Image-Based Place Recognition on Bucolic Environment Across Seasons From Semantic Edge Description

Assia Benbihi\textsuperscript{1,2}, Matthieu Geist\textsuperscript{3}, and Cédric Pradalier\textsuperscript{1,4}

\textsuperscript{1}UMI2958 GeorgiaTech-CNRS
\textsuperscript{2}Centrale Supélec, Université Paris Saclay, Metz
\textsuperscript{3}Google Research, Brain Team
\textsuperscript{4}GeorgiaTech Lorraine

Abstract

Most of the research effort on image-based place recognition is designed for urban environments. In bucolic environments such as natural scenes with low texture and little semantic content, the main challenge is to handle the variations in visual appearance across time such as illumination, weather, vegetation state or viewpoints. The nature of the variations is different and this leads to a different approach to describing a bucolic scene. We introduce a global image descriptor computed from its semantic and topological information. It is built from the wavelet transforms of the image semantic edges. Matching two images is then equivalent to matching their semantic edge descriptors. We show that this method reaches state-of-the-art image retrieval performance on two multi-season environment-monitoring datasets: the CMU-Seasons and the Symphony Lake dataset. It also generalises to urban scenes on which it is on par with the current baselines NetVLAD and DELF.

1 INTRODUCTION

Place recognition is the process by which a place that has been observed before can be identified when revisited. Image-based place recognition achieves this task using images taken with similar viewpoints at different times. This is particularly challenging for images captured in natural environments over multiple seasons (e.g. [1] or [2]) because their appearance is modified as a result of weather, sun position, vegetation state in addition to view-point and lighting, as usual in indoor or urban environments. In robotics, place recognition is used for the loop-closure stage of most large scale SLAM systems where its reliability is critical [3]. It is also an important part of any long-term monitoring system operating outdoor over many seasons [1, 4].

In practice, place recognition is usually cast as an image retrieval task where a query image is matched to the most similar image available in a database. The search is computed on a projection of the image content on a much lower-dimensional space. The challenge is then to compute a compact image encoding such that images of the same location are near to each other despite their change of appearance due to environmental changes.

Figure 1: WASABI computes a global image descriptor for place recognition over bucolic environments across seasons. It builds upon the image semantics and its edge geometry that are robust to strong appearance variations caused by illumination and season changes. While existing methods are tailored for urban-like scenes, our approach generalises to bucolic scenes that offer distinct challenges.

Most of the existing methods start with detecting and describing local features over the image before aggregating them into a low-dimensional vector. The methods differ on the local feature detection, description and aggregation. Most of the research efforts have focused on environments with rich semantics such as cities or touristic landmarks [5, 6]. Early methods relied on hand-crafted feature description (e.g. SIFT [7]) and simple aggregation.
based on histograms constructed on a clustering of the feature space [8]. Recent breakthroughs use deep-learning to learn retrieval-specific detection [6], description [9] and aggregation [5]. Another line of work relies on the geometric distribution of the image semantic elements to characterise it [10]. However, all of these approaches assume that the images have rich semantics or strong textures and focus on urban environments. To the contrary, we are interested in scenes described by images depicting nature or structures with few semantic or textured elements. In the following, such environments, including lakeshores and parks, will be qualified as ‘bucolic’.

In this paper, we show that an image descriptor based on the geometry of semantic edges is discriminative enough to reach State-of-the-Art (SoA) image-retrieval performance on bucolic environments. The detection step consists in extracting semantic edges and sorting them by their label. Continuous edges are then described with the wavelet transform [11] over a fixed-sized subsampling of the edge. This constitutes the local description step. The aggregation is a simple concatenation of the edge descriptors and their labels which, together, make the global image descriptor. Figure 1 illustrates the image retrieval pipeline with our novel descriptor dubbed WASABI:\footnote{Wavelet SemAntic edge descriptor for Bucolic environment\textit{}}: A collection of images is recorded along a road during the Spring. Global image descriptors are computed and stored in a database. Later in the year, in Autumn, while we traverse the same road, we describe the image at the current location. Place recognition consists in retrieving the database image which descriptor is the nearest to the current one. To compute the image distance, we associate each edge from one image to the nearest edge with the same semantic label in the other image. The distance between two edges is the Euclidean distance between descriptors. The image distance is the sum of the distances between edge descriptors of associated edges.

WASABI is compared to existing image retrieval methods on two outdoor bucolic datasets: the park slices of the CMU-Seasons\cite{2} and Symphony\cite{1}, recorded over a period of 1 year and 3 years respectively. Experiments show that it outperforms existing methods even when the latter are finetuned for these datasets. It is also on par with NetVLAD, the current SoA on urban scenes, which is specifically optimised for city environments. This shows that WASABI can generalise across environments.

The contribution of this paper is a novel global image descriptor based on semantics edge geometry for image-based place recognition in bucolic environment. Experiments show that it is also suitable for urban settings. The descriptor as well as the evaluation code are available at \url{https://github.com/abenbihi/wasabi.git}.

2 RELATED WORK

This section reviews the current state-of-the-art on place recognition. A common approach to place recognition is to perform global image descriptor matching. The main challenge is defining a compact yet discriminative representation of the image that also has to be robust to illumination, viewpoint and appearance variations.

Early methods built such a global descriptor by aggregating locally invariant features such as SIFT \cite{7}. A first step consists in generating a codebook of visual words by clustering local feature descriptors over a training dataset. This dataset must be different from the place recognition one to generalise well. An image is then described with the statistics of its local features with respect to this codebook. In \cite{8}, the local features of the image are assigned to the codebook clusters and the descriptor is simply the clustering histogram. \cite{12} improves over the previous clustering by fitting a mixture of Gaussians over the training dataset local features. Then, for each local feature of the image of interest, they concatenate the gradient of the probability of this feature to belong to one of the gaussian. This high-dimensional vector is then reduced with Principal Component Analysis (PCA). This approach is simplified in \cite{13} by computing clusters as in \cite{8} even though it does not use cluster-histogram for aggregation. Instead, they concatenate the distance vector between each local feature and its nearest cluster which is a specific case of the derivation in \cite{12}. All these methods rely on features based on pixel distribution that assumes that images have strong textures, which is not the case for bucolic image. They are also sensitive to variations in the image appearance such as seasonal changes. In contrast, we rely on the geometry of semantic elements and that proves to be robust to strong appearance changes.

Recent works also aim at disentangling local features and pixel intensity through learned feature descriptions. \cite{9} uses pre-trained Convolutional Neural Network (CNN) feature maps as local descriptor and aggregates them using the previous methods. The current SoA in place recognition, NetVLAD\cite{5} also takes advantage of the rich representation space of CNN but specifically trains a CNN to generate local feature descriptors relevant for image retrieval. They transformed the hand-crafted aggregation from VLAD\cite{14} into an end-to-end learning pipeline and reached top performances on urban scenes such as the Pittsburg or the Tokyo time machine datasets \cite{15,16}. DELF\cite{6} tackles the selection of local features and trains a network to sample only features relevant to the image retrieval through an attention mechanism on a landmark dataset. We also rely on CNNs to segment images but not to describe them. Instead, we use this high-level information to hand-craft a novel descriptor that relies on a combination of segmentation and image geome-
try. Segmentation is indeed robust to appearance changes but bucolic environments are typically not rich enough for the segmentation to be sufficient for place recognition. By combining it with information from the scene geometry we succeed in augmenting the discriminative power of the representation.

Another recent approach leverages the image geometry. [10] converts images into a semantic graph, uses temporal information to fuse the graphs over time and generates a global database graph. Then, given a new image expressed as a semantic graph, image retrieval is reduced to a graph matching problem. However, this approach assumes again that the environment is rich in semantic elements to avoid ambiguous graphs. This is not the case in bucolic environments which leads us to leverage edges as another robust and discriminative image element. Edge-based image retrieval is not novel [17] and the literature offers a wide range of edge descriptors [18]. But these local descriptors are usually less robust to illumination and viewpoint variations than their pixel-based counterparts.

In this work, we fuse edge description with semantic information to reach SoA performance on bucolic image retrieval across seasons. We rely on the wavelet descriptor for its compact representation while offering uniqueness and invariance properties [11].

3 METHOD

This section details the three steps of image retrieval: the detection and description of local features, their aggregation and the image distance computation. In this paper, a local feature is constructed as a vector that embeds the geometry of semantic edges.

3.1 Local feature detection and extraction.

The local feature detection stage takes a color image as input, and outputs a list of continuous edges together with their semantic labels. Two equivalent approaches can be considered. The first is to extract edges from the semantic segmentation of the image, i.e. its pixel-wise classification. The SoA relies on CNN trained on labeled data [19, 20]. The second approach is also based on CNN but directly outputs the edges together with their labels [21, 22]. The first approach is favored as there are many more public segmentation models than semantic edges ones.

Hence, starting from the semantic segmentation, a post-processing stage is necessary to reduce the labelling noise. Most of this noise consists in labeling errors around edges or small holes inside bigger semantic units. To reduce the influence of these errors, semantic blobs smaller than \( \text{min\_blob\_size} \) are merged with their nearest neighbours.

Furthermore, to make semantic edges robust over long periods of time, it is necessary to ignore classes corresponding to dynamic objects such as cars or pedestrians. Otherwise, they would alter the semantic edges and modify the global image descriptor. These classes are removed from the segmentation maps and the resulting hole is filled with the nearest semantic labels.

Taking the cleaned-up semantic segmentation as input, a simple Canny-based edge detection is performed and edges smaller than \( \text{min\_edge\_size} \) pixels are filtered out.

Segmentation noise may also break continuous edges. So the remaining edges are processed so as to re-connect edges actually belonging with each other. For each class, if two edge extremities are below a pixel distance \( \text{min\_neighbour\_gap} \), the corresponding edges are grouped together into a unique edge.

The parameters are chosen empirically based on the segmentation noise of the images. We use the segmentation model from [19]. It features a PSP-Net [20] network trained on Cityscapes [23]. As a result, we fine-tuned it on CMU-Seasons. In this case, the relevant detection parameters were \( \text{min\_blob\_size}=50, \text{min\_edge\_size}=50 \) and \( \text{min\_neighbour\_gap}=5 \).

3.2 Local feature description

Among the many existing edge descriptor, we favor the wavelet descriptor [11] for its properties relevant to image retrieval. It consists in projecting a signal over a basis of known function and is often used to generate compact yet unique representation of a signal. Wavelet description is not the only transform to generates a unique representation for a signal. The Fourier descriptors [24, 25] also provides such a unique embedding. However, the wavelet description is more compact than the Fourier one due to its multiple-scale decomposition. Empirically, we confirmed that the former was more discriminative than the latter for the same number of coefficients.

![Figure 2: Symphony. Semantic edge association across strong seasonal and weather variations.](image-url)

Given a 2D contour extracted at the previous step, we
subsample the edge at regular steps and collect $N$ pixels. Their $(x, y)$ locations in the image are concatenated into a 2D vector. We compute the discrete Haar-wavelet decomposition over each axis separately and concatenate the output that we L2 normalise. In the experiments, we set $N = 64$ and keep only the even coefficients of the wavelet transforms. This does not destroy information as the coefficients are redundant. The final edge descriptor is a 128-dimension vector.

3.3 Aggregation and Image distance

Aggregation is a simple accumulation of the edge descriptors together with their label. Given two images and using the aggregated edge descriptors, the image distance is the average distance between matching edges. More precisely, edges belonging to the same semantic class are associated between the images solving an assignment problem (see Fig. 2). The distance used is the Euclidean distance between edge descriptors and the image distance is the average of the associated descriptor distances. In a retrieval setting, we compute such a distance between the query image and every image in the database and return the database entry with the lowest distance.

4 EXPERIMENTS

This section details the experimental setup and presents results for our approach against methods for which public code is available: BoW[8], VLAD[26], NetVLAD[5], DELF[6]. We demonstrate the retrieval performance on two outdoor bucolic datasets: CMU-Seasons[2] and Symphony[1], recorded over a period of 1 year and 3 years respectively. Although existing methods reach SoA performance on urban environment, our approach proves to outperform them on bucolic scenes, and so, even when they are finetuned. It also shows better generalisation as it achieves near SoA performance of the urban slices on the CMU-Seasons dataset.

4.1 Datasets

Extended CMU-Seasons The Extended CMU-Seasons dataset (Fig. 3) is an extended version of the CMU-Seasons [27] dataset. It depicts urban, suburban, and park scenes in the area of Pittsburgh, USA. Two front-facing cameras are mounted on a car pointing to the left/right of the vehicle at approximately 45 degrees. Eleven traversals are recorded over a period of 1 year and the images from the two cameras do not overlap. The traversals are divided into 24 spatially disjoint slices, with slices [2-8] for urban scenes, [9-17] for suburban and [18-25] for park scenes respectively. All retrieval methods are evaluated on the park scenes for which ground-truth poses are available [22-25]. The other park scenes [18-21] can be used to train learned approaches. For each slice in [22-25], one traversal is used as the image database and the 10 other traversals are the queries. In total, there are 78 image sets of roughly 200 images with ground-truth camera poses. Figure 3 shows examples of matching images over multiple seasons with significant variations.

Lake The Symphony [1] dataset consists of 121 visual traversals of the shore of Symphony Lake in Metz, France. The 1.3 km shore is surveyed using a pan-tilt-zoom (PTZ) camera and a 2D LiDAR mounted on an unmanned surface vehicle. The camera faces starboard as the boat moves along the shore while maintaining a constant distance. The boat is deployed on average every 10 days from Jan 6, 2014 to April 3, 2017. In comparison to the roadway datasets, it holds a wider range of illumination and seasonal variations and much less texture and semantic features, which challenges existing place recognition methods.

We generate 10 traversals over one side of the lake from the ground-truth poses computed with the recorded 2D laser scans [28]. The other side of the lake can be used for training. One of the 121 recorded traversal is used as the reference from which we sample images at regular locations to generate the database. For each database image, the matching images are sampled from 10 random traversals out of the 120 left. Note that contrary to the CMU-Seasons dataset, this means that there is no light and appearance continuity over one traversal (Fig. 4).
Figure 4: Symphony dataset. Top-Down: images and their segmentation. First line: reference traversal at several locations. Each column $k$ depicts one location $\text{Pos}_k$. Each line depicts $\text{Pos}_k$ over random traversals noted $T$. Note that contrary to CMU-Seasons, we generate mixed-conditions evaluation traversals from the actual lake traversals. So there is no constant illumination or seasonal condition over one query traversal $T$.

4.2 Experimental setup

This section describes the rationale behind the evaluation. On CMU-Seasons we evaluate place recognition methods with respect to ($w.r.t$) the semantic elements of a traversal on one hand, and $w.r.t$ the lighting and seasonal conditions on the other hand. On the Symphony dataset, we assess their robustness to low texture images with few semantic elements and even harsher lighting and seasonal variations.

The first CMU-Seasons evaluation, $w.r.t$ the semantic elements, consists in running independent place recognition over each slice and average the performance over the traversals. Since the slices are spatially disjoint, they hold different semantic elements that challenge the image retrieval in various ways. For example, slice 23 seen from camera 0 holds mostly repetitive patterns of trees that are harder to differentiate than the building skyline seen from camera 1 on slice 25. Averaging over the traversals is a way to put aside the influence of the lighting and season for each traversal.

The second evaluation, $w.r.t$ the lighting and season, starts the same way with independent place recognition over each slice. But the scores are averaged over the slices for each traversal. This way, the semantic content is the same for all the traversals and only the lighting and season change.

On Symphony, only the lighting and seasonal robustness are assessed as the semantic content is constant over the lake.

As mentioned previously, our approach is evaluated against BoW, VLAD, NetVLAD and DELF. In their version available online, these methods are mostly tailored for rich semantic environments: the codebook for BoW and VLAD is trained on Flickr60k [29], NetVLAD is trained on the Pittsburg dataset [15] and DELF on the Google landmark one [6]. For fair comparison, we finetune them on CMU-Seasons and Symphony when possible, and report both original scores and the finetuned ones noted with (*). A new codebook generated for BOW an VLAD, using the CMU park slices 18-21. The NetVLAD training requires images with ground-truth poses, which is not the cases for these slices. So we train it on three slices from 22-25 and evaluate it on the remaining one. On Symphony, images together with their ground-truth poses are sampled from the west side of the lake that is spatially disjoint from the evaluation traversals. The DELF learned local features are not finetuned as the training code is not available even though the model is.

Finally, our approach is tested against the original available methods on the three urban CMU-Seasons slices for which ground-truth poses are available [6-8]. This assesses whether our approach is also relevant for urban settings and hence better generalise across environments than methods tailored specifically for urban scenes.

4.3 Metrics

The place recognition metrics are the recall@$N$ and the mean Average Precision ($mAP$)[30]. Both depend on a distance threshold $\epsilon$: a retrieved database image matches the query if the distance between their camera center is below $\epsilon$. The recall@$N$ is the percentage of queries for which there is at least one matching database image in the first $N$ retrieved images. We set $N \in \{1, 5, 10, 20\}$, and $\epsilon$ to $5m$ and $2m$ for the CMU-Seasons and the Symphony datasets respectively. Both metrics are available in the code.

4.4 Results

WASABI shows better performance on bucolic scenes than existing methods while only slightly underperforming the SoA NetVLAD and DELF on urban environments. This is expected as the SoA is optimised for such settings. Still, this shows that our method generalises to both types of environments. Finetuning existing methods to the bucolic scenes proves to be useful for VLAD but does not improve the overall performance for BoW and NetVLAD. A plausible explanation is that these methods require more data than the one available. Investigating the finetuning of these methods is out of the scope of this paper. The rest of this section details the results.
Recall $\hat{N}$ for the 8 park slices with different semantic appearance. Our approach tends to outperform SoA methods unless they are not enough edges $(23,c1)$ or the slice holds many urban elements $(25,c1)$.

Performance w.r.t semantic elements on CMU-Seasons
Fig 5 plots the recognition recall@N averaged over several seasonal conditions for 8 locations. All methods show the same sensitivity to the retrieval error tolerance: the recall@N curves increase similarly as a function of $N$. On half the locations $(22,c0, 22,c1, 24,c0, 25,c0)$, WASABI doubles the SoA NetVLAD score. Elsewhere, it reaches similar scores except for $25,c1$ for which it is slightly outperformed. This is expected as this slice holds urban elements for which DELF and NetVLAD are specifically trained.

All recalls drop on slice $23,c1$ that holds dense trees along the road. The images not only have few features usually leveraged by other methods, but also few semantic edges on which WASABI relies. This limit suggests that we exploit multiple levels of edges in future work and not only semantic edges. While finetuning NetVLAD shows no advantage over the original one on other slices, here it reaches the retrieval score. This suggests that retrieval could be learned on challenging bucolic environments. Still, the performance on the remaining slices shows that a simple finetuning may not be enough to transfer this method and additional research is necessary. These observations are confirmed with the mAP results, not plotted here for the sake of page limits.

Performance w.r.t light/season on CMU-Seasons
Fig 6 assesses the influence of various light/season conditions averaged over various semantic scenes. Only 9 plots are displayed for the sake of visualisation. This can be explained by the presence of a strong sunglare (Fig. 8) that makes the segmentation noisy. The resulting semantic edges are then less reliable even though they lead to similar retrieval performance as SoA.

Global performance on Symphony
Fig 7-left plots the Recall@5 with respect to the query image located at the same spot over all traversals. There is no correlation between the performance and the query location. As expected, the score shows no bias on a location, which backs our claim that a semantic analysis on Symphony is pointless. Recall@N for other values of N and other methods support this claim.

The right plot shows the Recall@N averaged over all the query images. WASABI presents a slight advantage over NetVLAD and DELF although one could have expected higher performance based on the previous solid re-
results on CMU-Seasons. One explanation is the segmentation noise induced by the image noise in one hand (e.g., sunglare) and the lack of domain adaptation on the other hand (Fig. 8). As there is no ground-truth segmentation for the Symphony dataset, finetuning the segmentation is currently not possible. However, the satisfying results on CMU-Seasons motivate future work to improve the Symphony segmentation as well as the robustness of the descriptor to failures of the segmentation stage.

Generalisation to urban setting

Figure 9 shows that WASABI compares with SoA NetVLAD scores even though it is not specifically tailored for urban environments. It is interesting to note that on slice 7, WASABI slightly outperforms NetVLAD on camera 0 whereas we observe the opposite on camera 1. The former mostly shows grass and trees along a parking lot whereas the latter images a building. This observation supports the bias that existing methods have toward urban environments.

Conclusions

In this paper, we presented WASABI, a novel image descriptor for place recognition across seasons in bucolic environments. It represents the image content through the geometry of its semantic edges, taking advantage of their invariance w.r.t seasons, weather and illumination. Experiments show that it outperforms existing image-retrieval approaches better suited for urban environments. Tuning these methods for bucolic datasets proves to be insufficient to reach the same performances as our approach. Conversely, WASABI generalises well to other settings and reach scores on par with SoA on urban scenes. Current research now focuses on improving the segmentation and disentangling the image description from the noise segmentation.
References

[1] S. Griffith, G. Chahine, and C. Pradalier, “Symphony lake dataset,” The International Journal of Robotics Research, vol. 36, no. 11, pp. 1151–1158, 2017.

[2] T. Sattler, W. Maddern, C. Toft, A. Torii, L. Hammarstrand, E. Stenborg, D. Safari, M. Okutomi, M. Pollefeys, J. Sivic et al., “Benchmarking 6dof outdoor visual localization in changing conditions,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 8601–8610.

[3] M. Cummins and P. Newman, “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.

[4] W. Churchill and P. Newman, “Experience-based navigation for long-term localisation,” The International Journal of Robotics Research, vol. 32, no. 14, pp. 1645–1661, 2013.

[5] R. Arandjelovic, P. Gronat, A. Torii, T. Pajdla, and J. Sivic, “Netvlad: Cnn architecture for weakly supervised place recognition,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 5297–5307.

[6] H. Noh, A. Araujo, J. Sim, T. Weyand, and B. Han, “Large-scale image retrieval with attentive deep local features,” in Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 3456–3465.

[7] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” International journal of computer vision, vol. 60, no. 2, pp. 91–110, 2004.

[8] J. Sivic and A. Zisserman, “Video google: A text retrieval approach to object matching in videos,” in null. IEEE, 2003, p. 1470.

[9] A. Babenko, A. Slesarev, A. Chigorin, and V. Lempitsky, “Neural codes for image retrieval,” in European conference on computer vision. Springer, 2014, pp. 584–599.

[10] A. Gawel, C. Del Don, R. Siegwart, J. Nieto, and C. Cadena, “X-view: Graph-based semantic multi-view localization,” IEEE Robotics and Automation Letters, vol. 3, no. 3, pp. 1687–1694, 2018.

[11] G.-H. Chuang and C.-C. Kuo, “Wavelet descriptor of planar curves: Theory and applications,” IEEE Transactions on Image Processing, vol. 5, no. 1, pp. 56–70, 1996.

[12] F. Perronnin, Y. Liu, J. Sánchez, and H. Poirier, “Large-scale image retrieval with compressed fisher vectors,” in 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, 2010, pp. 3384–3391.

[13] H. Jégou, M. Douze, C. Schmid, and P. Pérez, “Aggregating local descriptors into a compact image representation,” in IEEE Conference on Computer Vision & Pattern Recognition, jun 2010. [Online]. Available: http://lear.inrialpes.fr/pubs/2010/JDSP10

[14] H. Jégou, M. Douze, C. Schmid, and P. Pérez, “Aggregating local descriptors into a compact image representation,” in CVPR 2010-23rd IEEE Conference on Computer Vision & Pattern Recognition. IEEE Computer Society, 2010, pp. 3304–3311.

[15] A. Torii, J. Sivic, T. Pajdla, and M. Okutomi, “Visual place recognition with repetitive structures,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2013, pp. 883–890.

[16] A. Torii, R. Arandjelovic, J. Sivic, M. Okutomi, and T. Pajdla, “24/7 place recognition by view synthesis,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1808–1817.

[17] X. S. Zhou and T. S. Huang, “Edge-based structural features for content-based image retrieval,” Pattern recognition letters, vol. 22, no. 5, pp. 457–468, 2001.

[18] M. a. Merhy et al., “Reconnaissance de formes basée géodésiques et déformations locales de formes,” Ph.D. dissertation, Brest, 2017.

[19] M. Larsson, E. Stenborg, L. Hammarstrand, M. Pollefeys, T. Sattler, and F. Kahl, “A cross-season correspondence dataset for robust semantic segmentation,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 9532–9542.

[20] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, “Pyramid scene parsing network,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 2881–2890.
[21] Z. Yu, C. Feng, M.-Y. Liu, and S. Ramalingam, “Casenet: Deep category-aware semantic edge detection,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 5964–5973.

[22] D. Acuna, A. Kar, and S. Fidler, “Devil is in the edges: Learning semantic boundaries from noisy annotations,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 11075–11083.

[23] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, “The cityscapes dataset for semantic urban scene understanding,” in Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[24] C. T. Zahn and R. Z. Roskies, “Fourier descriptors for plane closed curves,” IEEE Transactions on computers, vol. 100, no. 3, pp. 269–281, 1972.

[25] G. H. Granlund, “Fourier preprocessing for hand print character recognition,” IEEE transactions on computers, vol. 100, no. 2, pp. 195–201, 1972.

[26] H. Jegou, F. Perronnin, M. Douze, J. Sánchez, P. Perez, and C. Schmid, “Aggregating local image descriptors into compact codes,” IEEE transactions on pattern analysis and machine intelligence, vol. 34, no. 9, pp. 1704–1716, 2011.

[27] H. Badino, D. Huber, and T. Kanade, “The CMU Visual Localization Data Set,” http://3dvis.ri.cmu.edu/data-sets/localization, 2011.

[28] C. Pradalier and F. Pomerleau, “Multi-session lakeshore monitoring in visually challenging conditions,” in Field and Service Robotics. Springer, 2019.

[29] H. Jegou, M. Douze, and C. Schmid, “Hamming embedding and weak geometric consistency for large scale image search,” in European conference on computer vision. Springer, 2008, pp. 304–317.

[30] J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman, “Object retrieval with large vocabularies and fast spatial matching,” in 2007 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2007, pp. 1–8.