A Brief Survey of Loop Closure Detection: A Case for Rethinking Evaluation of Intelligent Systems

Samer B. Nashed
College of Information and Computer Sciences
University of Massachusetts Amherst
Amherst, MA 01003
snashed@cs.umass.edu

Abstract

The process of determining if an agent has returned to a previously visited location by analyzing data from its sensors, known variably as loop closure detection and place recognition, is an essential component of modern mobile robotic systems. It is also one of the most challenging instances of the canonical 'data association problem', the problem of determining if and how two different sensor readings may or may not originate from the same physical entity. Until recently, hand-crafted solutions have dominated research on loop closure detection. As approaches based on deep learning begin to outperform more classical techniques, we have an excellent opportunity to rethink the definition of success, what different failure cases can teach us about the frontiers of robotic intelligence, and how to generate more interesting hypotheses in a more targeted manner. Here, we present an overview of loop closure detection research, focusing on its role within the robotics stack, why it is still an unsolved problem, and some of the successes and failures of research to-date. We go on to propose a new experimental paradigm and argue that loop closure detection offers a unique vehicle for studying machine learning in embodied systems, and embodied intelligent systems more broadly.

1 Introduction

Many of us have, at one point or another, wandered through a city we were not completely familiar with only to find ourselves not entirely sure of our location. Our belief about which corner we were on was spread out over several intersections. Then, after continuing to wander one or two more blocks, we would pass a familiar shop or catch a glimpse of a large building we had seen previously and our belief would collapse onto a much smaller area. Additionally, not only did we become more certain of our current location, but we also had a more accurate estimate of the path we had taken between that moment and the last time we had seen a familiar landmark.

Robots rely on a similar process, known as loop closure, to reduce the effect of error in their location estimate which builds up over time due to the inherent noise and un-modeled dynamics in their sensors and actuators. Loop closure events, where a set of features assumed to be stationary are re-observed, provide information about a robot's location in the form of a constraint on the estimated trajectory. Essentially, these events require the cumulative effect of all motor commands issued between observations to result in a specific affine transformation of the robot's coordinate system. Effective loop closure detection (LCD) is critical for the realization of mobile robotic systems capable of long-term autonomy. This fact is not lost on the research community, and there has been a wealth of research on LCD and other related problems. However, robots do not enjoy the same deep contextual, semantic, or commonsense knowledge that humans employ when attempting to loop-close. Although

NeurIPS 2020 Workshop: ML Retrospectives, Surveys & Meta-Analyses (ML-RSA).

this presents an obvious hurdle for researchers trying to implement functional systems, we argue this offers a unique academic opportunity and will revisit this trait in a more optimistic context.

Loop closure detection shares many similarities with other topics in machine learning and when operating on robotic systems using cameras can be thought of as a special case of image retrieval. Since nearly all LCD systems compute a similarity score as a proxy for the probability that two sensor readings are from the same location or bear some geometric relationship, LCD is fundamentally linked to all other topics in computing where a notion of similarity is ambiguous. A wide array of models and algorithms are theoretically applicable, and this is evidenced by the diversity of the resultant body of work on LCD. The heterogeneity of sensors, compute limitations, and operating environments of different robots generate further specialization and customization.

Custom systems may not seem generally interesting, but they are often required, since loop closure systems must be more performant than their image retrieval counterparts. Mobile robots depend on accurate, online location estimates, and LCD systems are a key component in generating these estimates. In addition to heightened performance requirements, LCD systems face a uniquely large set of confounding factors that make their design, deployment, and evaluation relatively challenging compared to many modern machine learning and perceptual robotics problems. For example, aerial robots inspecting the structural integrity of a building or bridge [14] might need to deal with greater variance in viewing angles compared to ground robots given the additional degrees of freedom. Underwater robots [113] building oceanographic models or inspecting shipping infrastructure may encounter lower levels of visibility, reducing the effective range and reliability of some sensors. Ground robots may operate indoors, as service robots in security or elder-care roles [62], where they encounter frequent occlusions, or outdoors, as automated farm equipment [37] or autonomous passenger cars [21], where they are exposed to a wide variety of lighting conditions.

In the first half of this paper, we provide an overview of the challenges in detecting loop closures and the resultant research aimed at meeting those challenges. In the latter half, we discuss evaluation of loop closure detection systems and intelligent systems more broadly. We start by addressing common shortcomings in contemporary LCD experimentation and highlighting some potential near-term changes in best practice. We then propose a new, human-centered experimental paradigm for LCD systems which may lead to richer hypotheses and more targeted approaches to developing state-of-the-art systems. LCD is an excellent example of a highly integrated behavior, and the fact that both humans and robots run essentially the same process could open new avenues for validating or interrogating LCD systems and providing direction and insight to researchers studying a broad spectrum of related artificial intelligence questions. Lastly, we argue that the recent ascension of machine learning systems for loop closure detection provides a valuable opportunity for rethinking the broad impact and implications of this domain. We also discuss how machine learning can advance the state-of-the-art, what the limits of current machine learning-based systems can tell us about the fundamental requirements of loop closure detection, and how our performance on loop closure may entail success or failure in complex tasks more broadly.

2 Loop Closure Detection: Context and Challenges

In practice, loop closure detection is not a standalone problem and, as with most other problems in robotics, the process of loop closure has spawned many related research topics. These include how to: use loop closures to increase map accuracy most effectively, adapt single-robot loop closure detection schemes to multi-robot settings [146, 128], efficiently communicate information about loop closures in multi-robot settings [116, 147, 53], plan actions for active loop closure [29, 27], prove trajectories contain loops by analyzing proprioceptive data [125], handle false loop closures robustly [115], and search efficiently for potential loop closures among past data instances in order to meet robots' real-time operating constraints [156, 89, 80]. We restrict our focus to LCD because it is a more commonly required capability than some of the above and because it lends itself more readily to the application of machine learning techniques. However, we should note that many of these issues, especially the first and last, are sometimes implicitly addressed in research nominally focused on improving LCD accuracy and robustness. Before we give an overview of existing research, we will provide some background and context regarding the role of LCD systems in mobile robotic systems and what makes LCD an especially challenging and interesting problem.

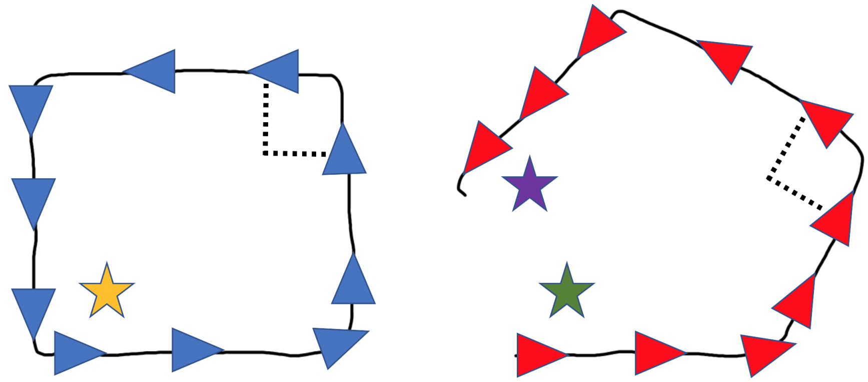

Figure 1: Example loop closure diagram. The robot's estimated trajectory is represented by the black line, and a subset of robot poses are represented by triangles. The stars represent recognizable features that may produce loop closures. On the left, the loop is detected and a constraint is added to the optimization problem representing the fact that the feature (star) must be in the same place in the global frame during its original and subsequent observations. On the right, the features are not recognized as the same, and the accumulated error in location estimate persists.

2.1 The Role of Loop Closure Detection

Loop closure detection is a sub-problem within one of the quintessential problems in robotics: Simultaneous Localization And Mapping (SLAM). SLAM systems use streams of sensor data from robots as they move around in new environments to simultaneously create a model of the environment and determine the robot's location with respect to that model. As the robot explores, the model (map) is built and refined incrementally, based on new data and the robot's previous estimate of its location. In many scenarios the robot does not have access to sensors with globally bounded error, such as GPS. Small amounts of noise in sensor readings can shift the minima or maxima of the robot's cost functions or estimators away from the ground truth, resulting in small errors in the robot's location estimate. Over time these errors can compound and if left unchecked will eventually produce location estimates which are so far from the ground truth that they prevent normal operation of the robot.

Figure 1 illustrates a simplified loop closure event. The error in the trajectory on the right is the result of the compounding of many small errors in the estimate from one pose (triangle) to the next. Since the estimate of the next pose depends on the estimate of the current pose, errors in previous estimates affect future estimates. Note that some segments of the trajectory may be internally accurate, such as the top right section which correctly estimates a 90 degree left turn, but due to previous errors, they result in an *increasing* error relative to the ground truth. Robust re-detection of stationary landmarks in the world can produce location information with a constant amount of error, regardless of how far the robot has travelled. This is in contrast to pure odometry, for example, the trajectory on the right, where the location estimate becomes increasingly inaccurate over time.

LCD systems are, so far, the most effective way to deal with localization drift. Assuming no false positives (which is not always a valid assumption), LCD systems do provide location estimates with globally bounded error. Furthermore, loops can be detected with sensors already available to the robot, and the natural trajectory of a robot exploring a new area will likely present opportunities for loop closure, even without explicitly planning for it. Loop closure events also provide indirect information about intermediate robot locations which occur between the original and subsequent, or loop closing, observation. Since a robot's trajectory unfolds sequentially, information from a loop closure event is applied selectively to the segment of the trajectory that occurs between the original and subsequent observation. When features are re-observed multiple times, the loop closure events can create multiple constraints, one for each pair of observations. This ability to use re-observations to refine the trajectory estimate for segments of trajectory which are not local to the observation is valuable for maintaining map accuracy.

Ultimately, expecting a similarity score computed on a single sensor reading, or even a sequence of sensor readings, to accurately and robustly represent the probability of revisiting the same location may be optimistic. However, although such scores are difficult to compute, similar efforts where similarity scores are used as proxies for much more complicated computations are extremely common across machine learning, including protein fold recognition [86], forensic ballistics [144], bilingual dictionary induction [129], knowledge graph evaluation [174], face recognition [141, 154, 155], personalized disease prediction [137], patent law [151], social network analysis [13], facility layout [78], ad relevance [163], image patch correspondence [61], question answering [25], UML diagram consistency [1], opponent modeling [112], paraphrase detection [45], behaviour novelty in swarm robotics [55], robotic grasp planning [143], and personality traits for human-robot interaction [4], among many others. LCD systems have the potential to benefit from the wealth of research on constructing robust similarity measures.

There are two key differences between LCD systems and other tasks performing similarity queries, such as image retrieval. First, the cost of false positives is high. State-of-the-art SLAM systems represent the estimate of the robot's trajectory as a dynamic Bayesian network (DBN) and from this representation construct a non-linear least-squares optimization problem where minimum cost represents maximum probability. Loop closures essentially add extra edges in the DBN, which manifest as additional cost terms in the resulting optimization problem. False loop closures add erroneous cost functions which can drastically alter the maximum likelihood trajectory estimate and thus severely distort the resulting map. In most loop closure situations, false positives are more costly than false negatives. Second, LCD systems must distinguish between different physical locations, even if those locations appear very similar qualitatively. The problem of distinct locations appearing nearly identical is known as 'perceptual aliasing', and is surprisingly common in man-made environments as well as some natural terrain like fields, forests, and sea floors.

Perceptual aliasing can have different degrees of impact depending on how the robot represents its operating environment. The two most common representations are topological maps and metric maps. Topological maps represent the environment as a graph, where distinct 'places' are nodes, and the edges represent navigability between places. A robot's trajectory through a topological map is represented by a sequence of nodes representing the order and identity of the places visited by the robot. Metric maps build a dense reconstruction of the environment using landmarks, features, or raw sensor readings with 2D or 3D positions all registered to the same global coordinate frame. A robot's trajectory through a metric map is represented by a sequence of robot positions and orientations, called 'poses', in $SE(2)$ or $SE(3)$ which describe the robot's motion through Cartesian space over time. Perceptual aliasing is often a larger problem for metric SLAM since even false loop closures within the same room can degrade the location estimate, while in topological SLAM the same association would likely result in a true positive. Moreover, loop closures in metric SLAM need to estimate the affine transform by which two sensor reading are related in addition to simply identifying that the readings originate from the same general place.

2.2 Challenges in Loop Closure Detection

Loop closure detection is a difficult problem in the general case. To begin, LCD systems rely on sensors which are inherently noisy, discrete, and often lack perfect calibration. An even greater number of challenges stem from the diversity of potential sensing conditions. For example, the images of the Eiffel Tower in Figure 2 represent a surprisingly wide variety of sensing conditions driven by several important underlying concepts and their dynamics. These include partial occlusions, field of view limitations, changing viewing angles, the presence or absence of vegetation due to seasonal changes, artificial light, changes in lighting conditions due to time of day and weather, reflective surfaces, and the presence of transient objects, such as boats on the Seine river. Some of these dynamics are difficult to model or predict, and the robot may face many of them simultaneously depending on its operating environment. Part of the difficulty these dynamics present is that research on measuring or characterizing these effects is still in its early stages.

Given this imposing slate of challenges, it seems likely that approaches more deeply integrated with knowledge or models of many of these processes will be superior in the long run. This hypothesis is motivated by two points. First, the data-ineffiency of modern machine learning systems is a serious problem in robotics. Robots are expensive to operate and deploy, and may have serious safety concerns which require increased robustness and the corresponding increase in data. Therefore, getting enough realistic data that represents all contingencies is arguably infeasible, especially for domains like extraterrestrial or deep sea exploration. Second, successful robotic systems are usually developed in many development cycles, as opposed to a single, inspired choice of architecture.

Figure 2: A variety of images of the Eiffel Tower and Seine river.

Although evaluating a single, large algorithm or network is easier from a testing infrastructure standpoint, it has significant disadvantages with respect to interpretability. This is a drawback for integrated systems like robots, since better models of performance have a positive feedback loop with respect to development by allowing more targeted algorithm improvements.

3 An Overview of Existing Research

Here, we present a brief summary of existing research on loop closure detection. Previously, surveys on LCD for depth-based [150] and camera-based [28, 153, 94] systems have offered useful summaries of existing work, explanations of various methods, and areas for potential research and innovation. In this section, our main contributions are 1) a combined accounting of both depth-based and camera-based techniques for loop closure, including research published since the last widely acknowledged surveys, and 2) providing an easily accessible directory linking LCD papers to particular sensing modalities, map types, technical details, and data sets. Regarding data set notation in the tables: entries with 'AV Driving' or 'Service Robot' indicate collection of a custom data set that most similarly resembles data gathered by autonomous vehicles (AVs) or service robots, respectively, and entries with 'dataset/+' indicate additional open source data was also used.

In general, we identify three primary axes along which engineering decisions are made when building LCD systems. First, should sensor input be represented by descriptors computed on subsets of data, such as image patches (often called local descriptors), or should descriptors represent data holistically (often called global descriptors)? Second, should the functions that produce descriptors from raw data be hand-crafted or learned? Third, should similarity functions for ultimately detecting loop closures be based on statistical measures, hand-crafted distance functions, or should they be learned, for instance by a neural network? All of these questions represent areas of active research, and a comprehensive understanding of these decisions and their affect on system performance remains elusive. Moreover, some recent research proposes combining elements of the above approaches to create hybrid systems with impressive performance.

3.1 Camera-based Systems

Tables 1 and 2 provide an overview of camera-based LCD systems. One of the more contested decisions, which we omit from the tables, is whether to use global image descriptors or local image descriptors [91]. In general the best choice may depend on the specific application. For instance, large changes in scene due to occlusions could result in large changes in global descriptors, whereas a subset of local descriptors which are not occluded may be robust enough to correctly identify the perceptual target. On the other hand, global effects to the image such as lighting changes due to the day/night cycle may be handled more robustly by global descriptors since the salience of individual image patches under these changes may be difficult to model. However, for determining loop closures and their metric constraints, local image descriptors are superior since they can encode information about viewing angle that global descriptors often cannot.

Work	Modality	Map Type	Learning	Technique	Dataset
[148]	RGB	Metric	No	HoG+RANSAC	Simulation
[69]	RGB	Metric	No	SIFT+RANSAC	AV Driving
[58]	RGB	Metric	No	Clustering+Histograms	fr2
[82]	RGB	Metric	No	SURF+RANSAC	Campus01/+
[161]	RGB	Metric	No	ORB+BoW+RANSAC	Service Robot
[169]	RGB	Metric	No	Vanish points+RANSAC	Service Robot
[35]	RGB	Metric	No	SIFT+RANSAC+MEWC	AUV
[47]	RGB	Topo-Metric	No	iBoW+RANSAC	KITTI/+
[110]	RGB	Topo-Metric	No	SURF+iBoW	Service Robot
[111]	RGB	Topo-Metric	No	BRISK+BoW	Service Robot
[173]	RGB	Topo-Metric	No	BoW+K-means	Service Robot
[109]	RGB	Topo-Metric	No	SIFT+BoW	AV Driving
[152]	RGB	Topo-Metric	Yes	Random Fern+RANSAC	Service Robot
[159]	RGB	Topo-Metric	Yes	ORB+RANSAC+PCANet	New College/+
[11]	RGB	Topological	Yes	LoCATe+DBoW2	Lip6/+
[139]	RGB	Topological	Yes	SIFT+BoW	AV Driving
[117]	RGB	Topological	Yes	Histograms+SVM	INDECS
[97]	RGB	Topological	Yes	BRISK+kNN	St. Lucia/+
[81]	RGB	Topological	Yes	GAN	Norland
[31]	RGB	Topological	Yes	Attentional CNN	St. Lucia/+
[104]	RGB	Topological	No	Patch Matching	Alderley
[30]	RGB	Topological	Yes	Multi-scale CNN	SPED/+
[114]	RGB	Topological	No	SMART	Alderley/+
[145]	RGB	Topological	No	AfRob	Service Robot
[107]	RGB	Topological	No	HOG+min-flow	AV Driving
[39]	RGB	Topological	No	Zernike Moments+NN	KITTI/+
[10]	RGB	Topological	No	BRIEF+BoW	New College/+
[102]	RGB	Topological	Yes	Autoencoder+BoW	KITTI/+
[172]	RGB	Topological	Yes	CNN+PCA	New College/+
[140]	RGB	Topological	No	Clustering+BoW+NN	KITTI/+
[170]	RGB	Topological	No	FAST+BoW	New College/+
[130]	RGB	Topological	No	SIFT+E^2LSH	Service Robot
[85]	RGB	Topological	No	ORB+V-kNN+BoW	New College/+
[103]	RGB	Topological	Yes	Autoencoder+HOG	KITTI/+
[33]	RGB	Topological	No	SURF+BoW+Chow Liu	AV Driving
[24]	RGB	Topological	Yes	CNN+Subgraph Match	New College/+
[72]	RGB	Topological	No	SURF+BoWP+RANSAC	New College/+
[98]	RGB	Topological	No	BRISK+BoW orthophoto	Service Robot
[171]	RGB	Topological	No	SIFT+BoRF	Service Robot
[7]	RGB	Topological	No	SIFT+iBoW+Bayes	AV Driving
[165]	RGB	Topological	No	BRIEF+LSD+LBD+BoW	New College/+
[90]	RGB	Topological	No	Gabor-Gist+PCA	Oxford City
[157]	RGB	Topological	No	BRISK+E^2LSH+NN	Service Robot
[67]	RGB	Topological	Yes	Places CNN	New College/+
[20]	RGB	Topological	No	SURF+BoW	Service Robot
[46]	RGB	Topological	Yes	Autoencoder	Frieburg
[73]	RGB	Topological	No	BRISK+iBoW	City Centre/+
[49]	RGB	Topo-Semantic	Yes	Places365+SeqSLAM	QUT Campus/+

Table 1: LCD systems for RGB images

Hand-crafted local feature descriptors, such as SIFT [92], SURF [12], BRIEF [23], BRISK [83], ORB [126], FREAK [2], and AKAZE [3], as well as methods for computing local image statistics,

Work	Modality	Map Type	Learning	Technique	Dataset
[38]	Gray	Topological	No	PCA+l_2-norm	Service Robot
[52]	Pan	Topological	No	Histograms	Service Robot
[9]	Pan	Topological	No	LDB+Ad-hoc Score	Ford Campus/+
[8]	Pan	Topological	Yes	NetVLAD+CNN	Time Machine
[42]	Omni	Topo-Metric	No	HOG	Service Robot
[5]	Omni	Topo-Metric	No	FS+HOG+l_2-norm	Service Robot
[101]	Omni	Topo-Metric	No	SIFT+kNN	Service Robot
[6]	Omni	Topological	Yes	SIFT+MRF	AV Driving
[142]	Omni	Topological	Yes	Histograms+kNN	Service Robot
[26]	Omni	Topological	No	SIFT+BoW+Bayes	AV Driving
[95]	Omni	Topological	No	Haar Wavelet	Service Robot
[71]	Omni	Topological	No	PIRF	City Centre/+
[149]	Omni	Topo-Semantic	Yes	HCT+SVM	COLD

Table 2: LCD systems for gray-scale, panoramic, or omni-directional images

such as histograms of oriented gradients (HOG) [34], discrete wavelet transforms, including Haar wavelets, discrete Fourier transforms, and Zernike moments [74], among others, are gradually being replaced by local image descriptors learned using convolutional neural networks (CNNs) or generative adversarial networks (GANs) [131]. There have also been some meta-studies that evaluate multiple deep learning approaches for LCD [160] and find that they perform well. We expect many of these methods to better fit given data sets and perhaps to generate more informative mappings from raw images to descriptors in some scenarios, but it seems unlikely that this evolution in feature descriptors alone will produce fully functional LCD systems in general. More impactful questions likely involve better understanding the un-modeled dynamics mentioned in section 2.2, such as investigating whether modeling seasonal change removal is more effective than change prediction [93].

One weakness of most camera-based LCD systems is their limitation to topological loop closure. Metric methods generalize topological methods, and with the rise in popularity of camera-based methods for SLAM [106, 118, 87], or methods which use both images as well as depth sensors, research on strictly topological LCD may deliver less impact. Identifying the correct topological location is informative, but it generally leaves the hardest part of the question unanswered.

3.2 Depth-based Systems

Similar to camera-based systems, there is some debate about whether point and corner descriptors, object descriptors, or learned similarity functions are superior. These methods are not necessarily mutually exclusive, and indeed the optimal choice likely depends on the deployment context. For instance, corner descriptors are likely an excellent choice for regular, man-made environments such as hospitals or warehouses, whereas learned similarity functions might perform well in environments with less structure, like agriculture or forestry applications. Studies on repeatability for different descriptor designs exist [16], and more work along these lines would likely benefit practitioners. Again though, while innovation on descriptor design does have potential to improve performance of depth-based LCD in some cases, it is more likely that large gains will come from addressing questions at a higher level of abstraction. Table 3 provides an overview of depth-based LCD systems.

Depth-based LCD systems have adopted and integrated tools from deep learning more slowly. However, there is a segment of the computer vision research community which has been generating local depth image descriptors using deep networks in the context of object recognition and model estimation. ShapeNet [158] was introduced to identify 3D objects by training on depth maps from CAD models. Similarly, there has been research on identifying 3D objects based on descriptors generated from one or more 2D views using a CNN [135]. Deep networks have also been trained to produce Fisher and Eigen shape descriptors for identifying simulated 3D objects [40]. PointNet [119] and PointNet++ [120] popularized the concept of using deep learning to generate single- and multi-scale descriptors for 3D depth patches, respectively. Other approaches, such as PCPNet [59] have proposed using deep learning to learn shape properties from depth observations. Another line of research uses deep learning to learn a general similarity function for depth data in an effort to improve data exploration and analysis when designing and debugging robotic perception algorithms [108].

Work	Modality	Map Type	Learning	Technique	Dataset
[22]	2D	Metric	No	Tree of Words	AV Driving
[138]	2D	Metric	No	FLIRT+RANSAC	Intel/fr-clinic
[70]	2D	Metric	No	FALKO+max-clique	Intel/fr-clinic/+
[65]	2D	Metric	No	Submap Matching	Intel/Frieburg/+
[122]	2D	Metric	No	Clustering	Service Robot
[164]	2D	Metric	No	Submap Matching	AV Driving
[56]	2D	Metric	Yes	AdaBoost	AV Driving
[84]	2D	Metric	Yes	CNN	Sim/Real UAV
[17]	2D	Metric	No	Moment Grids+kNN	AV Driving
[100]	2D	Metric	No	max-clique	NCLT
[77]	2D	Topological	Yes	Clustering+NN	Service Robot
[66]	2D+RGB	Metric	No	SIFT+scan stats	AV Driving
[121]	2D+RGB	Metric	Yes	Submaps+CNN	MIT Stata Center
[54]	3D+RGB	Metric	No	Bitmaps+kNN	7-scenes
[50]	3D+RGB	Topo-Metric	No	NBLD/ORB+Voting	KITTI/+
[19]	3D	Metric	No	Gestalt+Voting	Hannover2/+
[127]	3D	Metric	No	FPFH+SAC-IA	AV Driving
[99]	3D	Metric	No	NDT+ad-hoc Score	Hannover2/+
[124]	3D	Metric	No	Histograms	Hannover2/+
[168]	3D	Metric	Yes	SA-NDT+PointNet++	KITTI
[134]	3D	Metric	Yes	NARF+BoW	Hannover2/+
[133]	3D	Metric	No	LoG+ad-hoc Score	Hannover2/+
[60]	3D	Metric	No	ISHOT+Voting	AV Driving
[166]	3D	Metric	No	Clustering	Service Robot
[36]	3D	Metric	Yes	Random Forest	KITTI
[105]	3D	Metric	No	Histogram of Normals	AV Driving
[44]	3D	Topo-Metric	No	Objects+Subgraph ID	Service Robot
[18]	3D	Topo-Metric	No	Moment Grids+Voting	AV Driving
[32]	3D	Topological	No	NBLD+Voting	KITTI/+
[175]	3D	Topological	No	SURF+ad-hoc Score	Service Robot
[63]	3D	Topological	No	M2DP	KITTI/Freiburg/+
[43]	3D	Topological	No	Planes+Subgraph ID	Service Robot

Table 3: LCD systems designed for 2D or 3D depth information

3.3 Takeaways and Trends

Through analysing these tables and reading the individual papers some trends become clear. First, the precision-recall curves for the state-of-the-art in topological place recognition or loop closure in some of the easier domains are impressive. However, there is still a lot of room for improvement in more challenging scenarios. Second, hand-crafted approaches are still extremely popular. Although on the rise, there is room for many more machine learning techniques, particularly those that combine different specialized models instead of creating a single end-to-end model or using machine learning only for feature extraction. Third, there are very few approaches which explicitly model world dynamics, such as lighting or seasonal change. Some data sets contain such changes in which case the model may learn some of these mappings implicitly. However, this leaves the question of how well the method generalizes to other data hard to even hypothesize about. Moreover, many of these dynamics are not tested simultaneously, whereas they will almost certainly be encountered simultaneously during deployment. This is understandable since many of the standard computer vision data sets were not designed with LCD systems in mind.

Nonetheless, standard data sets are invaluable tools for LCD researchers. Current prominent data sets such as KITTI [51] and New College [132] record timestamped camera and laser data, but do not offer many labels that describe the underlying conditions that generated the sensor reading. KITTI offers high-resolution color stereo images as well as velodyne data taken on a vehicle driving around Karlsruhe on both highway and through rural areas. The New College data set provides stereo and omnidirectional camera data and laser range data collected by driving around New College campus and park. These data sets were designed to test trajectory estimation and SLAM among other things but were not designed specifically to stress test LCD systems. We want to develop LCD systems to deal with multiple forms of dynamics, which is a requirement for embodied agents, but we lack explicitly multi-dynamic data sets. This puts LCD researchers who want to study robustness and generalizability at a disadvantage since these and other common data sets represent such a small fraction of sensor signals likely to be encountered in the wild.

Although there are a wide variety of data sets already in circulation, we advocate for yet another data set - one specifically designed for testing LCD systems. This new data set may well be a conglomeration of existing data sets with revamped or expanded labeling. Specifically, we hypothesize that a data set that is easily searchable by each major challenge listed in section 2.2 would drastically improve both the ease of comparisons as well as the ability of researchers to understand success and failure causally. Currently, most major data sets do not have labels for meta-data such as night vs day, season, level of occlusion, weather, etc. at the scale required to support a modular approach to generally capable LCD.

4 Rethinking the Definition of Success

Currently, topological LCD systems are evaluated based on their precision-recall curves on a subset of the available open-source data. As other fields in machine learning have realized, including literature in fields as varied as fairness and bias [57, 48, 15] and long-term autonomy [62, 96], simple precision-recall curves are not typically sufficient to fully understand the characteristics of a given method, especially when in the real-world it is always tightly integrated with a larger system. This is doubly so if it may be used in a high-stakes situation, where its success or failure may impact other agents in a meaningful way. Below, we outline several obvious ways to improve experiments commonly run on LCD systems.

4.1 Not All Loop Closures are Created Equal

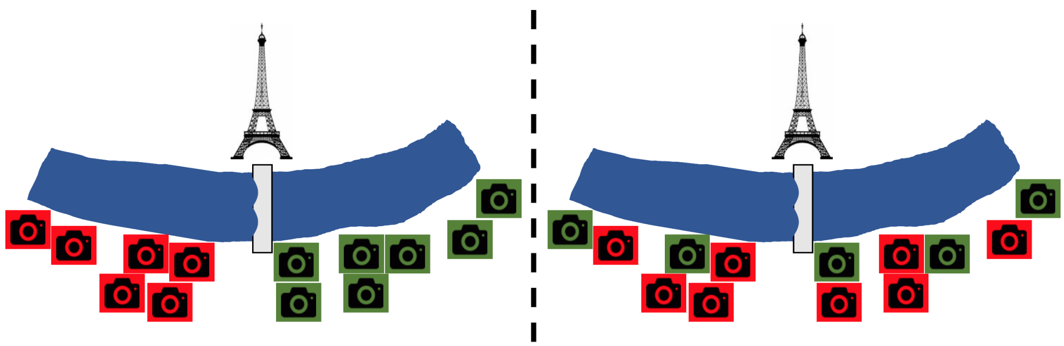

Figure 3: Example loop closure performance. Green and red cameras represent true positives and false negatives, respectively. The position of cameras roughly follows the vantage points in Figure 2.

Given the relationship between LCD systems and the larger SLAM systems that typically invoke them, there is a large asymmetry in the importance of different types of failure cases LCD systems may encounter. As highlighted earlier, false positives are usually more difficult to deal with than false negatives. Additionally, practitioners are often less worried about the precision and recall scores over the entire database, and instead worried about the success or failure rate on a place by place basis. For instance, consider Figure 3. The recall score for the system on the left is about 0.54, while the recall rate for the system on the right is about 0.38. However, in practice the system on the right will likely experience lower error since the time between loop closure events is generally lower. By measuring precision and recall over the entire data set, not only does it make it difficult to evaluate the quality of the algorithm, but it also obfuscates obvious areas of improvement for future researchers.

The misalignment between precision-recall scores and an accurate evaluation of a system's likely effectiveness in the real-world is emblematic of a longstanding tension between explicit or direct tests and implicit or indirect tests. These notions often correlate strongly with unit tests (testing a small part of code for specific functionality) and integration tests (testing how multiple large pieces of code interact), and this tension is by no means unique to LCD. What we are really interested in is the system's overall performance; can it complete the tasks asked of it safely and efficiently? This is the integration test, testing whether all components are working together harmoniously. However, what we are often conducting during experiments are forms of unit tests; does the system realize that these two specific images contain the same object? Unit tests are incredibly useful, but they alone are not sufficient. Often, integration tests can uncover systematic shortcomings that are not visible to individual unit tests. As robotic systems grow in complexity and their applications grow in scale, researchers will need to become adept at designing and running integration-style experiments in addition to baseline unit experiments.

4.2 What Does it Mean to Recognize an Invariant?

Loop closure detection, at its core, is about detecting invariants in an environment external to the agent - an environment largely outside the agent's control. The collection of data sets under many different sensing conditions is an excellent start to empirically investigating invariants and their distribution in different operating environments. There have been some truly massive data collection efforts, and we acknowledge their utility in bringing LCD research to its present state. However, there are limits to this strategy. Gathering enough training data to cover all of the different sensing conditions outlined earlier is becoming unsustainable.

We hypothesize that developing generative models of performance, or, at the very least, enabling systems to generate explanations grounded in world dynamics regarding why the system succeeded or failed in a particular instance, will be an important step in developing generalizable systems. As data sets be come larger and algorithms more complex, the random fluctuations in performance between huge models trained on slightly different data will present an obstacle to effective evaluation, as has already happened in reinforcement learning [64, 68].

Some papers engage this challenge, comparing different methods for place recognition on data from a UAV instead of a car [167]. They found that viewing angle changes contributed most to decreases in performance, whereas the methods seemed to be fairly robust to illumination changes. However, we believe that these questions regarding generalizability are indicative of other fundamental questions about what the system is actually learning. Whether systems are learning the underlying reasons that something appears to be invariant or whether they are just learning by counting is an incredibly important distinction with respect to the system's ability to generalize.

One of our primary arguments for the claim that multiple systems reasoning about different affects are superior to one big network is that multiple specialized systems make it easier to generate explanations. The provenance of different factors in the reasoning process becomes clearer. Moreover, it is easier to conduct proper integration and ablation tests in such setups since there are fewer shared components. Smaller models have many other advantages, including needing less data to train, using less memory, faster evaluation times, and faster training time.

5 Navigating a Human World

Long-term autonomy (LTA) is widely regarded as one of the goals of research on robotics and artificial intelligence, and some have argued for a re-prioritization of efforts in LTA systems towards 'learning via interaction' and 'systems-level integration' [79]. In modern robotics architectures, there are few direct analogs between robotic sub-tasks and human sub-tasks. LCD is a notable exception, and this presents a unique opportunity to conduct a new breed of experiment. The goal of many LCD systems is to recognize invariants in human environments, and it therefore seems likely that we can learn a great deal by studying human navigation and exploration.

We propose a set of new research initiatives which aim to improve the performance of LCD systems by comparing the relative importance of different features in the LCD system to the importance as indicated by a human navigating the same environment. This type of experiment has many variations, including having a human tele-operate a robot and perceiving only the robot's raw sensor data, or only the robot's memoized feature representation, and analyzing the gaze patterns of the humans against different features visible to the robot. Other versions include experiments where different sets of data are given to different groups of humans and their ability to reconstruct the topological or geometric structure of the environment is measured as a function of the different features available.

Situational awareness for human subjects in some experiments is expected to be critical and because of this there is also potential for collaboration with researchers studying virtual and augmented reality and developing new human-robot interfaces. This adds an unfortunate level of complexity to these experiments but has the potential to benefit both LCD researchers as well as HRI and HCI scholars. A Particular challenge we foresee is representing memoized data appropriately, specifically for visualizations. Experiments with communicating memoizations also open up possibilities for non-visual interfaces, such as those using audio or haptic signals.

This line of experimentation may lead to many new validation methods for understanding what LCD systems are learning. In particular, it could lead to sample efficient learning and development cycles since information from humans is usually very directed. Moreover, the benefits of these experiments are likely to influence other human-related fields as well, such as semantic mapping [76, 88], where robots try to infer the purpose of a given location in addition to its geometric or topological properties with respect to the rest of its environment.

6 Evaluating Intelligent Systems

Many sub-fields of artificial intelligence point to a problem or collection of problems as being the holy grail for that sub-field. For example, many natural language processing researchers consider understanding figurative language to represent a significant milestone in natural language understanding [41, 123, 75]. Here, we argue that LCD can and should be the measuring stick for many machine learning, artificial intelligence, and robotics initiatives.

First, however, we want to address a growing body of work focuses on learning end-to-end visual motor policies for indoor navigation via deep neural networks [162]. In these systems, localization and mapping are not represented explicitly and are instead latent via recurrent networks, or are represented in the input itself in the form of large sequences of sensor data. There are several ways to view the relationship between these systems and traditional SLAM systems. First, when initially exploring an environment, there is no available data and so these types of systems cannot be used. However, in a thoroughly explored and mapped environment, such systems may be more robust to noise and dynamic scene changes than traditional localization approaches. In this view, they are less at odds and more complimentary. Second, and more importantly, is the question of generalizability.

It is possible that these end-to-end navigation policies are implicitly solving loop closure detection or re-localization. However, if we want to benefit from this occurrence, more research is necessary in order to understand how these representations develop and under what conditions we can export them to other settings. The benefit of such approaches is that these concepts are seemingly learned automatically, however the downside is that we aren't able to benefit from their having been learned in any other context. Both approaches motivate LCD as an important problem to understand. In one case we would like to know how to extract concepts learned by networks in a more refined manner, and in the other case we would like to know which concepts we can learn directly and how well we can learn them.

To truly solve loop closure, with human-level precision, recall, and *generalizability*, what kind of concepts about the world would our LCD system need to understand, either statistically or explicitly? Certainly, all of the concepts in section 2.2 must be mastered. However, these concepts are just specific instances of general types of abstractions and happen to span a relatively large set, including geometric abstractions, temporal abstraction, and abstraction over dynamic processes in nature. We hypothesize that loop closure detection offers a sort of 'sweet spot' for targeting the next generation of intelligent systems and has many positive characteristics in this respect.

6.1 Accessibility and Control Over Experiments

Although non-trivial to evaluate holistically, it is easy to measure success or failure on individual problem instances. Compared to, for instance, determining if an image was properly captioned, an article was properly summarized, or a piece of text was transformed from one literary style to another, LCD performance at the instance level is well-defined and straightforward. It is also easy to modulate the difficulty of the test environment or test data. This could be made significantly easier with the availability of additional data sets, but nonetheless this is also straightforward. Concepts such as weather, seasons, day vs. night, etc. can be tested independently. This is an important property that is difficult to control for in many other domains and also supports the development of modular LCD systems where different concepts are addressed by different sub-systems.

6.2 Scope and Specificity of Sub-problems

The set of concepts required is large but not overly broad. There are many concepts affecting appearance in Figure 2, but individually their effects on the images are well-defined. These concepts are also specific instances of abstractions that we want intelligent systems in general to master. Generalizing over various timescales and over various geometries is a central component of human problem solving, and LCD presents an excellent test bed for developing those capacities in machines in a limited sense. LCD also provides opportunities for experimenting with methods for both modeling and systems integration since each concept could be modeled independently and the resultant models combined to form a working system. Research on the formulation of novel individual models as well as architectures for model composition would be applicable.

6.3 Potential for Impact and Applicability of Existing Research

Most abstractions required for general loop closure have received considerable attention as components of other problems. We have mentioned several times that LCD bears resemblance to many other machine learning problems and we cannot stress this enough. Many of the sub-tasks in LCD systems appear in other areas of research, and we are confident that advances in LCD algorithms can impact other machine learning endeavors and vice-versa. LCD offers opportunities for discovery with respect to both artificial and biological intelligence. Processes which have such strong analogs between humans and robots are not common, and LCD provides a well-defined task for comparative studies. Lastly, LCD is incredibly useful in its own right given its importance for performant mobile robots. Mobile robots simply cannot operate for extended periods of time in large, complex environments without some form of loop closure or re-localization.

In summary, we believe LCD requires a diverse set of capabilities that are well-enough defined to design a research agenda around, but also ambitious enough to make it a plausible starting point for a truly generally intelligent system. We want to motivate the application of advances in other domains to LCD, and in particular we want to encourage research that probes the underlying capabilities of modern machine learning systems. An example of such work investigates what convolutional neural networks are learning (semantic appearance vs specific location) and where within the architecture they are learning it [136].

7 Conclusion

We want machine learning researchers to view LCD not as an odd by-product of a robotics optimization problem, but as a unique opportunity to test and explore the boundaries of the state-of-the-art for learning many different abstract concepts. This paper presents a brief overview and contextualization of the loop closure detection problem and its situation within mobile robotics and goes on to explore new ways to improve experimental design of LCD systems and generate more interesting research questions. Finally, we make an argument for the importance and uniqueness of loop closure detection as a stepping stone to building generally capable, intelligent, embodied systems.

References

[1] A. Adamu and W. M. N. W. Zainon. Multiview similarity assessment technique of uml diagrams. *Procedia Computer Science*, 124:311–318, 2017.

[2] A. Alahi, R. Ortiz, and P. Vandergheynst. Freak: Fast retina keypoint. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 510–517. Ieee, 2012.

[3] P. F. Alcantarilla and T. Solutions. Fast explicit diffusion for accelerated features in nonlinear scale spaces. *IEEE Trans. Patt. Anal. Mach. Intell*, 34(7):1281–1298, 2011.

[4] A. Aly and A. Tapus. A model for synthesizing a combined verbal and nonverbal behavior based on personality traits in human-robot interaction. In *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 325–332. IEEE, 2013.

[5] F. Amorós, L. Payá, J. M. Marín, and O. Reinoso. Trajectory estimation and optimization through loop closure detection, using omnidirectional imaging and global-appearance descriptors. *Expert Systems with Applications*, 102:273–290, 2018.

[6] R. Anati and K. Daniilidis. Constructing topological maps using markov random fields and loop-closure detection. In *Advances in Neural Information Processing Systems*, pages 37–45, 2009.

[7] A. Angeli, D. Filliat, S. Doncieux, and J.-A. Meyer. Fast and incremental method for loop-closure detection using bags of visual words. *IEEE Transactions on Robotics*, 24(5):1027–1037, 2008.

[8] 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*, pages 5297–5307, 2016.

[9] R. Arroyo, P. F. Alcantarilla, L. M. Bergasa, J. J. Yebes, and S. Gámez. Bidirectional loop closure detection on panoramas for visual navigation. In *2014 IEEE Intelligent Vehicles Symposium Proceedings*, pages 1378–1383. IEEE, 2014.

[10] L. Bampis, A. Amanatiadis, and A. Gasteratos. Encoding the description of image sequences: A two-layered pipeline for loop closure detection. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4530–4536. IEEE, 2016.

[11] L. Bampis, S. Chatzichristofis, C. Iakovidou, A. Amanatiadis, Y. Boutalis, and A. Gasteratos. A locate-based visual place recognition system for mobile robotics and gpgpus. *Concurrency and Computation: Practice and Experience*, 30(7):e4146, 2018.

[12] H. Bay, T. Tuytelaars, and L. Van Gool. Surf: Speeded up robust features. In *European conference on computer vision*, pages 404–417. Springer, 2006.

[13] M. Berlingerio, D. Koutra, T. Eliassi-Rad, and C. Faloutsos. Network similarity via multiple social theories. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 1439–1440, 2013.

[14] A. Bircher, K. Alexis, M. Burri, P. Oettershagen, S. Omari, T. Mantel, and R. Siegwart. Structural inspection path planning via iterative viewpoint resampling with application to aerial robotics. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6423–6430. IEEE, 2015.

[15] S. L. Blodgett, S. Barocas, H. Daumé III, and H. Wallach. Language (technology) is power: A critical survey of" bias" in nlp. *arXiv preprint arXiv:2005.14050*, 2020.

[16] E. R. Boroson and N. Ayanian. 3d keypoint repeatability for heterogeneous multi-robot slam. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 6337–6343. IEEE, 2019.

[17] M. Bosse and R. Zlot. Keypoint design and evaluation for place recognition in 2d lidar maps. *Robotics and Autonomous Systems*, 57(12):1211–1224, 2009.

[18] M. Bosse and R. Zlot. Place recognition using regional point descriptors for 3d mapping. In *Field and Service Robotics*, pages 195–204. Springer, 2010.

[19] M. Bosse and R. Zlot. Place recognition using keypoint voting in large 3d lidar datasets. In *2013 IEEE International Conference on Robotics and Automation*, pages 2677–2684. IEEE, 2013.

[20] T. Botterill, S. Mills, and R. Green. Speeded-up bag-of-words algorithm for robot localisation through scene recognition. In *2008 23rd International Conference Image and Vision Computing New Zealand*, pages 1–6. IEEE, 2008.

[21] G. Bresson, Z. Alsayed, L. Yu, and S. Glaser. Simultaneous localization and mapping: A survey of current trends in autonomous driving. *IEEE Transactions on Intelligent Vehicles*, 2(3):194–220, 2017.

[22] J. Callmer, K. Granström, J. Nieto, and F. Ramos. Tree of words for visual loop closure detection in urban slam. In *Proceedings of the 2008 Australasian conference on robotics and automation*, page 8, 2008.

[23] M. Calonder, V. Lepetit, C. Strecha, and P. Fua. Brief: Binary robust independent elementary features. In *European conference on computer vision*, pages 778–792. Springer, 2010.

[24] S. Cascianelli, G. Costante, E. Bellocchio, P. Valigi, M. L. Fravolini, and T. A. Ciarfuglia. Robust visual semi-semantic loop closure detection by a covisibility graph and cnn features. *Robotics and Autonomous Systems*, 92:53–65, 2017.

[25] A. Celikyilmaz, D. Hakkani-Tur, and G. Tur. Lda based similarity modeling for question answering. In *Proceedings of the NAACL HLT 2010 Workshop on Semantic Search*, pages 1–9, 2010.

[26] A. Chapoulie, P. Rives, and D. Filliat. A spherical representation for efficient visual loop closing. In *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, pages 335–342. IEEE, 2011.

[27] S. M. Chaves, A. Kim, E. Galceran, and R. M. Eustice. Opportunistic sampling-based active visual slam for underwater inspection. *Autonomous Robots*, 40(7):1245–1265, 2016.

[28] T. Chen, K. Wu, K.-H. Yap, Z. Li, and F. S. Tsai. A survey on mobile landmark recognition for information retrieval. In *2009 Tenth International Conference on Mobile Data Management: Systems, Services and Middleware*, pages 625–630. IEEE, 2009.

[29] X. Chen, H. Lu, J. Xiao, H. Zhang, and P. Wang. Robust relocalization based on active loop closure for real-time monocular slam. In *International Conference on Computer Vision Systems*, pages 131–143. Springer, 2017.

[30] 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 *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3223–3230. IEEE, 2017.

[31] Z. Chen, L. Liu, I. Sa, Z. Ge, and M. Chli. Learning context flexible attention model for long-term visual place recognition. *IEEE Robotics and Automation Letters*, 3(4):4015–4022, 2018.

[32] T. Cieslewski, E. Stumm, A. Gawel, M. Bosse, S. Lynen, and R. Siegwart. Point cloud descriptors for place recognition using sparse visual information. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4830–4836. IEEE, 2016.

[33] M. Cummins and P. Newman. Fab-map: Probabilistic localization and mapping in the space of appearance. *The International Journal of Robotics Research*, 27(6):647–665, 2008.

[34] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, volume 1, pages 886–893. IEEE, 2005.

[35] H. Do, S. Hong, and J. Kim. Robust loop closure method for multi-robot map fusion by integration of consistency and data similarity. *IEEE Robotics and Automation Letters*, 5(4):5701–5708, 2020.

[36] R. Dubé, D. Dugas, E. Stumm, J. Nieto, R. Siegwart, and C. Cadena. Segmatch: Segment based place recognition in 3d point clouds. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 5266–5272. IEEE, 2017.

[37] T. Duckett, S. Pearson, S. Blackmore, B. Grieve, W.-H. Chen, G. Cielniak, J. Cleaversmith, J. Dai, S. Davis, C. Fox, et al. Agricultural robotics: the future of robotic agriculture. *arXiv preprint arXiv:1806.06762*, 2018.

[38] G. Dudek and D. Jugessur. Robust place recognition using local appearance based methods. In *Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No. 00CH37065)*, volume 2, pages 1030–1035. IEEE, 2000.

[39] C. Erhan, E. Sariyanidi, O. Sencan, and H. Temeltas. Patterns of approximated localised moments for visual loop closure detection. *IET Computer Vision*, 11(3):237–245, 2016.

[40] Y. Fang, J. Xie, G. Dai, M. Wang, F. Zhu, T. Xu, and E. Wong. 3d deep shape descriptor. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2319–2328, 2015.

[41] A. Feldman and X. Lu. Proceedings of the workshop on computational approaches to figurative language. In *Proceedings of the Workshop on Computational Approaches to Figurative Language*, 2007.

[42] L. Fernández, L. Payá, O. Reinoso, A. Gil, and D. Valiente. Visual hybrid slam: an appearance-based approach to loop closure. In *ROBOT2013: First Iberian Robotics Conference*, pages 693–701. Springer, 2014.

[43] E. Fernández-Moral, W. Mayol-Cuevas, V. Arevalo, and J. Gonzalez-Jimenez. Fast place recognition with plane-based maps. In *2013 IEEE International Conference on Robotics and Automation*, pages 2719–2724. IEEE, 2013.

[44] R. Finman, L. Paull, and J. J. Leonard. Toward object-based place recognition in dense rgb-d maps. In *ICRA Workshop Visual Place Recognition in Changing Environments, Seattle, WA*, volume 76, 2015.

[45] J. Ganitkevitch, B. Van Durme, and C. Callison-Burch. Ppdb: The paraphrase database. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 758–764, 2013.

[46] X. Gao and T. Zhang. Loop closure detection for visual slam systems using deep neural networks. In *2015 34th Chinese Control Conference (CCC)*, pages 5851–5856. IEEE, 2015.

[47] E. Garcia-Fidalgo and A. Ortiz. ibow-lcd: An appearance-based loop-closure detection approach using incremental bags of binary words. *IEEE Robotics and Automation Letters*, 3(4):3051–3057, 2018.

[48] J. Gardner, C. Brooks, and R. Baker. Evaluating the fairness of predictive student models through slicing analysis. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, pages 225–234, 2019.

[49] S. Garg, N. Suenderhauf, and M. Milford. Don't look back: Robustifying place categorization for viewpoint-and condition-invariant place recognition. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3645–3652. IEEE, 2018.

[50] A. Gawel, T. Cieslewski, R. Dubé, M. Bosse, R. Siegwart, and J. Nieto. Structure-based vision-laser matching. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 182–188. IEEE, 2016.

[51] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun. Vision meets robotics: The kitti dataset. *International Journal of Robotics Research (IJRR)*, 2013.

[52] L. Gerstmayr-Hillen, O. Schlüter, M. Krzykawski, and R. Möller. Parsimonious loop-closure detection based on global image-descriptors of panoramic images. In *2011 15th International Conference on Advanced Robotics (ICAR)*, pages 576–581. IEEE, 2011.

[53] M. Giamou, K. Khosoussi, and J. P. How. Talk resource-efficiently to me: Optimal communication planning for distributed loop closure detection. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1–9. IEEE, 2018.

[54] B. Glocker, S. Izadi, J. Shotton, and A. Criminisi. Real-time rgb-d camera relocalization. In *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 173–179. IEEE, 2013.

[55] J. Gomes and A. L. Christensen. Generic behaviour similarity measures for evolutionary swarm robotics. In *Proceedings of the 15th annual conference on Genetic and evolutionary computation*, pages 199–206, 2013.

[56] K. Granstrom, J. Callmer, F. Ramos, and J. Nieto. Learning to detect loop closure from range data. In *2009 IEEE International Conference on Robotics and Automation*, pages 15–22. IEEE, 2009.

[57] N. Grgic-Hlaca, M. B. Zafar, K. P. Gummadi, and A. Weller. Beyond distributive fairness in algorithmic decision making: Feature selection for procedurally fair learning. In *AAAI*, pages 51–60, 2018.

[58] O. Guclu and A. B. Can. Fast and effective loop closure detection to improve slam performance. *Journal of Intelligent & Robotic Systems*, 93(3-4):495–517, 2019.

[59] P. Guerrero, Y. Kleiman, M. Ovsjanikov, and N. J. Mitra. Pcpnet learning local shape properties from raw point clouds. In *Computer Graphics Forum*, volume 37, pages 75–85. Wiley Online Library, 2018.

[60] J. Guo, P. V. Borges, C. Park, and A. Gawel. Local descriptor for robust place recognition using lidar intensity. *IEEE Robotics and Automation Letters*, 4(2):1470–1477, 2019.

[61] W. Hartmann, S. Galliani, M. Havlena, L. Van Gool, and K. Schindler. Learned multi-patch similarity. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1586–1594, 2017.

[62] N. Hawes, C. Burbridge, F. Jovan, L. Kunze, B. Lacerda, L. Mudrova, J. Young, J. Wyatt, D. Hebesberger, T. Kortner, et al. The strands project: Long-term autonomy in everyday environments. *IEEE Robotics & Automation Magazine*, 24(3):146–156, 2017.

[63] L. He, X. Wang, and H. Zhang. M2dp: A novel 3d point cloud descriptor and its application in loop closure detection. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 231–237. IEEE, 2016.

[64] P. Henderson, R. Islam, P. Bachman, J. Pineau, D. Precup, and D. Meger. Deep reinforcement learning that matters. *arXiv preprint arXiv:1709.06560*, 2017.

[65] W. Hess, D. Kohler, H. Rapp, and D. Andor. Real-time loop closure in 2d lidar slam. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1271–1278. IEEE, 2016.

[66] K. L. Ho and P. Newman. Loop closure detection in slam by combining visual and spatial appearance. *Robotics and Autonomous Systems*, 54(9):740–749, 2006.

[67] Y. Hou, H. Zhang, and S. Zhou. Convolutional neural network-based image representation for visual loop closure detection. In *2015 IEEE international conference on information and automation*, pages 2238–2245. IEEE, 2015.

[68] M. Hutson. Artificial intelligence faces reproducibility crisis, 2018.

[69] V. Ila, J. Andrade-Cetto, R. Valencia, and A. Sanfeliu. Vision-based loop closing for delayed state robot mapping. In *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3892–3897. IEEE, 2007.

[70] F. Kallasi and D. L. Rizzini. Efficient loop closure based on falko lidar features for online robot localization and mapping. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1206–1213. IEEE, 2016.

[71] A. Kawewong, N. Tongprasit, and O. Hasegawa. Pirf-nav 2.0: Fast and online incremental appearance-based loop-closure detection in an indoor environment. *Robotics and Autonomous Systems*, 59(10):727–739, 2011.

[72] N. Kejriwal, S. Kumar, and T. Shibata. High performance loop closure detection using bag of word pairs. *Robotics and Autonomous Systems*, 77:55–65, 2016.

[73] S. Khan and D. Wollherr. Ibuild: Incremental bag of binary words for appearance based loop closure detection. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pages 5441–5447. IEEE, 2015.

[74] A. Khotanzad and Y. H. Hong. Invariant image recognition by zernike moments. *IEEE Transactions on pattern analysis and machine intelligence*, 12(5):489–497, 1990.

[75] B. B. Klebanov, E. Shutova, P. Lichtenstein, S. Muresan, C. Wee, A. Feldman, and D. Ghosh. Proceedings of the second workshop on figurative language processing. In *Proceedings of the Second Workshop on Figurative Language Processing*, 2020.

[76] I. Kostavelis and A. Gasteratos. Semantic mapping for mobile robotics tasks: A survey. *Robotics and Autonomous Systems*, 66:86–103, 2015.

[77] B. Kuipers and P. Beeson. Bootstrap learning for place recognition. In *AAAI/IAAI*, pages 174–180, 2002.

[78] R. Kumar and S. P. Singh. A similarity score-based two-phase heuristic approach to solve the dynamic cellular facility layout for manufacturing systems. *Engineering Optimization*, 49(11):1848–1867, 2017.

[79] L. Kunze, N. Hawes, T. Duckett, M. Hanheide, and T. Krajník. Artificial intelligence for long-term robot autonomy: A survey. *IEEE Robotics and Automation Letters*, 3(4):4023–4030, 2018.

[80] M. Labbe and F. Michaud. Appearance-based loop closure detection for online large-scale and long-term operation. *IEEE Transactions on Robotics*, 29(3):734–745, 2013.

[81] Y. Latif, R. Garg, M. Milford, and I. Reid. Addressing challenging place recognition tasks using generative adversarial networks. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2349–2355. IEEE, 2018.

[82] G. H. Lee, F. Fraundorfer, and M. Pollefeys. Structureless pose-graph loop-closure with a multi-camera system on a self-driving car. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 564–571. IEEE, 2013.

[83] S. Leutenegger, M. Chli, and R. Y. Siegwart. Brisk: Binary robust invariant scalable keypoints. In *2011 International conference on computer vision*, pages 2548–2555. Ieee, 2011.

[84] J. Li, H. Zhan, B. M. Chen, I. Reid, and G. H. Lee. Deep learning for 2d scan matching and loop closure. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 763–768. IEEE, 2017.

[85] Y. Li, Z. Hu, G. Huang, Z. Li, and M. A. Sotelo. Image sequence matching using both holistic and local features for loop closure detection. *IEEE Access*, 5:13835–13846, 2017.

[86] B. Liu, C.-C. Li, and K. Yan. Deepsvm-fold: protein fold recognition by combining support vector machines and pairwise sequence similarity scores generated by deep learning networks. *Briefings in bioinformatics*, 21(5):1733–1741, 2020.

[87] H. Liu, M. Chen, G. Zhang, H. Bao, and Y. Bao. Ice-ba: Incremental, consistent and efficient bundle adjustment for visual-inertial slam. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1974–1982, 2018.

[88] Q. Liu, R. Li, H. Hu, and D. Gu. Extracting semantic information from visual data: A survey. *Robotics*, 5(1):8, 2016.

[89] Y. Liu and H. Zhang. Indexing visual features: Real-time loop closure detection using a tree structure. In *2012 IEEE International conference on robotics and automation*, pages 3613–3618. IEEE, 2012.

[90] Y. Liu and H. Zhang. Visual loop closure detection with a compact image descriptor. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1051–1056. IEEE, 2012.

[91] Y. Liu and H. Zhang. Performance evaluation of whole-image descriptors in visual loop closure detection. In *2013 IEEE International Conference on Information and Automation (ICIA)*, pages 716–722. IEEE, 2013.

[92] D. G. Lowe. Object recognition from local scale-invariant features. In *Proceedings of the seventh IEEE international conference on computer vision*, volume 2, pages 1150–1157. Ieee, 1999.

[93] S. Lowry and M. J. Milford. Supervised and unsupervised linear learning techniques for visual place recognition in changing environments. *IEEE Transactions on Robotics*, 32(3):600–613, 2016.

[94] S. Lowry, N. Sünderhauf, P. Newman, J. J. Leonard, D. Cox, P. Corke, and M. J. Milford. Visual place recognition: A survey. *IEEE Transactions on Robotics*, 32(1):1–19, 2015.

[95] W. L. D. Lui and R. Jarvis. A pure vision-based approach to topological slam. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3784–3791. IEEE, 2010.

[96] M. Luperto, J. Monroy, J. R. Ruiz-Sarmiento, F.-A. Moreno, N. Basilico, J. Gonzalez-Jimenez, and N. A. Borghese. Towards long-term deployment of a mobile robot for at-home ambient assisted living of the elderly. In *2019 European Conference on Mobile Robots (ECMR)*, pages 1–6. IEEE, 2019.

[97] S. Lynen, M. Bosse, P. Furgale, and R. Siegwart. Placeless place-recognition. In *2014 2nd International Conference on 3D Vision*, volume 1, pages 303–310. IEEE, 2014.

[98] F. Maffra, L. Teixeira, Z. Chen, and M. Chli. Loop-closure detection in urban scenes for autonomous robot navigation. In *2017 International Conference on 3D Vision (3DV)*, pages 356–364. IEEE, 2017.

[99] M. Magnusson, H. Andreasson, A. Nuchter, and A. J. Lilienthal. Appearance-based loop detection from 3d laser data using the normal distributions transform. In *2009 IEEE International Conference on Robotics and Automation*, pages 23–28. IEEE, 2009.

[100] J. G. Mangelson, D. Dominic, R. M. Eustice, and R. Vasudevan. Pairwise consistent measurement set maximization for robust multi-robot map merging. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2916–2923. IEEE, 2018.

[101] L. Maohai, W. Han, S. Lining, and C. Zesu. Robust omnidirectional mobile robot topological navigation system using omnidirectional vision. *Engineering applications of artificial intelligence*, 26(8):1942–1952, 2013.

[102] A. R. Memon, H. Wang, and A. Hussain. Loop closure detection using supervised and unsupervised deep neural networks for monocular slam systems. *Robotics and Autonomous Systems*, 126:103470, 2020.

[103] N. Merrill and G. Huang. Lightweight unsupervised deep loop closure. *arXiv preprint arXiv:1805.07703*, 2018.

[104] M. Milford, W. Scheirer, E. Vig, A. Glover, O. Baumann, J. Mattingley, and D. Cox. Condition-invariant, top-down visual place recognition. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 5571–5577. IEEE, 2014.

[105] N. Muhammad and S. Lacroix. Loop closure detection using small-sized signatures from 3d lidar data. In *2011 IEEE International Symposium on Safety, Security, and Rescue Robotics*, pages 333–338. IEEE, 2011.

[106] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos. Orb-slam: a versatile and accurate monocular slam system. *IEEE transactions on robotics*, 31(5):1147–1163, 2015.

[107] T. Naseer, L. Spinello, W. Burgard, and C. Stachniss. Robust visual robot localization across seasons using network flows. In *AAAI*, pages 2564–2570, 2014.

[108] S. B. Nashed. Laser2vec: Similarity-based retrieval for robotic perception data. *arXiv preprint arXiv:2007.15746*, 2020.

[109] P. Newman, D. Cole, and K. Ho. Outdoor slam using visual appearance and laser ranging. In *Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006.*, pages 1180–1187. IEEE, 2006.

[110] T. Nicosevici and R. Garcia. On-line visual vocabularies for robot navigation and mapping. In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 205–212. IEEE, 2009.

[111] J. H. Oh, G. Eoh, and B. H. Lee. Appearance-based place recognition using whole-image brisk for collaborative multi-robot localization. *International Journal of Mechanical Engineering and Robotics Research*, 4(3):264–268, 2015.

[112] J. C. Osborn and M. Mateas. A game-independent play trace dissimilarity metric. In *FDG*, 2014.

[113] L. Paull, S. Saeedi, M. Seto, and H. Li. Auv navigation and localization: A review. *IEEE Journal of Oceanic Engineering*, 39(1):131–149, 2013.

[114] E. Pepperell, P. I. Corke, and M. J. Milford. All-environment visual place recognition with smart. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1612–1618, 2014.

[115] M. Pfingsthorn and A. Birk. Generalized graph slam: Solving local and global ambiguities through multimodal and hyperedge constraints. *The International Journal of Robotics Research*, 35(6):601–630, 2016.

[116] M. Pfingsthorn, A. Birk, and H. Bülow. An efficient strategy for data exchange in multi-robot mapping under underwater communication constraints. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 4886–4893. IEEE, 2010.

[117] A. Pronobis, B. Caputo, P. Jensfelt, and H. I. Christensen. A discriminative approach to robust visual place recognition. In *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3829–3836. IEEE, 2006.

[118] A. Pumarola, A. Vakhitov, A. Agudo, A. Sanfeliu, and F. Moreno-Noguer. Pl-slam: Real-time monocular visual slam with points and lines. In *2017 IEEE international conference on robotics and automation (ICRA)*, pages 4503–4508. IEEE, 2017.

[119] C. R. Qi, H. Su, K. Mo, and L. J. Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 652–660, 2017.

[120] C. R. Qi, L. Yi, H. Su, and L. J. Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In *Advances in neural information processing systems*, pages 5099–5108, 2017.

[121] H. Qin, M. Huang, J. Cao, and X. Zhang. Loop closure detection in slam by combining visual cnn features and submaps. In *2018 4th International Conference on Control, Automation and Robotics (ICCAR)*, pages 426–430. IEEE, 2018.

[122] A. A. Ravankar, Y. Kobayashi, and T. Emaru. Clustering based loop closure technique for 2d robot mapping based on ekf-slam. In *2013 7th Asia Modelling Symposium*, pages 72–77. IEEE, 2013.

[123] A. Reyes, P. Rosso, and D. Buscaldi. From humor recognition to irony detection: The figurative language of social media. *Data & Knowledge Engineering*, 74:1–12, 2012.

[124] T. Röhling, J. Mack, and D. Schulz. A fast histogram-based similarity measure for detecting loop closures in 3-d lidar data. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 736–741. IEEE, 2015.

[125] S. Rohou, P. Franek, C. Aubry, and L. Jaulin. Proving the existence of loops in robot trajectories. *The International Journal of Robotics Research*, 37(12):1500–1516, 2018.

[126] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski. Orb: An efficient alternative to sift or surf. In *2011 International conference on computer vision*, pages 2564–2571. Ieee, 2011.

[127] R. B. Rusu, N. Blodow, and M. Beetz. Fast point feature histograms (fpfh) for 3d registration. In *2009 IEEE international conference on robotics and automation*, pages 3212–3217. IEEE, 2009.

[128] M. J. Schuster, K. Schmid, C. Brand, and M. Beetz. Distributed stereo vision-based 6d localization and mapping for multi-robot teams. *Journal of Field Robotics*, 36(2):305–332, 2019.

[129] S. Severini, V. Hangya, A. Fraser, and H. Schütze. Lmu bilingual dictionary induction system with word surface similarity scores for bucc 2020. In *Proceedings of the 13th Workshop on Building and Using Comparable Corpora*, pages 49–55, 2020.

[130] H. Shahbazi and H. Zhang. Application of locality sensitive hashing to realtime loop closure detection. In *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1228–1233. IEEE, 2011.

[131] D.-W. Shin and Y.-S. Ho. Local patch descriptor using deep convolutional generative adversarial network for loop closure detection in slam. In *2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pages 546–549. IEEE, 2017.

[132] M. Smith, I. Baldwin, W. Churchill, R. Paul, and P. Newman. The new college vision and laser data set. *The International Journal of Robotics Research*, 28(5):595–599, 2009.

[133] B. Steder, G. Grisetti, and W. Burgard. Robust place recognition for 3d range data based on point features. In *2010 IEEE International Conference on Robotics and Automation*, pages 1400–1405. IEEE, 2010.

[134] B. Steder, M. Ruhnke, S. Grzonka, and W. Burgard. Place recognition in 3d scans using a combination of bag of words and point feature based relative pose estimation. In *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1249–1255, 2011.

[135] H. Su, S. Maji, E. Kalogerakis, and E. Learned-Miller. Multi-view convolutional neural networks for 3d shape recognition. In *Proceedings of the IEEE international conference on computer vision*, pages 945–953, 2015.

[136] N. Sünderhauf, S. Shirazi, F. Dayoub, B. Upcroft, and M. Milford. On the performance of convnet features for place recognition. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4297–4304. IEEE, 2015.

[137] Q. Suo, F. Ma, Y. Yuan, M. Huai, W. Zhong, A. Zhang, and J. Gao. Personalized disease prediction using a cnn-based similarity learning method. In *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 811–816. IEEE, 2017.

[138] G. D. Tipaldi, M. Braun, and K. O. Arras. Flirt: Interest regions for 2d range data with applications to robot navigation. In *Experimental Robotics*, pages 695–710. Springer, 2014.

[139] 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*, pages 883–890, 2013.

[140] K. A. Tsintotas, L. Bampis, and A. Gasteratos. Assigning visual words to places for loop closure detection. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1–7. IEEE, 2018.

[141] M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. In *Proceedings. 1991 IEEE computer society conference on computer vision and pattern recognition*, pages 586–587. IEEE Computer Society, 1991.

[142] I. Ulrich and I. Nourbakhsh. Appearance-based place recognition for topological localization. In *Proceedings 2000 ICRA. Millennium Conference. IEEE International Conference on Robotics and Automation. Symposia Proceedings (Cat. No. 00CH37065)*, volume 2, pages 1023–1029. Ieee, 2000.

[143] J. J. van Vuuren, L. Tang, I. Al-Bahadly, and K. Arif. Towards the autonomous robotic gripping and handling of novel objects. In *2019 14th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, pages 1812–1817. IEEE, 2019.

[144] S. Vanderplas, M. Nally, T. Klep, C. Cadevall, and H. Hofmann. Comparison of three similarity scores for bullet lea matching. *Forensic Science International*, 308:110167, 2020.

[145] K. M. Varadarajan. Topological mapping for robot navigation using affordance features. In *2015 6th International Conference on Automation, Robotics and Applications (ICARA)*, pages 42–49. IEEE, 2015.

[146] T. A. Vidal-Calleja, C. Berger, and S. Lacroix. Event-driven loop closure in multi-robot mapping. In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1535–1540. IEEE, 2009.

[147] A. Visser and B. A. Slamet. Including communication success in the estimation of information gain for multi-robot exploration. In *2008 6th International symposium on modeling and optimization in mobile, ad hoc, and wireless networks and workshops*, pages 680–687. IEEE, 2008.

[148] A. Vokhmintcev, M. Timchenko, and K. Alina. Real-time visual loop-closure detection using fused iterative close point algorithm and extended kalman filter. In *2017 International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM)*, pages 1–6. IEEE, 2017.

[149] M.-L. Wang and H.-Y. Lin. An extended-hct semantic description for visual place recognition. *The International Journal of Robotics Research*, 30(11):1403–1420, 2011.

[150] Z. Wang, Y. Shen, B. Cai, and M. T. Saleem. A brief review on loop closure detection with 3d point cloud. In *2019 IEEE International Conference on Real-time Computing and Robotics (RCAR)*, pages 929–934. IEEE, 2019.

[151] R. Whalen, A. Lungeanu, L. DeChurch, and N. Contractor. Patent similarity data and innovation metrics. *Journal of Empirical Legal Studies*, 17(3):615–639, 2020.

[152] B. Williams, M. Cummins, J. Neira, P. Newman, I. Reid, and J. Tardós. An image-to-map loop closing method for monocular slam. In *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2053–2059. IEEE, 2008.

[153] B. Williams, M. Cummins, J. Neira, P. Newman, I. Reid, and J. Tardós. A comparison of loop closing techniques in monocular slam. *Robotics and Autonomous Systems*, 57(12):1188–1197, 2009.

[154] L. Wolf, T. Hassner, and I. Maoz. Face recognition in unconstrained videos with matched background similarity. In *CVPR 2011*, pages 529–534. IEEE, 2011.

[155] L. Wolf and N. Levy. The svm-minus similarity score for video face recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3523–3530, 2013.

[156] J. Wu, H. Zhang, and Y. Guan. An efficient visual loop closure detection method in a map of 20 million key locations. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 861–866. IEEE, 2014.

[157] J. Wu, H. Zhang, and Y. Guan. Visual loop closure detection by matching binary visual features using locality sensitive hashing. In *Proceeding of the 11th World Congress on Intelligent Control and Automation*, pages 940–945. IEEE, 2014.

[158] Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, and J. Xiao. 3d shapenets: A deep representation for volumetric shapes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1912–1920, 2015.

[159] Y. Xia, J. Li, L. Qi, and H. Fan. Loop closure detection for visual slam using pcanet features. In *2016 international joint conference on neural networks (IJCNN)*, pages 2274–2281. IEEE, 2016.

[160] Y. Xia, J. Li, L. Qi, H. Yu, and J. Dong. An evaluation of deep learning in loop closure detection for visual slam. In *2017 IEEE international conference on internet of things (iThings) and IEEE green computing and communications (GreenCom) and IEEE cyber, physical and social computing (CPSCom) and IEEE smart data (SmartData)*, pages 85–91. IEEE, 2017.

[161] G. Yang, Z. Chen, Y. Li, and Z. Su. Rapid relocation method for mobile robot based on improved orb-slam2 algorithm. *Remote Sensing*, 11(2):149, 2019.

[162] X. Ye and Y. Yang. From seeing to moving: A survey on learning for visual indoor navigation (vin). *arXiv preprint arXiv:2002.11310*, 2020.

[163] W.-t. Yih, K. Toutanova, J. C. Platt, and C. Meek. Learning discriminative projections for text similarity measures. In *Proceedings of the fifteenth conference on computational natural language learning*, pages 247–256, 2011.

[164] G. Yin, Z. Sun, and Y. Dai. Global loop closure detection of multi-robot based on particle filter. In *2018 3rd International Conference on Robotics and Automation Engineering (ICRAE)*, pages 92–96. IEEE, 2018.

[165] J. Yin, D. Li, and G. He. Mobile robot loop closure detection using endpoint and line feature visual dictionary. In *2017 2nd International Conference on Robotics and Automation Engineering (ICRAE)*, pages 73–78. IEEE, 2017.

[166] H. Yu, H.-W. Chae, and J.-B. Song. Place recognition based on surface graph for a mobile robot. In *2017 14th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI)*, pages 342–346. IEEE, 2017.

[167] 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? *arXiv preprint arXiv:1904.07967*, 2019.

[168] A. Zaganidis, A. Zerntev, T. Duckett, G. Cielniak, et al. Semantically assisted loop closure in slam using ndt histograms. 2019.

[169] G. Zhang, D. H. Kang, and I. H. Suh. Loop closure through vanishing points in a line-based monocular slam. In *2012 IEEE International Conference on Robotics and Automation*, pages 4565–4570. IEEE, 2012.

[170] G. Zhang, M. J. Lilly, and P. A. Vela. Learning binary features online from motion dynamics for incremental loop-closure detection and place recognition. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pages 765–772. IEEE, 2016.

[171] H. Zhang. Borf: Loop-closure detection with scale invariant visual features. In *2011 IEEE International conference on robotics and automation*, pages 3125–3130. IEEE, 2011.

[172] X. Zhang, Y. Su, and X. Zhu. Loop closure detection for visual slam systems using convolutional neural network. In *2017 23rd International Conference on Automation and Computing (ICAC)*, pages 1–6. IEEE, 2017.

[173] L. Zhiwei, G. Xiang, C. Yanyan, and Z. Songhao. A novel loop closure detection method in monocular slam. *Intelligent Service Robotics*, 6(2):79–87, 2013.

[174] G. Zhu and C. A. Iglesias. Sematch: Semantic similarity framework for knowledge graphs. *Knowledge-Based Systems*, 130:30–32, 2017.

[175] Y. Zhuang, N. Jiang, H. Hu, and F. Yan. 3-d-laser-based scene measurement and place recognition for mobile robots in dynamic indoor environments. *IEEE Transactions on Instrumentation and Measurement*, 62(2):438–450, 2012.