be constructed. Figure 5a illustrates the similarity matrix at the end of this experimentation.

Each black dot represents a positive match detected in a pair of images, i.e, that they belong to the same site. In order to build a ground truth matrix (Figure 5b), all pairing

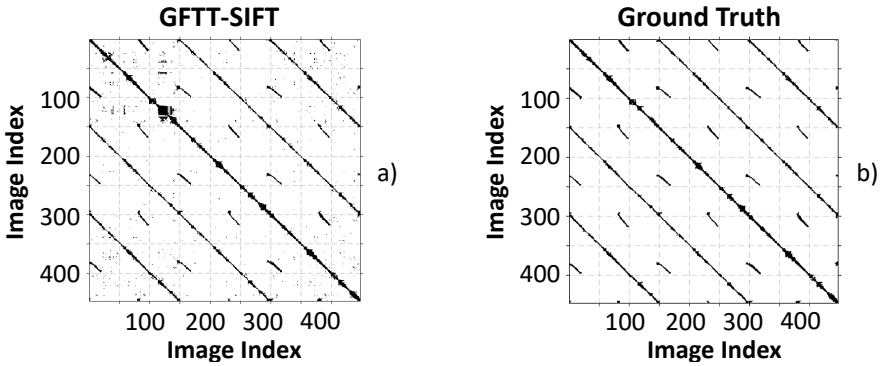

Fig. 5. Similarity matrices. Each point (dark regions) within the matrix represents a correspondence detected in a pair of images from different sequences. On the left side, figure 5a illustrates the results of our method. Figure 5b, on the right, plots the expected result.

images were carefully examined by the main author of this work, who visually decided which pairs of images were positive or negative matches. From Figure 5, it is to note that the similarity matrix derived from our method significantly resembles the ground truth. However, two special cases arise: False Positives and False Negatives. The former allude to additional points found in the figure's white zone, where matches would be assumed to be null.

Further, our results reveal a black box at the upper left side of the matrix. These artifacts, although located on the main diagonal, are made up of dots that should not exist, namely, False Positives.

False Negatives, on the other hand, refer to those images in which, despite depicting the same place, no correspondence between pairs of images was detected. Both cases constitute specific problems in the performance of the proposed method. In order to evaluate the presence of false positives and negatives in the performance or our method, we used a Precision-Recall (PR) curve, shown in Figure 6.

From the curve, it can be clearly perceived that as recall increases, precision decreases, though in a very low proportion. The accuracy achieved is considerably high, obtaining a maximum value of 0.9523, dropping only to 0.8371. The sensitivity factor also produces favorable results. In spite of the minimum value of 0.0519 being quite low, it reaches the recall limit of 1 with a still high precision. Taking these data into account, added to the fact that the area under the PR curve is of great dimension, it is possible to establish that our methodology achieved a strong classification capacity.

For comparative purposes, our dataset was additionally evaluated under a methodology with different *modus operandi*: the fast appearance-based mapping algorithm, or FAB-MAP [5].

FAB-MAP is one of the most popular solutions for VPR based on local image features. This approach turns to probability for the identification of locations and also employs a BoW model built upon appearence-based features, e.g, SIFT and SURF.

Despite representing an important milestone within the state of the art, its performance struggled to obtain favorable results in our database. As evidenced in

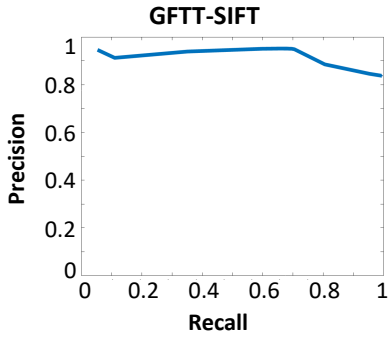

Fig. 6. PR curve produced by GFTT-SIFT detector-descriptor combination. It is noteworthy how, while the recall increases (up to a value of 1), precision decays only to a rate close to 0.85.

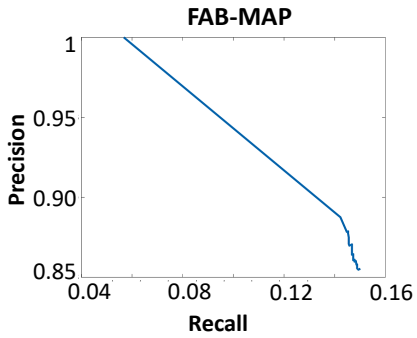

Fig. 7. FAB-MAP Precision-Recall curve. The prevalence of a high accuracy index (0.85) is notable. However, as Precision diminishes, Recall reaches just an index near 0.15.

Figure 7, whereas Precision drops close to 0.85, only a 0.15 Recall is reached. Such score indicates that although this methodology was able to discriminate most of the possible False Positive cases, there were a large number of False Negatives.

The latter is verified through the generated similarity matrix, depicted at the left of Figure 8). From the analysis of Figures 5 through 8 of this section, we can establish that the proposed algorithm, based on a detector-descriptor combination (GFTT-SIFT) under a geometric strategy, outperforms a learning-based method, such as FAB-MAP, for recognizing previously visited places subject to dynamic conditions (spatial, lighting and occlusions).

6 Conclusions and Future Work

In this work we presented a novel database for previously visited places in the context of VPR. The main particularity of these images is the set of challenging conditions for computer vision techniques.

Our video sequences were captured at different times of the day, yielding a combination of spatial and illumination changes, occlusions, similar elements with

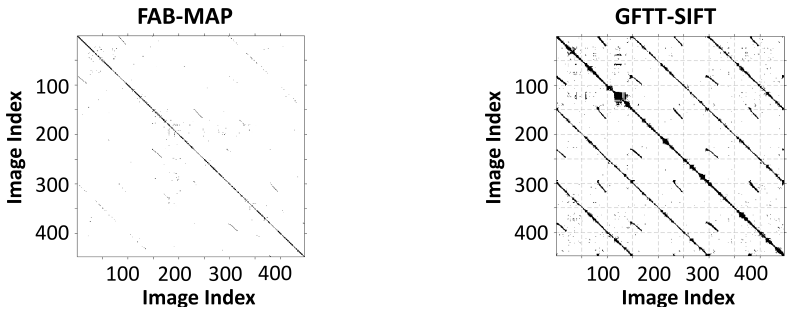

Fig. 8. Similarity matrix generated by FAB-MAP (a) compared to the similarity matrix obtained through our method (b). The presence of a large number of False Negatives can be seen, reinforcing the low recall shown in the PR curve of Figure 7.

high repeatability and even environmental factors that may cause confusion, such as the sky. On the basis of our experiments, we realize that classical computational algorithms, as combinations of detectors and local feature descriptors, generate sufficiently good results in location recognition with greater speed and simplicity and without compromising reliability, in comparison with a state of the art methods that uses BoW.

We conducted an exhaustive search for the best detector-descriptor combination for VPR. The Good Features To Track feature detector, along with the SIFT descriptor, exhibited high robustness in identifying reliable features that corresponded to important regions in the environment even in adverse situations such as changes in lighting due to the different day hours.

The designation of the fundamental matrix as a geometric constraint was a relevant addition for frame classification. Although it is most often applied to tasks such as visual odometry, it proved to be a solid and accurate method for the identification of previously visited sites. Our methodology, by itself, was able to produce highly successful results.

From a Precision-Recall Curve, a high measure of sensitivity (recall) was achieved with a very low decrease in pecision. Finally, these results are supported by a comparison against a learning method considered a milestone in the state of the art: FAB-MAP. The obtained plots exhibit a very low recall for this probabilistic algorithm, as well as a larger drop in precision. Thus, it is demonstrated that our appraoch is able to perform accurate, fast and simple VPR, without the need to rely on large quantities of training data, nor consuming high computational time.

As future work we intend to analyze the incorporation of techniques that provide different perspectives to the geometric ones, for instance, detector-descriptor pairs used for visual attention. In this way, we could address the highly challenging cases that could not be completely solved under the proposed methodology. We also aim to test our methodology under public and commonly used databases related to the VPR problem, for instance [29, 30].

Similarly, we are looking forward to publishing our novel dataset in a public repository so that other researchers can make use of it.

References

1. Atapour-Abarghouei, A., Akcay, S., de La-Garanderie, G.P., Breckon, T.P.: Generative Adversarial Framework for Depth Filling Via Wasserstein Metric, Cosine Transform and Domain Transfer. Pattern Recognition, vol. 91, pp. 232–244 (2019). DOI: 10.1016/j.patcog.2019.02.010.
2. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: Speeded-Up Robust Features (SURF). Computer Vision and Image Understanding, vol. 110, no. 3, pp. 346–359 (2008). DOI: 10.1016/j.cviu.2007.09.014.
3. Brown, J.W.: Neuropsychology of Visual Perception. Psychology Press, vol. 2 (1989)
4. Chen, Z., Lam, O., Jacobson, A., Milford, M.: Convolutional Neural Network-Based Place Recognition. In: Australasian Conference on Robotics and Automation (ACRA) (2014)
5. Cummins, M., Newman, P.: FAB-MAP: Probabilistic Localization and Mapping in the Space of Appearance. The International Journal of Robotics Research, vol. 27, no. 6, pp. 647–665 (2008). DOI: 10.1177/0278364908090961.
6. Cummins, M., Newman, P.: Appearance-only SLAM at Large Scale with FAB-MAP 2.0. The International Journal of Robotics Research, vol. 30, no. 9, pp. 1100–1123 (2011). DOI: 10.1177/0278364908090961.
7. Dalal, N., Triggs, B.: 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 (2005). DOI: 10.1109/CVPR.2005.177.
8. Furgale, P., Barfoot, T.: Visual Teach and Repeat for Long Range Rover Autonomy. Journal of Field Robotics, vol. 27, no. 5, pp. 534–560 (2010). DOI: 10.1002/rob.20342.
9. Gálvez-López, D., Tardos, J.D.: Bags of Binary Words for Fast Place Recognition in Image Sequences. IEEE Transactions on Robotics, vol. 28, no. 5, pp. 1188–1197 (2012). DOI: 10.1109/TRO.2012.2197158.
10. Garg, S., Milford, M.: SeqNet: Learning Descriptors for Sequence-Based Hierarchical Place Recognition. IEEE Robotics and Automation Letters, vol. 6, no. 3, pp. 4305–4312 (2021). DOI: 10.1109/LRA.2021.3067633.
11. Golledge, R.G.: Do People Understand Spatial Concepts: The Case of First-Order Primitives. Theories and Methods of Spatio-Temporal Reasoning in Geographic Space, pp. 1–21 (1992). DOI: 10.1007/3-540-55966-3_1.
12. Hartley, R.I., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press, (2004)
13. Lowe, D.G.: Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision, vol. 60, no. 2, pp. 91–110 (2004). DOI: 10.1023/B:VISI.0000029664.99615.94.
14. Lowry, S., Sünderhauf, N., Newman, P., Leonard, J., Cox, D., Corke, P., Milford, M.: Visual Place Recognition: A Survey. IEEE Transactions on Robotics, vol. 32, no. 1, pp. 1–19 (2016). DOI: 10.1109/TRO.2015.2496823.
15. Maldonado-Ramírez, A., Torres-Méndez, L.A.: A Bag of Relevant Regions for Visual Place Recognition in Challenging Environments. In: 23rd International Conference on Pattern Recognition, pp. 1358–1363 (2016). DOI: 10.1109/ICPR.2016.7899826.
16. Maldonado-Ramírez, A., Torres-Méndez, L.A.: Robotic Visual Tracking of Relevant Cues in Underwater Environments with Poor Visibility Conditions. Journal of Sensors, vol. 2016, pp. 1–16 (2016). DOI: 10.1155/2016/4265042.
17. Martínez-Miwa, C.A., Castelán, M., Torres-Méndez, L.A., Maldonado-Ramírez, A.: Human and Machine Capabilities for Place Recognition: A Comparison Study. In: The Tenth International Conference on Advanced Cognitive Technologies and Applications, pp. 72–77 (2018)

18. Mei, C., Sibley, G., Cummins, M., Newman, P., Reid, I.: A Constant-Time Efficient Stereo SLAM System. In: Proceedings of the 20th British Machine Vision Conference, pp. 1–11 (2009)
19. Mendez, O., Hadfield, S., Pugeault, N., Bowden, R.: SeDAR: Reading Floorplans Like a Human-Using Deep Learning to Enable Human-Inspired Localisation. International Journal of Computer Vision, vol. 128, no. 5, pp. 1286–1310 (2020). DOI: 10.1007/s11263-019-01239-4.
20. Newman, P., Chuchill, W.: Experience-Based Navigation for Long-Term Localization. The International Journal of Robotics Research, vol. 32, no. 14, pp. 1645–1661 (2013). DOI: 10.1177/0278364913499193.
21. Ojala, T., Pietikäinen, M., Harwood, D.: A Comparative Study of Texture Measures with Classification Based on Featured Distributions. Pattern Recognition, vol. 29, no. 1, pp. 51–59 (1996). DOI: 10.1016/0031-3203(95)00067-4.
22. O'keefe, J., Nadel, L.: The Hippocampus as a Cognitive Map. Behavioral and Brain Sciences, vol. 2, no. 4, pp. 487–494 (1979). DOI: 10.1017/S0140525X00063949.
23. Qiao, Y., Cappelle, C., Ruichek, Y.: Place Recognition Based Visual Localization Using LBP Feature and SVM. In: Mexican International Conference on Artificial Intelligence, vol. 9414. pp. 393–404 (2015). DOI: 10.1007/978-3-319-27101-9_30.
24. Redish, A.D., Touretzky, D.S.: Cognitive Maps Beyond the Hippocampus. Hippocampus, vol. 7, no. 1, pp. 15–35 (1997) DOI: 10.1002/(SICI)1098-1063(1997)7:1 ¡15::AID-HIPO3¿3.0.CO;2-6.
25. Shi, J., Tomasi, C.: Good Features to Track. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 593–600 (1994). DOI: 10.1109/CVPR.1994.323794.
26. Siagian, C., Itti, L.: Impact of Neuroscience in Robotic Vision Localization and Navigation. Computational and Cognitive Neuroscience of Vision, Cognitive Science and Technology, pp. 235–276 (2017). DOI: 10.1007/978-981-10-0213-7_11.
27. Sowa, J.F.: Conceptual Structures: Information Processing in Mind and Machine. Addison-Wesley Longman Publishing Co., Inc, (1984)
28. Sünderhauf, N., Shirazi, S., Dayoub, F., Upcroft, B., Milford, M.: On the Performance of ConvNet features for Place Recognition. In: 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 4297–4304 (2015). DOI: 10.1109/IROS.2015.7353986.
29. Torii, A., Arandjelovic, R., Sivic, J., Okutomi, M., Pajdla, T.: 24/7 Place Recognition by View Synthesis. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1808–1817 (2015). DOI: 10.1109/CVPR.2015.7298790.
30. Torii, A., Sivic, J., Okutomi, M., Pajdla, T.: Visual Place Recognition with Repetitive Structures, pp. 883–890 (2015). DOI: 10.1109/CVPR.2013.119.
31. Zhang, Y., Bai, Y., Ding, M., Ghanem, B.: Multi-Task Generative Adversarial Network for Detecting Small Objects in the Wild. International Journal of Computer Vision, vol. 128, no. 6, pp. 1810–1828 (2020). DOI: 10.1007/s11263-020-01301-6.