ConvSequential-SLAM

A Sequence-Based, Training-Less Visual Place Recognition Technique for Changing Environments

Tomia, Mihnea Alexandru; Zaffar, Mubariz; Milford, Michael J.; McDonald-Maier, Klaus D.; Ehsan, Shoaib

DOI
10.1109/ACCESS.2021.3107778

Publication date
2021

Document Version
Final published version

Published in
IEEE Access

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 30, 2021, accepted August 18, 2021, date of publication August 24, 2021, date of current version September 1, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3107778

ConvSequential-SLAM: A Sequence-Based, Training-Less Visual Place Recognition Technique for Changing Environments

MIHNEA-ALEXANDRU TOMIȚĂ[1], MUBARIZ ZAFFAR[2], MICHAEL J. MILFORD[3], (Senior Member, IEEE), KLAUS D. MCDONALD-MAIER[1], (Senior Member, IEEE), AND SHOAIB EHSAN[1], (Senior Member, IEEE)

[1]School of Computer Science and Electronic Engineering, University of Essex, Colchester, Essex CO4 3SQ, U.K.
[2]Cognitive Robotics Department, Delft University of Technology, 2628 CD Delft, The Netherlands
[3]School of Electrical Engineering and Computer Science, Queensland University of Technology, Brisbane, QLD 4000, Australia

Corresponding author: Mihnea-Alexandru Tomiță (matomi@essex.ac.uk)

This work was supported by the U.K. Engineering and Physical Sciences Research Council under Grant EP/R02572X/1 and Grant EP/P017487/1.

ABSTRACT Visual Place Recognition (VPR) is the ability to correctly recall a previously visited place under changing viewpoints and appearances. A large number of handcrafted and deep-learning-based VPR techniques exist, where the former suffer from appearance changes and the latter have significant computational needs. In this paper, we present a new handcrafted VPR technique, namely ConvSequential-SLAM, that achieves state-of-the-art place matching performance under challenging conditions. We utilise sequential information and block-normalisation to handle appearance changes, while using regional-convolutional matching to achieve viewpoint-invariance. We analyse content-overlap in-between query frames to find a minimum sequence length, while also re-using the image entropy information for environment-based sequence length tuning. State-of-the-art performance is reported in contrast to 9 contemporary VPR techniques on 4 public datasets. Qualitative insights and an ablation study on sequence length are also provided.

INDEX TERMS SLAM, sequence-based filtering, visual localization, visual place recognition.

I. INTRODUCTION

VPR is the ability of a robot to correctly identify a previously visited place using visual information. It is a challenging problem due to variations in viewpoint, scale, illumination, seasons, clutter background and dynamic objects. Confusing and feature-less frames can also drastically increase the difficulty in place matching [1] due to the lack of distinct features to distinguish a given place from a geographically-different but visually-similar place (perceptual-aliasing).

VPR is usually cast as an image retrieval problem. Prior to the usage of deep-learning in VPR systems, handcrafted local and global feature descriptors were used to perform place recognition. Local feature descriptors only process salient parts (keypoints) of the image, while global feature descriptors process the entire image regardless of its content. The performance of local feature descriptors suffers under illumination changes in the environment, while global feature descriptors cannot handle viewpoint variation [2]. The application of deep-learning, especially Convolutional Neural Networks (CNNs), was first studied by Chen *et al.* in [3] and since then most of the advances in VPR have been primarily due to deep-learning-based techniques. CNNs are systems capable of learning features extracted from images using supervised training on labeled datasets. Such CNN-based VPR techniques have achieved state-of-the-art performance on the most challenging VPR datasets, as evaluated in [4] and [5]. However, in order to train a CNN for VPR tasks, one needs a large-scale dataset of labeled images taken from different environments, under various angles, seasons and illumination conditions. Although labelled VPR datasets exist, such as Oxford Robot Car dataset [6], SPED dataset [7] and Pittsburgh dataset [8], they represent a particular environment under limited

The associate editor coordinating the review of this manuscript and approving it for publication was Charalambos Poullis.

conditional and viewpoint changes. Therefore, the creation of a large-scale, labeled dataset representing all the different possible variations is not feasible and requires significant time and resources. Furthermore, training a CNN within reasonable times to adjust to a new environment will require dedicated Graphics Processing Units (GPUs) and may take several days/weeks in order to be trained. Because of the CNN's intense computational nature, their encoding-time and run-time memory are also significantly higher than those needed for the handcrafted feature descriptors. Although the CNN-based VPR techniques have largely outperformed handcrafted feature descriptor-based techniques on the image matching front, their intense computational requirements make them harder to use in this field. As a result of these demanding requirements, the deployment of CNN-based techniques for VPR are restricted for resource-constrained vehicles such as battery-powered aerial, micro-aerial and ground vehicles, as discussed in [5] and [9].

In this work, we propose a novel sequence-based and training-free VPR technique, namely ConvSequential-SLAM, that is successfully able to perform Visual Place Recognition (VPR) under changing viewpoint and appearance conditions. In contrast to other sequence-based VPR systems such as [10] and [11] that use a constant sequence length of images, our technique is using a dynamic sequence-based matching approach that is able to determine the most representative sequence length for each sequence of images. The resulting system is a training-less and light-weight VPR system, successfully able to adapt to any environment. We report state-of-the-art performance on both viewpoint and conditionally-variant datasets. Fig. 1 shows the block-diagram of ConvSequential-SLAM.

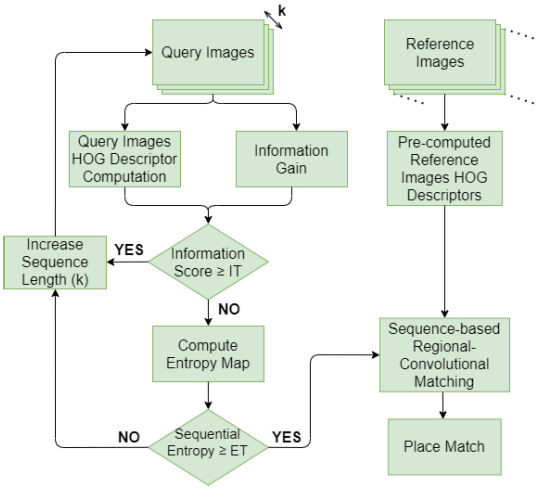

FIGURE 1. The block diagram of our framework is given here, which presents all the major components of the system.

We make the following main contributions in this paper: 1) We integrate convolutional scanning into our system, achieving robustness to moderate viewpoint variations; 2) We achieve conditional invariance by using regional, block-normalised Histogram-of-Oriented-Gradients (HOG) descriptors instead of contrast-enhanced pixel-matching; 3) We developed an analysis based on *information-gain* from consecutive query images to determine the minimum sequence length needed. This approach can only be used by this particular algorithm presented in the paper, and cannot be applied to all possible algorithms; 4) Building upon the sequence length generated by analysing consecutive query images, we use the entropy computation for salient region extraction to formulate an optimal dynamic sequence length, instead of a constant sequence length, as used in sequence-based VPR techniques.

The remainder of the paper is organised as follows: Section II presents an overview of the existing literature in VPR. Section III presents detailed information about ConvSequential-SLAM, while in Section IV we discuss the experimental setup. Section V presents the results and analysis obtained by evaluating the performance of our algorithm against other VPR techniques on public VPR datasets. The conclusion and future work are given in Section VI.

II. LITERATURE REVIEW

A comprehensive review of the existing challenges and research in the field of VPR is presented in [2]. Local feature descriptors such as scale-invariant feature transform (SIFT) [12] and speeded-up robust features (SURF) [13] make use of the most notable features in the image for extraction (keypoints), followed by description. These local descriptors have been widely used to perform VPR such as in [14]–[17], [18]. FAB-MAP [19] is an appearance based place recognition system based on local feature descriptors integrated within a SLAM system. It represents visual places as words and uses SURF for feature detection. CAT-SLAM [20], extends the work of FAB-MAP by including odometry information. Center Surround Extremas (CenSurE) [21] introduces a suite of new feature detectors that outperforms the previously mentioned local feature descriptors, performing real-time detection and matching of image features. CenSurE has been used by FrameSLAM in [22]. The Bag-of-Words model (BoW) [23], [24] has been used for VPR tasks such as in [25].

A very popular global feature descriptor is Gist [26], [27]. The work done in [28]-[30] shows some examples of Gist whole-image descriptor used in place recognition. Histogram-of-Oriented-Gradients (HOG) [31] is a global descriptor which creates a histogram by calculating the gradient of all pixels present in the image. McManus *et al.* used HOG for VPR in [32]. In SeqSLAM [10], sequences of camera frames are compared instead of single frames, thus achieving increased performance in VPR compared to traditional feature-based techniques, when the place is subject to drastic changes. More recently in CoHOG [33], the authors proposed a training-free technique based on the HOG descriptor that is able to achieve state-of-the-art performance in VPR.

CNNs are known to be robust feature extractors and their performance on VPR related tasks showed promising results,

thus being extensively explored in the field of place recognition in challenging environments. The authors of [3], combined all 21 layers of the Overfeat network [34] trained on ImageNet 2012 dataset together with the spatial and sequential filter of SeqSLAM. Chen *et al.* [7] trained two neural network architectures, namely HybridNet and AMOSNet, on the Specific PlaceEs Dataset (SPED). Arandjelović *et al.* [35] introduced a new layer based on a generalised Vector of Locally Aggregated Descriptors (VLAD) entitled NetVLAD, that can be incorporated in any CNN architecture for VPR training. The authors of [4] tested the performance of NetVLAD on multiple datasets, including: Berlin Kudamm, Gardens Point and Nordland datasets, showing its robust performance given various VPR scenarios. In Cross-Region-BoW [36], the authors identified the ROI of a query image using the CNN's layers acting as high-level feature extractors, then described these regions using low-level convolutional layers. Khaliq *et al.* [37] presents a lightweight VPR approach, based on ROI extraction combined with VLAD in order to achieve state-of-the-art performance under severe viewpoint and conditional variations. CALC [38] trained a Convolutional Auto-Encoder to output illumination-invariant Histogram-of-Oriented-Gradients (HOG) descriptors, where instead of using the original version of the image, laterally shifted and distorted versions of the image are used as input to output the same HOG descriptor for all distorted inputs. This results in a very light-weight system robust to variations in viewpoint and illumination. Torii *et al.* proposed in [39] a place recognition approach, entitled DenseVLAD, that successfully combines synthesis of novel virtual views with a densely sampled but compact image descriptor. Khaliq *et al.* present RegionVLAD [40], a light-weight CNN-based VPR technique that is able to detect salient features from images, while filtering out confusing elements.

The authors of [41] proposed a single-image and training-free VPR system, robust to appearance and viewpoint variation. Neubert and Protzel [42] present a novel local region detector, SP-Grid, which is able to ameliorate the effects of viewpoint variation. The work of Pepperell *et al.* [43] on SMART extended SeqSLAM by incorporating odometry into its calculations. More recently, the authors of [44] proposed a new sequence-based VPR system for aerial robots. In [45], the authors present a highly-scalable VPR pipeline that uses coarse scalar-quantisation based hashing. Sequence-based matching is used to resolve the collisions in the hash space. Johns and Yang [46] show a new method for appearance-based localisation, namely Feature Co-occurrence Maps. The authors of DeepSeqSLAM [11] proposed to integrate a RNN model on top of a CNN. The resulting system is successfully able to learn both visual and positional representations from a single monocular image sequence of a route. The authors of [47] propose a system with robust localisation, by creating a data association graph that is able to relate images from sequences. Vysotska and Stachniss presented in [48] a new approach based on graph-based image sequence matching that is swiftly able to retrieve the correspondences between a sequence of query and reference images, under severe appearance changes. The authors of [49] use a minicolumn network (MCN) approach that is able to create an internal representation that encodes sequential information.

As opposed to using VPR techniques with constant sequence lengths such as SLAM [10] and DeepSeqSLAM [11], our technique is able to adapt to each unique environment by employing a dynamic sequence-based matching approach. This is important as using an unchanged number of images in a sequence can either limit the performance of a VPR technique if the chosen sequence length is too small, or it can unnecessarily increase the computational load of the system, otherwise. ConvSequential-SLAM achieves state-of-the-art performance or comparable place matching performance in VPR related scenarios, while having a lower computational load which is important especially for resource constrained platforms.

III. METHODOLOGY

This section presents the methodology proposed in our work. The query images represent the visual data received from the camera, while the reference images represent the stored map of the environment in the form of RGB images. The block diagram showing each step of the ConvSequential-SLAM system is presented in Fig. 1.

A. INFORMATION GAIN

The first major innovation in our work is the ability of our technique to determine the information-gain resulted from analysing consecutive query images. This allows a more robust understanding of the environment, while it also gives enough information about different textures and properties found in successive query images. This approach is used to determine the local change-point in consecutive query images, thus enabling a minimum sequence length (*min_k*) for each sequence of images to be determined (see subsection III-E).

The information-gain is calculated as follows. Firstly, we compute the Histogram-of-Oriented-Gradients of the first and second query image that are part of a sequence. Secondly, we proceed to compare these two images together using regional convolutional matching (see subsection III-D), generating a similarity score. Finally, we compare this score with the Information Threshold (IT) to determine if the similarity between the two query images provides sufficient information gain. We then proceed to compare the first query image with the third and so on, repeating the above steps, until we find a representative minimum sequence length. The information gain can be easily summarised as in equation (1) and (2) below:

$$Information\ Gain = 1 - Similarity\ Score \tag{1}$$

$$Initial\ Seq = \begin{cases} k+1, & \text{if } Info\ Gain \geq IT. \\ entropy\ map, & \text{otherwise.} \end{cases} \tag{2}$$

In the above equation, $min\_k \leq k \leq max\_k\_IG$, IT represents the Information Threshold and it is in range [0,1], min_k is the minimum sequence length (set to 1) and max_k_IG is the maximum sequence length. The Initial Sequence Length (*Initial Seq*) in equation (2) represents the number of query images that are part of the query list generated by this approach. When the Information Gain module provides its best sequence length (e.g. $Info\ Gain < IT$), we proceed to calculate the sequential entropy (see subsection III-F) for that sequence of query images and determine whether this has the optimal length.

B. ENTROPY MAP AND ROI EXTRACTION

The second step in the ConvSequential-SLAM framework is to create the entropy map representing the salient regions in each query image. This is similar to [33], where the entropy map creation is based on estimating the local pixel intensity variation within the grayscale image and computing the base-2 logarithm of the histogram of pixel intensity values within each local region. This entropy map is represented by a matrix of size *W1*H1*, the elements of which are values in the range {0-8}, due to the pixel intensities in the range of 2^0 to $2^8 - 1$. The dimensions *W1*H1* represent the fixed size dimensions of the input image. The following matrix represents the entropy map for a query image:

$$Entropy = \begin{bmatrix} e_{11} & e_{12} & e_{13} & \dots & e_{1W_1} \\ e_{21} & e_{22} & e_{23} & \dots & e_{2W_1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ e_{H_1 1} & e_{H_1 2} & e_{H_1 3} & \dots & e_{H_1 W_1} \end{bmatrix} \quad \text{where } e_{ij} \in \{0-8\}. \tag{3}$$

Using the entropy map of an image, we extract Regions-of-Interest (ROIs) by computing the average entropy of a region of size *W2*H2*. If this entropy is above a threshold ET, it reflects that a region in informative and is selected as an ROI. The total number of regions (non-overlapping) in an image is $N = W1/W2 \times H1/H2$ and the total number of ROIs is G which can vary from one query image to another. Further details about the effect of ET and entropy map creation have been discussed at length in [33] and not provided here to avoid redundancy.

Moreover, in order to get a single entropy value for the entire image, we sum all the elements of the entropy matrix and divide them by *W1*H1*8* to get the re-scaled value. This is useful for the computation of sequential entropy of a sequence of query images to determine the dynamic sequence length (see subsection III-F).

C. REGIONAL HOG COMPUTATION

The process of regional HOG computation takes place as follows. In the first instance, we compute a gradient map of a grayscale image of size *W1*H1*. Following this, a histogram of oriented-gradients is computed for all N regions of the image, with each region having the size of *W2*H2*. Furthermore, each histogram of every region has L bins, where each bin is labelled with equally spaced gradient angles between 0-180 degrees. Lastly, we use L2-normalisation to achieve illumination invariance. This is done at a block level of size *(W2*2)*(H2*2)*.

D. REGIONAL CONVOLUTIONAL MATCHING

Following the regional HOG computation, we proceed to regional convolutional matching, given each query image is represented as N regions, each being described by a HOG-descriptor of depth $4*L$. Using the information from the Region of Interest (ROI) evaluation, these N regions are reduced to G salient regions. By doing so, the query image HOG-descriptor can be represented as a 2D matrix of dimensions $[G, 4*L]$. The reference image has N regions with the descriptor size of $4*L$, therefore its resulting matrix has the dimensions of $[N, 4*L]$. We then proceed to multiply the query and reference matrices, and the result is a matrix of dimensions $[G, N]$. Each row of this matrix represents a salient region of a query image, while each column represents the cosine-matching scores for that region with all the N regions of a reference image. Max-pooling is used across the rows of the aforementioned matrix in order to determine the best matched regions between the query and reference images. The final score is computed as the arithmetic mean of matching scores of all G regions and is in the range of 0 - 1, such that the higher the score, the higher the similarity between the two images. Finally, the reference image that has the highest score is chosen to be the best match for a given query image.

E. CREATING THE QUERY IMAGES SEQUENCE

Query images are added into a 1D list in a sequential manner, such that the length of this list is dependent on the sequential entropy (explained in subsection III-F). Even if the sequential entropy's value for the first k images is higher than the Entropy Threshold (ET), where $0 \leq ET \leq 1$, the minimum sequence length will be determined using the information-gain resulted from analysing consecutive query images (see subsection III-A). Thus, we will not end up with non-optimal sequence lengths, that will ultimately result in poor performance. The 1D query list containing a sequence of query images is represented as:

$$Sequential\ Query\ List = \begin{bmatrix} q_1 & q_2 & q_3 & \dots & q_k \end{bmatrix} \tag{4}$$

In the above equation q_1 is the first query image, q_k is the last query image, and k is the total number of images that are part of a sequence. As previously mentioned, the length of this list will constantly change, but all the images will be in a sequential order, starting from the first image to the *k-th* image. When computing the second sequence of query images, we start with the second image (q_2) and so on. It is important to note that for any N images read, the number of query images sequence lists created will be $N - k + 1$, where k will contain the length of the last list created. That is, for

Algorithm 1 Matching Query and Reference Sequences

```
Given: Query Images Sequence (Q_Seq)
Given: Reference_Images_List (R_List)
ref_matching_scores = [ ]
iterator = 0
k = Length (Q_Seq)
while itr + k ≤ Length(R_List) do
    Sequential_Reference_List = R_Seq = [ ]
    for ref_itr in range(itr, itr + k) do
        APPEND R_List[ref_itr] to R_Seq
    match_score = Sequence_Matching_Func(Q_Seq,
    R_Seq)
    ADD match_score to ref_matching_scores
    iterator = iterator + 1
Best Match = Max (ref_matching_scores)
```

any N query images, the algorithm will only match the first N - $k + 1$ images.

F. ENTROPY-BASED DYNAMIC QUERY IMAGES SEQUENCE

The second key innovation is incorporating the ability of our technique to reuse entropy as measure of the overall information content found in a sequence of query images, to decide an optimal sequence length of each query list. Building upon the sequence length generated by analysing consecutive query images (see subsection III-A), we use the entropy to maximize the efficiency of this length. To achieve this, our technique first looks at the information content (entropy score) of the query sequence list generated in subsection III-A (first *min_k_IG* query images). If the information content within this sequence of images is less than a threshold (ET), we increase the sequence length by a constant step, then recompute the information content for this new increased sequence of images. If the information content (*Seq Entropy*) for this increased sequence of images reaches a reasonable value (ET), the corresponding length of the query images sequence is used, otherwise we keep increasing it (up till maximum sequence length) to find a suitable sequence length. *Seq Entropy* represents the arithmetic mean of the entropy scores of the query images within the sequence. The entire iterative process is summarised in equation (5) below:

$$Seq\ Length = \begin{cases} min\_k\_IG, & \text{if } Seq\ Entropy \geq ET. \\ k + 1, & \text{otherwise.} \end{cases} \quad (5)$$

In the above equation $min\_k\_IG \leq k \leq max\_k$, ET represents the Entropy Threshold, *min_k_IG* is the minimum sequence length (generated in subsection III-A) and *max_k* is the maximum sequence length. The Sequence Length (*Seq Length*) in equation (5) represents the number of images that are part of the query list at a given time, thus being dependent on the value of k. In the same equation, the Sequential Entropy (*Seq Entropy*) refers to the average entropy value (see subsection III-B) of k images that are part of this query list.

G. DYNAMIC SEQUENCE MATCHING

This subsection presents how we achieve the matching between dynamic sequence length of images. As discussed in subsection III-F, our technique creates a dynamic list of query images, i.e., the length of the query sequence list will vary for different sets of query images. During the matching phase, we create a sequential 1D reference list of the same length as the sequential query list. These sequential 1D reference lists are created for all the images in the reference map. Because the size of our reference list is dependent on the sequential query list's length, this simplifies the matching of the query and reference image sequences. The algorithm that retrieves a correct match for a sequence of query images given a reference map can be found in Algorithm 1. The function *Sequence_Matching_Func* in Algorithm 1 takes k corresponding pairs (1-to-1 matching) from the query image sequence (*Q_Seq*) and reference image sequence (*R_Seq*) and matches them using Regional Convolutional Matching, as explained in subsection III-D. The matching score of the query and reference sequences is the arithmetic mean of the matching scores of the pairs within these sequences. This function returns the matching score of the query image sequence and the reference image sequence. Given all the reference images and their corresponding sequences from the reference map, the sequence with the highest matching score is selected as the best match.

IV. EXPERIMENTAL SETUP

A. SEQUENTIAL DATASETS

To evaluate the proposed technique, we have used the following public VPR datasets: Gardens Point dataset [50] containing images taken from different angles (viewpoint variation). This dataset consists of a total of 600 images, divided into 200 query images (day images) and 400 reference images equally divided into both day and night images. In this paper, we have used *day left* as query images and both *day right* and *night right* as reference images; Nordland dataset [51] containing drastic appearance changes of a place in different seasons (spring, summer, autumn and winter). We have used 172 query images taken form the summer dataset and 172 reference images taken from the winter dataset; Campus Loop dataset [38] containing viewpoint variation, seasonal variation and also the presence of statically-occluded frames. This dataset has 100 query and 100 reference images. Apart from using these datasets to show the performance of our technique, we also use the Alderley (night-to-day) dataset solely to show the variation in sequence length due to sequential entropy. This sequential dataset consists of 201 query images (night images) and 201 reference images (day images).

B. STATE-OF-THE-ART VPR TECHNIQUES

We compare the performance of ConvSequential-SLAM with other VPR techniques, such as CoHOG [33], HOG [31], CALC [38], HybridNet [7], AMOSNet [7], SeqSLAM [10],

RegionVLAD [40], NetVLAD [35] and DeepSeqSLAM [11] on the datasets mentioned in Subsection IV-A. We have used SeqSLAM with a sequence length of 5 and 10 images respectively, while DeepSeqSLAM was tested with a sequence length of 10 images only. The remaining VPR techniques are single-image-based and are provided for completeness.

C. PARAMETERS

In this work, we have used W1 = H1 = 512, W2 = H2 = 16, L = 8 bins, G (can take different values for different query images depending on the scene that is represented), $ET = 0.5$, $IT = 0.9$, $min\_k = 1$, $max\_k\_IG = 15$ and $max\_k = 25$ for ConvSequential-SLAM. These values represent the backbone of our system, as they are responsible for determining the optimal sequence length. The above values were specifically chosen as they provide overall good results (in terms of Accuracy, AUC-PR and PCU) while also providing comparable results to our static sequence length ($k = 10$ images) version of ConvSequential-SLAM. An ablation study showing the performance of our technique in terms of accuracy and AUC-PR with various sequence lengths ($1 \leq k \leq 20$) is provided later in this paper.

D. PERFORMANCE METRICS

The authors of [2] suggested that Precision-Recall curves are a key evaluation metric for VPR techniques. Therefore, an ideal system would achieve 100% precision at 100% recall. The authors of [52] used the Area-under-the-Precision-Recall-Curve (AUC-PR) to compare the performance of VPR techniques which has therefore been adopted in our work as well. AUC-PR is computed by plotting the Precision-Recall curve at different confidence thresholds.

In [52], the authors highlighted that an ideal AUC-PR score of 1 is achievable even if the system contains false-positives and therefore suggested that the accuracy A (refer to equation (6)) of a system should also be provided. Moreover, AUC-PR is only focusing on the matching performance of a given VPR system, thus not incorporating the computational intensity of that technique. This is essential in real-world scenarios, where for resource-constrained platforms, the matching performance has to be directly related with the computational intensity. For this reason, the authors of [4], [5], [37] and [38] determined the feature encoding time (t_e) to be an important performance indicator. In [33], the Performance-per-Compute-Unit (PCU) is defined by combining P_{R100} with t_e as in equation (7):

$$A = \frac{Total\ No.\ of\ Correctly\ Matched\ Query\ Images}{Total\ No.\ of\ Query\ Images\ in\ Database} \tag{6}$$

$$PCU = P_{R100} \times \log\left(\frac{t_{e\_max}}{t_e} + 9\right) \tag{7}$$

As it can be seen in the above equation, higher precision will always lead to a higher PCU value. The maximum feature encoding time (t_{e_max}) is chosen to represent the most resource intensive VPR technique. In our case, this is NetVLAD, with the feature encoding time of $t_{e\_max} = 0.77$ sec. Fig. 4 shows the feature encoding time (t_e) for various VPR technique tested. The scalar 9 is added in equation (7) so that the VPR technique that has $t_e = t_{e\_max}$ does not have a PCU of 0, but instead that it provides a more interpretable range. We present the PCU of ConvSequential-SLAM in Fig. 3 and discuss it further in subsection V-C.

TABLE 1. The AUC-PR of VPR techniques on the 4 datasets.

VPR Technique	AUC-PR			
	Campus Loop Dataset	*Gardens Point (day-to-day) Dataset*	*Gardens Point (day-to-night) Dataset*	*Nordland Dataset*
Ours Static k	0.999	1	0.754	0.533
Ours Dynamic k	1	1	0.8	0.6
CoHOG	0.776	0.928	0.43	0.151
HOG	0.301	0.431	0.294	0.036
CALC	0.597	0.738	0.403	0.104
HybridNet	0.889	0.933	0.595	0.214
AMOSNet	0.872	0.907	0.571	0.132
SeqSLAM k = 5	0.273	0.296	0.037	0.059
SeqSLAM k = 10	0.371	0.333	0.071	0.16
RegionVLAD	0.412	0.739	0.642	0.563
NetVLAD	0.998	0.959	0.698	0.733
DeepSeqSLAM	0.999	1	0.952	0.736

V. RESULTS AND ANALYSIS

In this section, we discuss results from a place matching performance point, in terms of accuracy, AUC-PR and PCU. We also present the performance of ConvSequential-SLAM for various sequence lengths and show how the sequence length varies between one dataset and another. Finally, we show some samples of correctly and incorrectly retrieved query and reference images by our technique for a qualitative insight. For all experiments presented below, we have used a PC equipped with an Intel Core i7-4790k CPU.

A. ACCURACY

This subsection presents the accuracy results of ConvSequential-SLAM against the performance of other VPR techniques. Fig. 2 shows the computed values of accuracy for all techniques, on all 4 datasets. As all the datasets tested contain consecutive images, there is a high possibility that each image is similar to the ones located in its immediate proximity. Therefore, if for any query image, the reference image found to be the best match is in the range ± 2, we consider it as a correct match, except for the Nordland dataset where we use the ± 1 range.

ConvSequential-SLAM achieves state-of-the-art accuracy on all datasets utilised in this work. Our approach also achieves state-of-the-art performance on the highly conditionally-variant Gardens Point (day-to-night) and Nordland datasets, followed by state-of-the-art deep learning-based techniques like NetVLAD, HybridNet and AMOSNet. However, DeepSeqSLAM with a sequence length of 10 images outperforms every other VPR technique on both Gardens Point (day-to-night) and Nordland datasets.

B. AREA-UNDER-THE-PRECISION-RECALL-CURVE

The performance of ConvSequential-SLAM in terms of AUC-PR on all datasets is reported in Table 1. It achieves

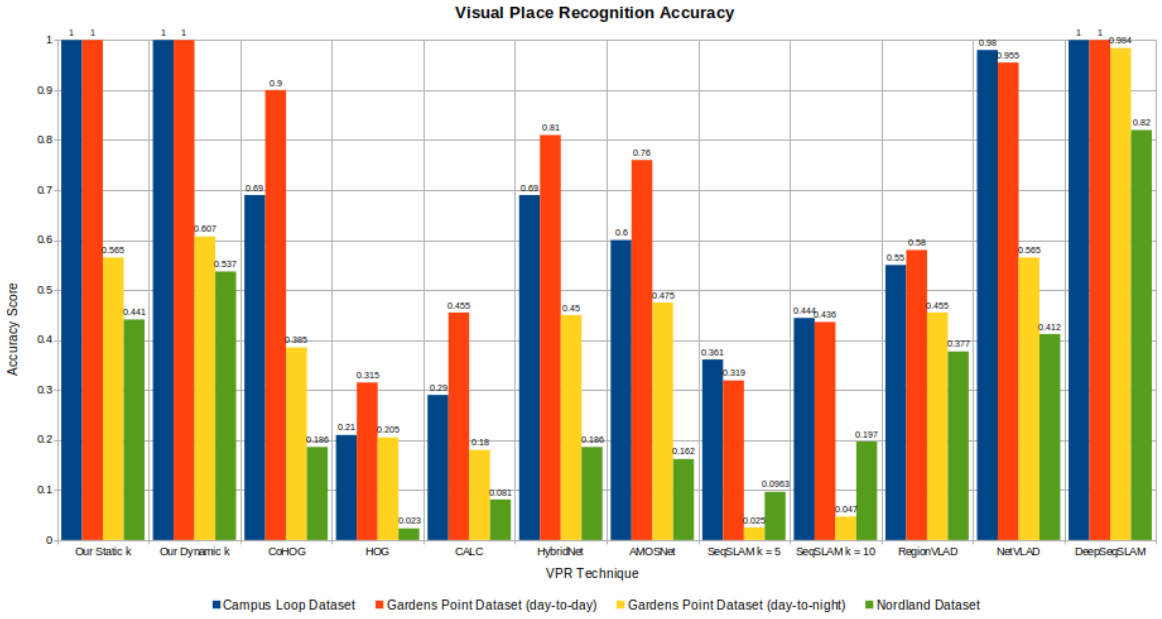

FIGURE 2. The accuracy of ConvSequential-SLAM (using a fixed sequence length of 10 images as well as a dynamic length determined by the system itself) is compared against the accuracy of other state-of-the-art VPR techniques on widely used public VPR datasets.

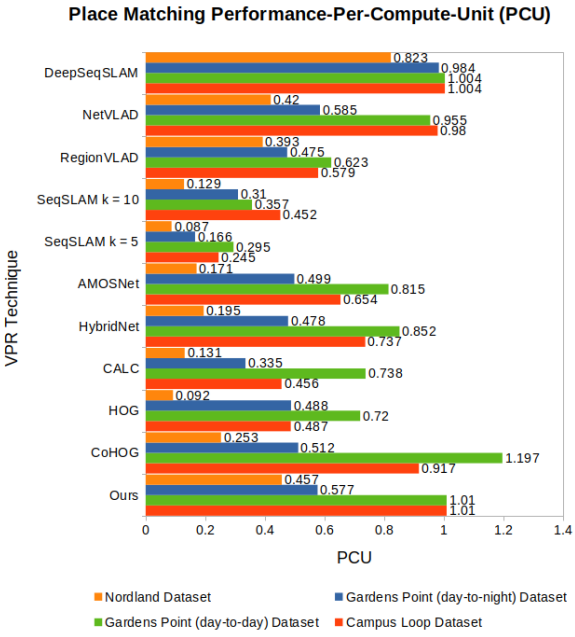

FIGURE 3. The PCU of ConvSequential-SLAM is compared with the PCU of other contemporary VPR techniques on all mentioned datasets.

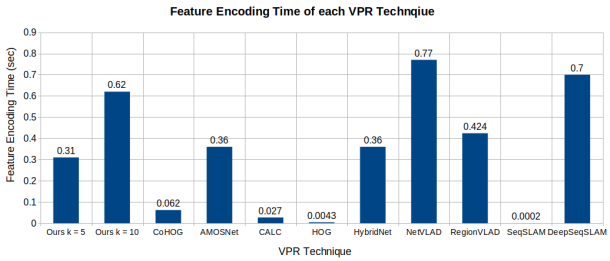

FIGURE 4. The feature encoding time of various VPR technique are presented in this graph. *Our* $k = 5$ and $k = 10$ represent the feature encoding time of ConvSequential-SLAM using fixed sequences of 5 and 10 images respectively.

state-of-the-art AUC-PR performance on the Campus Loop, Gardens Point (day-to-day), and Gardens Point (day-to-night) datasets. When compared to NetVLAD, ConvSequential-SLAM achieves better performance on all datasets tested, except on Nordland dataset, as reported in Table 1. We can see a small boost in performance between our algorithm using a fixed sequence length of 10 images and a dynamic sequence length respectively. When compared to DeepSeqSLAM, the proposed technique achieves comparable results on both Campus Loop and Gardens Point (day-to-day) datasets. However, DeepSeqSLAM outperforms all VPR techniques on the remaining two datasets (Gardens Point (day-to-night) and Nordland). In Fig. 5, we present the Precision-Recall curves of all the VPR techniques tested in our work on all 4 datasets introduced in Section IV.

C. PERFORMANCE-PER-COMPUTE-UNIT (PCU)

Fig. 3 presents the PCU of ConvSequential-SLAM using a fixed sequence length of 10 images. Because we match sequences of images instead of single frames, the feature encoding time will also be increased with each image that is part of that sequence, as shown in Fig. 4. The feature encoding time for dynamic k will vary between the lowest value (that is for a minimum sequence length determined by the information-gain) and the maximum value for a sequence length of 25 images.

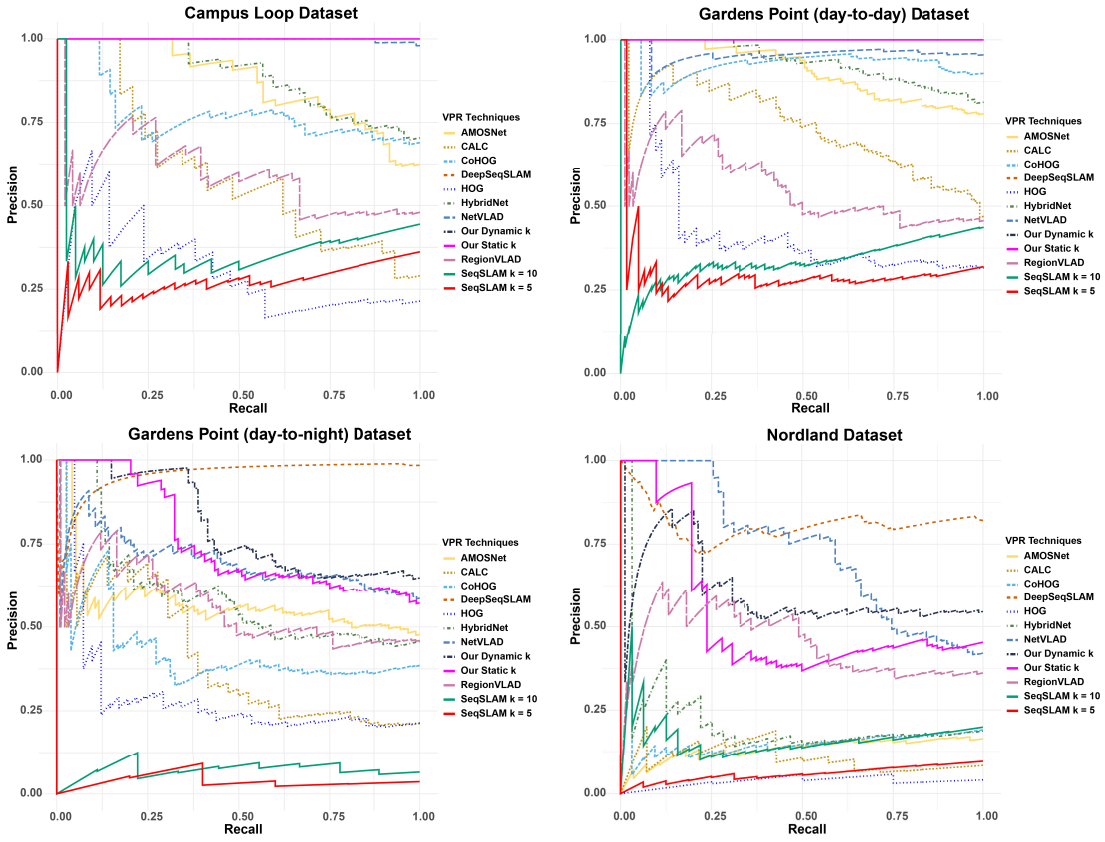

FIGURE 5. The precision-recall curves for all VPR techniques on each of the 4 datasets used in our work are enclosed here.

In this subsection, we use the average sequence length computed by our methodology within the dataset for encoding time computation. Even though using a higher sequence length will result in higher encoding times, the performance boost in place matching that is gained greatly benefits ConvSequential-SLAM. This can be seen in Fig. 3, where we present the PCU value of our technique against other VPR techniques, on all 4 datasets. As mentioned in subsection IV-D, a system that achieves high precision will have a high PCU value. This is also the case for ConvSequential-SLAM, which achieves high PCU values due to its high precision.

D. VARIATION IN SEQUENCE LENGTH

It is well known the fact that by incorporating sequence-based filtering into a VPR system, the overall performance is greatly improved. However, sequence matching requires a static sequence length to be provided for each environment that the robotic platform is operating in. Furthermore, this sequence length cannot always be constant as different VPR techniques have different place matching performances. This is an important factor especially for resource constrained platforms, in which the computational intensity of a VPR technique needs to be carefully considered. We show that ConvSequential-SLAM is successfully able to adapt its sequence length depending on the environment.

Fig. 6 presents the different sequence lengths (k) that can be achieved when the program was run on all datasets. It is important to note that by varying both the Entropy Threshold (ET) and Information Threshold (IT) we can achieve lower or higher sequence lengths. Also, it is worth noting that we use day images as query images for both Gardens Point (day-to-day) and Gardens Point (day-to-night) datasets, so we only include one instance of the dataset in Fig. 6 in order to avoid redundancy.

In addition to the datasets mentioned above, we also use the Alderley (night-to-day) dataset to show how the sequence length is modifying because of entropy. However, all VPR techniques poorly perform on this dataset, because the query images (night images) provide little to no information about the environment. This is due to the poor lighting condition in the environment as well as the presence of rain, which increase the difficulty in place matching.

Because in the Campus Loop, Gardens Point and Nordland datasets the entropy across each dataset is too high, the sequence length would not have increased in most cases, therefore we would end up with non-optimal sequence lengths. By using information-gain resulted from analysing consecutive query images, we are able to increase the minimum sequence length even though the salient information found in any given query image is above the threshold set (e.g. $ET \geq 0.5$). However, in contrast with the

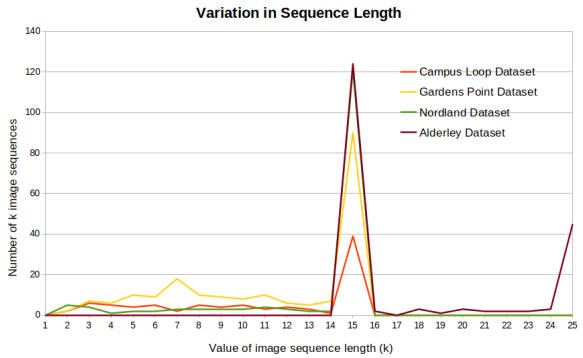

FIGURE 6. The variation in sequence length of ConvSequential-SLAM on all four datasets is shown here.

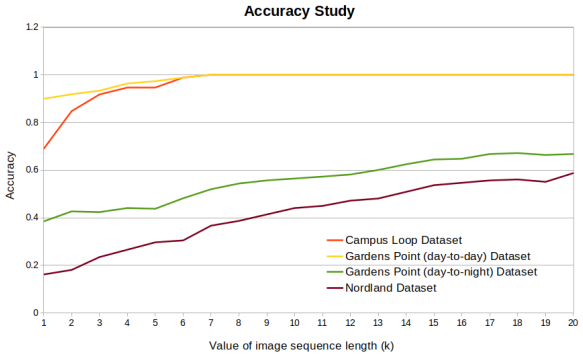

FIGURE 7. The ablation study showing the accuracy of ConvSequential-SLAM for different k values ($1 \leq k \leq 20$).

previously mentioned datasets where the sequence length would not increase due to high information content in query images, on the Alderley dataset query images (night images) do not contain salient information due to poor illumination, therefore the sequence length will increase up to the maximum sequence length of 25 as shown in Fig. 6.

E. ABLATION STUDY

Fig. 7 and Fig. 8 presents the performance of our approach in terms of Accuracy and AUC-PR values, when using dynamic length of images. The program was tested with fixed k lengths between 1 and 20 images respectively. Increasing the value of k leads to an increase in both Accuracy and AUC-PR performance. In both Campus Loop and Gardens Point (day-to-day) datasets, a sequence length of $k = 7$ images will result in the best performance, whilst a higher sequence length is needed in order to achieve desirable results for the remaining two datasets: Gardens Point (day-to-night) and Nordland.

F. EXEMPLAR MATCHES

Fig. 9 shows some correctly matched sequences of query and reference images, taken from each dataset. Some failure cases for the proposed technique are shown in Fig. 10. These are primarily due to the presence of confusing features coming from trees and vegetation that can be found in most images

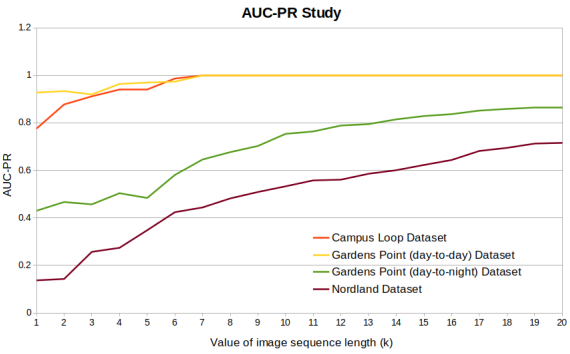

FIGURE 8. The ablation study showing the AUC-PR of ConvSequential-SLAM for different k values ($1 \leq k \leq 20$).

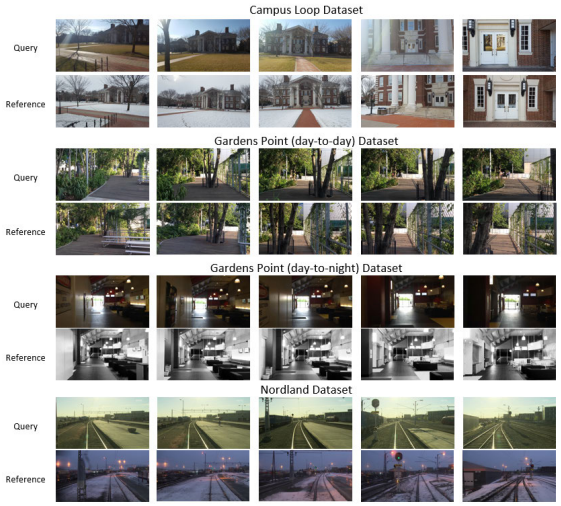

FIGURE 9. Some correctly matched sequences of query and reference frames.

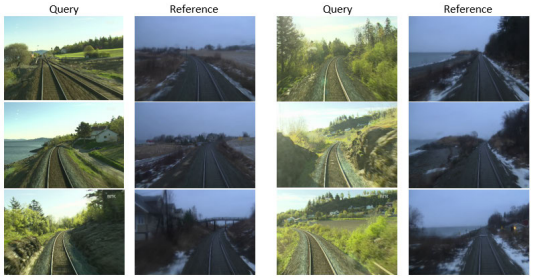

FIGURE 10. Some incorrectly matched query and reference frames.

throughout the Nordland dataset, increasing the difficulty in place matching.

VI. CONCLUSION AND FUTURE DIRECTIONS

We have presented ConvSequential-SLAM, a system that successfully performs VPR in challenging environments, with zero training requirements. The proposed technique achieves state-of-the-art performance on public VPR datasets that contain both viewpoint and appearance variations.

A possible future direction for improving this work is to cater for dynamic objects and confusing features coming from trees, vegetation etc. in outdoor environment.

REFERENCES

[1] M. Zaffar, S. Ehsan, M. Milford, and K. D. McDonald-Maier, "Memorable maps: A framework for re-defining places in visual place recognition," *IEEE Trans. Intell. Transp. Syst.*, early access, Jun. 25, 2020, doi: 10.1109/TITS.2020.3001228.

[2] S. Lowry, N. Sünderhauf, P. Newman, J. J. Leonard, D. Cox, P. Corke, and M. J. Milford, "Visual place recognition: A survey," *IEEE Trans. Robot.*, vol. 32, no. 1, pp. 1–19, Feb. 2016.

[3] Z. Chen, O. Lam, A. Jacobson, and M. Milford, "Convolutional neural network-based place recognition," 2014, *arXiv:1411.1509*. [Online]. Available: http://arxiv.org/abs/1411.1509

[4] M. Zaffar, A. Khaliq, S. Ehsan, M. Milford, and K. McDonald-Maier, "Levelling the playing field: A comprehensive comparison of visual place recognition approaches under changing conditions," 2019, *arXiv:1903.09107*. [Online]. Available: http://arxiv.org/abs/1903.09107

[5] M. Zaffar, A. Khaliq, S. Ehsan, M. Milford, K. Alexis, and K. McDonald-Maier, "Are state-of-the-art visual place recognition techniques any good for aerial robotics?" 2019, *arXiv:1904.07967*. [Online]. Available: http://arxiv.org/abs/1904.07967

[6] W. Maddern, G. Pascoe, C. Linegar, and P. Newman, "1 year, 1000 km: The Oxford RobotCar dataset," *IJ Robot. Res.*, vol. 36, no. 1, pp. 3–15, 2016.

[7] Z. Chen, A. Jacobson, N. Sünderhauf, B. Upcroft, L. Liu, C. Shen, I. Reid, and M. Milford, "Deep learning features at scale for visual place recognition," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2017, pp. 3223–3230.

[8] A. Torii, J. Sivic, T. Pajdla, and M. Okutomi, "Visual place recognition with repetitive structures," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 883–890.

[9] F. Maffra, L. Teixeira, Z. Chen, and M. Chli, "Real-time wide-baseline place recognition using depth completion," *IEEE Robot. Autom. Lett.*, vol. 4, no. 2, pp. 1525–1532, Jan. 2019.

[10] M. J. Milford and G. F. Wyeth, "SeqSLAM: Visual route-based navigation for sunny summer days and stormy winter nights," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2012, pp. 1643–1649.

[11] M. Chancán and M. Milford, "DeepSeqSLAM: A trainable CNN+RNN for joint global description and sequence-based place recognition," 2020, *arXiv:2011.08518*. [Online]. Available: http://arxiv.org/abs/2011.08518

[12] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2, Sep. 1999, pp. 1150–1157.

[13] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded up robust features," *Comput. Vis. Image Understand.*, vol. 110, pp. 404–417, Jan. 2006.

[14] S. Se, D. Lowe, and J. Little, "Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks," *Int. J. Robot. Res.*, vol. 21, no. 8, pp. 735–758, 2002.

[15] H. Andreasson and T. Duckett, "Topological localization for mobile robots using omni-directional vision and local features," *IFAC Proc. Volumes*, vol. 37, no. 8, pp. 36–41, Jul. 2004.

[16] E. Stumm, C. Mei, and S. Lacroix, "Probabilistic place recognition with covisibility maps," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Nov. 2013, pp. 4158–4163.

[17] J. Košecká, F. Li, and X. Yang, "Global localization and relative positioning based on scale-invariant keypoints," *Robot. Auton. Syst.*, vol. 52, no. 1, pp. 27–38, 2005.

[18] A. C. Murillo, J. J. Guerrero, and C. Sagues, "SURF features for efficient robot localization with omnidirectional images," in *Proc. IEEE Int. Conf. Robot. Autom.*, Apr. 2007, pp. 3901–3907.

[19] M. Cummins and P. Newman, "Appearance-only SLAM at large scale with FAB-MAP 2.0," *Int. J. Robot. Res.*, vol. 30, no. 9, pp. 1100–1123, Jun. 2011.

[20] W. Maddern, M. Milford, and G. Wyeth, "CAT-SLAM: Probabilistic localisation and mapping using a continuous appearance-based trajectory," *Int. J. Robot. Res.*, vol. 31, no. 4, pp. 429–451, Apr. 2012.

[21] M. Agrawal, K. Konolige, and M. R. Blas, "CenSurE: Center surround extremas for realtime feature detection and matching," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2008, pp. 102–115.

[22] K. Konolige and M. Agrawal, "FrameSLAM: From bundle adjustment to real-time visual mapping," *IEEE Trans. Robot.*, vol. 24, no. 5, pp. 1066–1077, Oct. 2008.

[23] J. Sivic and A. Zisserman, "Video Google: A text retrieval approach to object matching in videos," in *Proc. 9th IEEE Int. Conf. Comput. Vis.*, Oct. 2003, p. 1470.

[24] L. Fei-Fei and P. Perona, "A Bayesian hierarchical model for learning natural scene categories," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2, Jun. 2005, pp. 524–531.

[25] A. Angeli, S. Doncieux, J.-A. Meyer, and D. Filliat, "Incremental vision-based topological SLAM," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Sep. 2008, pp. 1031–1036.

[26] A. Oliva and A. Torralba, "Building the gist of a scene: The role of global image features in recognition," in *Visual Perception*, vol. 155, S. Martinez-Conde, S. Macknik, L. Martinez, J.-M. Alonso, and P. Tse, Eds. Amsterdam, The Netherlands: Elsevier, 2006, ch. 2, pp. 23–36.

[27] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *Int. J. Comput. Vis.*, vol. 42, no. 3, pp. 145–175, 2001.

[28] A. C. Murillo and J. Kosecka, "Experiments in place recognition using gist panoramas," in *Proc. IEEE 12th Int. Conf. Comput. Vis. Workshops (ICCV Workshops)*, Sep. 2009, pp. 2196–2203.

[29] G. Singh and J. Kosecka, "Visual loop closing using gist descriptors in Manhattan world," in *Proc. Omnidirectional Vis. Workshop (ICRA)*, 2010, pp. 4042–4047.

[30] C. Siagian and L. Itti, "Biologically inspired mobile robot vision localization," *IEEE Trans. Robot.*, vol. 25, no. 4, pp. 861–873, Aug. 2009.

[31] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1, Jun. 2005, pp. 886–893.

[32] C. McManus, B. Upcroft, and P. Newmann, "Scene signatures: Localised and point-less features for localisation," *Proc. Robot.: Sci. Syst*, vol. 39, no. 3, pp. 1–9, Jul. 2014.

[33] M. Zaffar, S. Ehsan, M. Milford, and K. McDonald-Maier, "CoHOG: A light-weight, compute-efficient, and training-free visual place recognition technique for changing environments," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 1835–1842, Apr. 2020.

[34] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "OverFeat: Integrated recognition, localization and detection using convolutional networks," 2013, *arXiv:1312.6229*. [Online]. Available: http://arxiv.org/abs/1312.6229

[35] R. Arandjelovic, P. Gronat, A. Torii, T. Pajdla, and J. Sivic, "NetVLAD: CNN architecture for weakly supervised place recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 5297–5307.

[36] Z. Chen, F. Maffra, I. Sa, and M. Chli, "Only look once, mining distinctive landmarks from ConvNet for visual place recognition," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2017, pp. 9–16.

[37] A. Khaliq, S. Ehsan, Z. Chen, M. Milford, and K. McDonald-Maier, "A holistic visual place recognition approach using lightweight CNNs for significant ViewPoint and appearance changes," *IEEE Trans. Robot.*, vol. 36, no. 2, pp. 561–569, Apr. 2020.

[38] N. Merrill and G. Huang, "Lightweight unsupervised deep loop closure," 2018, *arXiv:1805.07703*. [Online]. Available: http://arxiv.org/abs/1805.07703

[39] A. Torii, R. Arandjelovic, J. Sivic, M. Okutomi, and T. Pajdla, "24/7 place recognition by view synthesis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1808–1817.

[40] A. Khaliq, S. Ehsan, M. Milford, and K. McDonald-Maier, "CAMAL: Context-aware multi-layer attention framework for lightweight environment invariant visual place recognition," 2019, *arXiv:1909.08153*. [Online]. Available: http://arxiv.org/abs/1909.08153

[41] N. Sünderhauf, S. Shirazi, A. Jacobson, F. Dayoub, E. Pepperell, B. Upcroft, and M. Milford, "Place recognition with convnet landmarks: Viewpoint-robust, condition-robust, training-free," *Robot., Sci. Syst. XI*, vol. 33, no. 9, pp. 1–10, Jul. 2015.

[42] P. Neubert and P. Protzel, "Beyond holistic descriptors, keypoints, and fixed patches: Multiscale superpixel grids for place recognition in changing environments," *IEEE Robot. Autom. Lett.*, vol. 1, no. 1, pp. 484–491, Jan. 2016.

[43] E. Pepperell, P. I. Corke, and M. J. Milford, "All-environment visual place recognition with SMART," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2014, pp. 1612–1618.

[44] M. Yang, J. Mao, X. He, L. Zhang, and X. Hu, "A sequence-based visual place recognition method for aerial mobile robots," *J. Phys., Conf. Ser.*, vol. 1654, Oct. 2020, Art. no. 012080.
[45] S. Garg and M. Milford, "Fast, compact and highly scalable visual place recognition through sequence-based matching of overloaded representations," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2020, pp. 3341–3348.
[46] E. Johns and G.-Z. Yang, "Feature co-occurrence maps: Appearance-based localisation throughout the day," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2013, pp. 3212–3218.
[47] T. Naseer, L. Spinello, W. Burgard, and C. Stachniss, "Robust visual robot localization across seasons using network flows," in *Proc. 28th AAAI Conf. Artif. Intell.*, 2014, pp. 2564–2570.
[48] O. Vysotska and C. Stachniss, "Relocalization under substantial appearance changes using hashing," in *Proc. Workshop Planning, Perception Navigat. Intell. Vehicles (IROS)*, Vancouver, BC, Canada, vol. 24, 2017.
[49] P. Neubert, S. Schubert, and P. Protzel, "A neurologically inspired sequence processing model for mobile robot place recognition," *IEEE Robot. Autom. Lett.*, vol. 4, no. 4, pp. 3200–3207, Oct. 2019.
[50] N. Sunderhauf, S. Shirazi, F. Dayoub, B. Upcroft, and M. Milford, "On the performance of ConvNet features for place recognition," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2015, pp. 4297–4304.
[51] S. Skrede. (2013). *Nordland Dataset*. [Online]. Available: https://nrkbeta.no/2013/01/15/nordlandsbanen-minute-by-minute-season-by-season/
[52] M. Zaffar, S. Ehsan, M. Milford, D. Flynn, and K. McDonald-Maier, "VPR-bench: An open-source visual place recognition evaluation framework with quantifiable viewpoint and appearance change," 2020, *arXiv:2005.08135*. [Online]. Available: http://arxiv.org/abs/2005.08135

MIHNEA-ALEXANDRU TOMIȚĂ received the B.Sc. degree in computer science from the University of Essex, Colchester, U.K., in 2019, where he is currently pursuing the Ph.D. degree with a part of the National Centre for Nuclear Robotics (NCNR).

His current research interests include computer vision, sequence-based filtering, deep learning, and SLAM.

MUBARIZ ZAFFAR received the B.E. degree in electrical engineering from the National University of Sciences and Technology (NUST), Pakistan, in 2016, and the M.Sc. degree in computer science and electronic engineering from the University of Essex, U.K., in 2020. He is currently pursuing the Ph.D. degree with the Cognitive Robotics Department, TU Delft, where he is a part of the Intelligent Vehicles Group.

His research interests include computer vision and deep learning for autonomous robotics, visual place recognition and robot navigation, SLAM, and embedded systems.

Mr. Zaffar was a recipient of the South-Asian Helix Innovation Award, the DICE Foundation Innovation Award, the IET Present-Around-The-World Regional Awards, and the NUST High Achiever's Award.

MICHAEL J. MILFORD (Senior Member, IEEE) received the Bachelor of Mechanical and Space Engineering degree and the Ph.D. degree in electrical engineering from The University of Queensland, Brisbane, QLD, Australia.

He was a Research Fellow on the Thinking Systems Project with Queensland Brain Institute, until 2010, when he became a Lecturer with Queensland University of Technology (QUT), Brisbane. He is currently an Associate Professor and an Australian Research Council Future Fellow with QUT and the Chief Investigator of the Australian Centre of Excellence for Robotic Vision. He conducts interdisciplinary research into navigation across the fields of robotics, neuroscience, and computer vision.

Dr. Milford was a recipient of an inaugural Australian Research Council Discovery Early Career Research Award, in 2012, and became a Microsoft Research Faculty Fellow, in 2013.

KLAUS D. MCDONALD-MAIER (Senior Member, IEEE) received the Dipl.-Ing. degree in electrical engineering from the University of Ulm, Ulm, Germany, the M.S. degree in electrical engineering from CPE Lyon, Villeurbanne, France, in 1995, and the Ph.D. degree in computer science from Friedrich Schiller University Jena, Germany, in 1999.

He was a Systems Architect on reusable micro-controller cores and modules with Infineon Technologies AG's Cores and Modules Division, Munich, Germany, and a Lecturer of electronics engineering with the University of Kent, Canterbury, U.K. In 2005, he joined the University of Essex, Colchester, U.K., where he is currently a Professor with the School of Computer Science and Electronic Engineering. His current research interests include embedded systems and system-on-a-chip design, security, development support and technology, parallel and energy efficient architectures, and the application of soft computing and image processing techniques for real-world problems.

Dr. McDonald-Maier is a member of the Verband der Elektrotechnik Elektronik Informationstechnik and the British Computer Society, and a fellow of the Institution of Engineering and Technology.

SHOAIB EHSAN (Senior Member, IEEE) received the B.Sc. degree in electrical engineering from the University of Engineering and Technology, Taxila, Pakistan, in 2003, and the Ph.D. degree in computing and electronic systems, with a focus on computer vision, from the University of Essex, Colchester, U.K., in 2012.

He has extensive industrial and academic experience in the areas of embedded systems, embedded software design, computer vision, and image processing. His current research interests include intrusion detection for embedded systems, local feature detection and description techniques, image feature matching, and performance analysis of vision systems.

Dr. Ehsan was a recipient of the University of Essex Post Graduate Research Scholarship and the Overseas Research Student Scholarship. He was a winner of the prestigious Sullivan Doctoral Thesis Prize by the British Machine Vision Association.