Evaluation criteria for appearance based maps

Gorkem Erinc, Stefano Carpin
School of Engineering
5200 N Lake Rd
University of California, Merced, USA
{gerinc,scarpin}@ucmerced.edu

ABSTRACT

We present a set of task-based performance evaluation criteria designed to measure the quality of appearance based maps. Instead of aiming to measure a map's overall goodness, metrics defined in this paper focus on individual tasks, namely localization, planning, and navigation, and the quality of the map with respect to the their successful execution. The performance of a map in terms of localization is measured by the amount of information captured from the environment and the accuracy of this information. The planning metric favors instead maps with high connectivity and measures the validity of these connections. The navigation criterion, on the other hand, computes the robustness and stability associated with the paths that a robot will extract from the map. These metrics are tested on appearance maps created in our lab and their distinctiveness is shown.

1. INTRODUCTION

Perception is one of the keys to build intelligent robotic systems operating in unstructured environments. Most often robots gather information through their sensors and build a map, i.e. spatial model of the environment they operate in. By using this internal representation they can accomplish complex tasks and autonomously operate in a given environment. Therefore, not surprisingly mapping is one of the most studied problems in robotics. Up to now most of the research efforts have been devoted to metric maps, i.e. maps providing metric information about the elements in the map [18]. However, thanks to recent developments in sensor technology and computer vision algorithms, appearance based maps have recently surfaced and are gaining momentum[1]. Systems based on omni-directional [20] and monocular cameras [15, 9] work in image space and do not necessarily need metric localization for mapping or navigation purposes. In other words, localization and mapping

---

[1]For sake of completeness one should also mention a third approach, namely topological maps. These however will not be further considered in this paper

hinges on an *appearance based map.*

An appearance map[2] is an undirected weighted graph $G = (V, E)$ in which each vertex $v \in V$ represents an image captured by a camera at a certain position in the workspace. An edge $e_{ij} \in E$ connects two distinct vertices v_i, v_j whenever the associated images are sufficiently similar, accordingly to a given similarity metric. A frequent way to define similarity between two images is to extract salient features from each image and count the number of common ones. The number of matching features is considered as the indication of similarity and is then set as the weight of that edge. Hence, the weight w_{ij} associated with an edge measures the similarity between images assigned to vertices v_i and v_j. An example appearance graph is shown in Fig. 1.
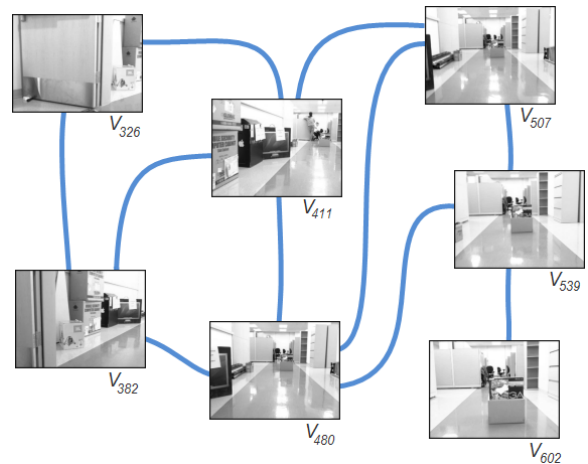

Figure 1: The figure shows an appearance based map with 7 vertices. Edges are added between sufficiently similar images.

Appearance based maps are gaining momentum because they can be built with the only aid of a monocular camera, thus offering a cheap alternative to solutions based on different sensors. Also, by incorporating visual elements, they may be easier to integrate into intelligent systems expected to interact with untrained human users.

In this paper we propose a set of evaluation criteria for appearance based maps. The topic of map benchmarking

---

[2]At the moment there is no universally accepted definition of appearance map, however the one we embrace in this manuscript is consistent with a significant fraction of formerly published papers on the topic.

has recently generated much interest also for metric maps. However, despite some proposals, no consensus exists yet for metric maps (see [2] and references therein). Due to their more recent introduction, no metrics have been proposed yet for appearance based maps. There are many reasons to justify the creation of metrics for appearance based maps, and for maps in general. The main one is that robotics researchers are in great need of objective quality measures to evaluate the impact of the countless ideas proposed in this quickly growing field. To date, for most problems it is impossible to compare two different solutions according to widely accepted criteria. The peculiar nature of robotics research also undermines the experimental replication of published results. However, given the recent dramatic changes in information technology, it is nowadays possible and needed to converge towards accepted performance measures, and to disseminate experimental data so that these assessments can be performed and repeated by an arbitrary third party.

The intrinsic purpose of a map evaluation metric is to measure the quality of different maps and determine which map is the best in general or in terms of some specific criteria. The maps to be evaluated can be generated by different agents using different algorithms. On the other hand, this metric can also be used to measure the effectiveness of one part of the mapping algorithm. For instance, by using the same set of images different appearance maps can be built by changing the parameters or the algorithms used in the map building procedure. In this paper, focusing on the former goal, we present evaluation criteria measuring the quality of a map independently of the algorithm used to build it. Hence, during the evaluation of a map the building blocks of the map creation algorithm are treated as black boxes, and the metric is designed in a way that minimizes the effects of these building blocks.

2. EVALUATION OF VISUAL MAPS

The ideal quality-assessment of a map would be performed by comparing it to the ground truth, i.e. the real value of the variable to be estimated. However, in most cases ground truth might not be easily accessible, or it may be time consuming to acquire. This is a well-known problem for any kind of map. In addition, approaches based on appearance graphs sample 3D environments with 2D images, and it is therefore impractical to generate a map serving as ground truth. In fact, the real environment is the only source which can be used as ground truth. Hence, the quality of an appearance graph can be best measured by evaluating how well it captures the desired properties of the environment it models.

A map by definition is a representation of the environment, and therefore, it may be natural to conclude that the map with most resemblance has the utmost quality. However, this point of view skips the basic motivation behind the need of a map, i.e. its utilization for the successful completion of a given task. A robot creates some form of an internal representation of the environment as a tool to successfully achieve its assigned mission. Thus, for the map to be most useful to the robot, it has to offer enough information to complete the assignment. Consider for example a robot performing inside a warehouse whose task is to quickly navigate between target locations revealed during the mission. One cannot rule out the possibility that a robot using a map with good geometric accuracy, but possibly cumber-

some, may be outperformed by one using a model with an inferior geometric accuracy but easier to process.

With this motivation, we advocate that the assessment shall not be detached form the task at hand. Despite the fact that this statement may seem fairly trivial, in the context of occupancy grid maps one faces a significant amount of scholar work where maps are most often treated as images, and therefore contrasted and evaluated using algorithms and metrics that have little to do with the ultimate task the maps are needed for. On the contrary, we propose three task-centric evaluation criteria, namely localization, navigation, and planning.

2.1 Localization

The ability to estimate its own position is one of the very fundamental robot abilities enabling the successful completion of a variety of tasks. Hence, it is natural to evaluate a map with respect to its usefulness for localization. Unlike in metric maps, localization in appearance graphs is realized by finding within the map the image most similar to the one perceived by the robot when it needs to localize itself[3]. The robot is declared lost if it cannot localize itself to any image in the map.

The utility of the map with respect to localization is measured by the robot's localization performance using that map. A good metric should assign high utility to maps providing good localization. The goodness of localization in appearance graphs can be defined in two ways: 1) coverage 2) accuracy. The *coverage* metric measures the amount of information in the environment captured by the images in the appearance graph. To this end, we propose to collect a set of pictures captured at random locations in the environment where the map is created and use them as query images.

This method of evaluation of a map's merit based on its localization performance requires the use of an image retrieval algorithm in order to find the vertex with the most similar image in the graph. Over the years, several nearest neighbor search based approaches using kd-trees and k-means clustering have been proposed as solutions for the image retrieval problem [9, 14]. Today, SIFT features [11] and the bag-of-words image representation [17] are at the core of state-of-art large scale image retrieval systems. Hence, several versions of bag-of-words algorithm have been proposed as the basis of the localization procedure within a visual map [1, 8, 10, 19]. Nevertheless, even the most successful image retrieval algorithms have sub-perfect performances. In other words, the image retrieval algorithm used for localization can return an invalid image/feature matching resulting in wrong localization. Therefore, the erroneous localization ratio depends on the performance of the image retrieval algorithm and less on the quality of the map being evaluated. In order to minimize the effect of the image retrieval algorithm on the quality score of the map, the proposed method will ignore false image matchings and focus only on valid image correspondences.

Having set the scene, the coverage of a map is defined as

---

[3]We acknowledge this standpoint is somewhat simplified because one would likely get better results by considering not just the current image, but a sequence. This extension is subject to further investigations.

the percentage of successful localizations:

$$l_{cov} = \frac{L_{success}}{L_{total}}. \qquad (1)$$

Localization accuracy, on the other hand, describes the closeness of the estimated location to the real location of the robot. From the appearance graph perspective the robot will benefit from an accurate localization when the returned and query image look alike. Since similarity between images is encoded by the number of matched features, the accuracy has a positive correlation with the number of correct feature matches. Besides sharing features originating from the same objects in the environment, in similar images these objects should appear in close proximity in terms of image coordinates. In order to capture this idea, the accuracy error is defined as the average distance between corresponding features' image coordinates.

$$l_{acc} = \frac{\sum_{i=1}^{N} d\left(a_i, q_i\right)}{N} \qquad (2)$$

where $d(.)$ is the Euclidean distance function, a and q are corresponding features from the retrieved and query image respectively, and N is the number of feature matches between these two images.

Then, the average localization accuracy error of all queries is assigned as the overall accuracy error of the map.

The localization accuracy metric favors appearance graphs with large number of vertices since the probability of having a similar image for a random query image increases with the number of images in the map. On the other hand, if the number of vertices in the map are limited due to the amount of data that can be stored, then a map consisting of images taken all around the environment will be preferred since it will reduce the number of times the robot gets lost. However, when the robot localizes itself, the localization will have less accuracy for a random query image since the images are spread around. Hence, for a fixed number of vertices allowed in the map, there is a balance between accuracy and coverage.

The reader should also note that neither accuracy nor coverage metrics consider the edges in the graph, but they both focus exclusively on the amount of information captured by the images encoded in the vertices.

2.2 Planning

In order to reach full autonomy, robots need to choose their actions by themselves. Planning can be realized using the so-called state space which provides an abstraction of the overall system. Similarly, an appearance graph with vertices corresponding to states and edges corresponding to actions can be used for path planning, i.e. finding the shortest path in the appearance space between two nodes in the graph. In the literature there are several search algorithms which work directly on undirected weighted graphs such as Dijkstra's algorithm. The time complexity of any planning algorithm working on graphs will increase with the number of vertices and edges. Hence, from the perspective of planning efficiency, maps with less vertices and edges will be preferred. Furthermore, the number of extracted features are correlated with the number of vertices and have a direct effect on the running time of the planner since most of the image matching algorithms utilize nearest neighbor search in the high dimensional feature space. Hence, it can be concluded that the amount of data required to store the map is an indication of its performance in planning tasks, and data size comparison should be a part of the general map evaluation process.

For planning tasks, it is important that a map is well connected and contains a good representation of passages between places. Such a map will have a higher utilization rate than a map which reflects an accurate and detailed representation of the environment in general, but also includes an edge between two images of two places that are physically separated. Hence, a metric to measure the usefulness of a map with respect to path planning should consider the ability to plan paths that are valid in the real environment. Inspired from the work by Collins et. al. [5], we would like to measure the validity of the paths in a map by comparing paths generated within both the appearance graph and the actual environment.

In this test representing the *exactness* of the map, the ratio of paths that are valid in the map, but invalid the real environment will be calculated. To measure the ratio of these paths, also known as *false positives*, two connected vertices are randomly selected from the appearance graph. Then, it is tested whether two places identified from these images are actually connected in the real environment. The physical connectivity assessment can be performed either visually or by teleoperating the robot which constructed the map from one place to another. Even though one can argue that the visual inspection introduces some subjectivity into the metric, for robots with well-known kinematics and structured environments it is expected it will provide an accurate approximation. A generated path failing this test is identified as the false positive. The false positive ratio is then estimated by repeatedly generating random paths and counting the ones that fail the connectivity test.

Alternatively, as proposed in [5] to evaluate metric maps, a map can also be evaluated in terms of *completeness* by the ratio of the paths that actually exist in the environment, but are not captured in the map, i.e. so-called *false negatives*. Borrowing the same idea that we used to compute false positive paths, two images from the map are randomly selected and a graph search algorithm is used to find a path connecting these two images. If this path cannot be computed due to the fact that two images lie in disconnected components of the graph but the path exists in the real environment, the path will be declared as a false negative. The ratio of false negatives will reflect the inability of the planner to find a path using the map.

One could argue that two random images should be randomly selected from the environment instead of the map if the completeness of the map is to be measured. However, in that case, first, each random image sampled from the environment should be localized in the map. As stated in section 2.1, this localization procedure measures the coverage of the map and may result in failure either if the map has low coverage score or if the localization algorithm fails to find the valid match in the map. Therefore, to measure the utility of each task separately, query images are sampled from the images in the map. Applying this criterion in the selection of query images, this metric, however, will return zero false negative scores for any appearance graph consisting of only one connected segment since there will always be a path between any two vertices of the graph.

2.3 Navigation

Navigation, defined as the ability to reach an assigned target location from the robot's current position, is another fundamental level of competence that is greatly influenced by the available spatial model.

The evaluation criterion based on the planning task defined in section 2.2 punishing invalid paths generated using the map also evaluates maps with respect to the robust navigation criterion. Maps receiving high scores in this metric will be the ones representing environments that consist of physically apart places with similar appearances. Due to the computed similarity between them, the mapping algorithm may create an edge between the images of these places and since an edge encoding a similarity is interpreted as traversable, a graph search will return a path in between which is in fact invalid. Thus, a map with high false positive ratio will provide the robot a medium in which the probability of bumping into an obstacle is high. However, the contrary is not necessarily true. In a different scenario, navigation between two vertices v_i and v_k sharing an edge due to their similar visual signatures may not be possible. In other words, a visual similarity criterion for edge creation depends on the map building algorithm and an edge may not necessarily encode an indication of traversibility. This path will not be distinguished by the planning metric and classified as a false positive since these vertices may be indeed physically connected. However, the robot trying to follow this direct path may not be able to navigate in between.

Motivated from this idea the quality of the map should also be evaluated with respect to navigational criterion. We therefore embraced the idea of measuring the navigability of a map by computing the stability and robustness of its internal paths in terms of robot navigation. This metric is applied to the random paths generated to evaluate the planning quality. Since most of the proposed algorithms [7, 4, 13, 6, 16] base their servoing algorithm on multi-view geometry or more specifically on the fundamental matrix, we grasped the idea of measuring the quality of the fundamental matrix, F, for each pair of consecutive images in the path. Given two matched images their corresponding features are extracted. Then, a RANSAC algorithm as described in [12] is utilized to compute the fundamental matrix based on a number of randomly selected tentative feature matches. Based on [3], the error in the computed fundamental matrix is computed as the mean distance of points to their epipolar lines in the second image.

$$Err_F = \frac{\sum_{i=1}^{N} d\left(\overline{\mathbf{x}}_i, F\mathbf{x}_i\right) + dist\left(\mathbf{x}_i, F^T\overline{\mathbf{x}}_i\right)}{2N}$$

where $(\mathbf{x}_i, \overline{\mathbf{x}}_i)$, $i = 1, \ldots, N$ are the N corresponding features and $d(.)$ is the distance between a point and a line.

The average error of the fundamental matrix along the path is computed and returned as navigation error of the path. The overall navigation error of the map is then assigned as the average navigation error of all randomly created paths.

3. RESULTS

In this section we exemplify how the proposed metric work on a set of appearance based maps. In order to capture a broader spectrum of maps, a P3AT robot equipped with a monocular camera is teleoperated in the environment twice resulting in image sequences from two independent runs. These images are fed as the input to our map creation algorithm described in [7]. The paths followed during these runs and the test runs are shown in figure 2. A total of 1064 test images are collected by teleoperating a P3AT robot in the environment. The test image sequence is divided into three subpaths as shown in different colors in Fig. 2.

In order to increase heterogeneity, different parameters are used while building the two maps. In the first run (represented with blue circles in Fig. 2), a higher image capturing frequency is used resulting in map, m_1 with a larger number of images. Furthermore, comparing to the second map m_2, a more conservative approach is taken in the image retrieval algorithm by increasing the required number of feature matches for an image in the map to be matched with a query image. As a result two very different appearance graphs are created.

3.1 Localization

The core component in localization procedure in appearance graphs is the image matching function. As mentioned before, the quality of the localization results are mainly dependent on the image retrieval algorithm. Therefore, image matches should be verified for each localization test. Since no ground truth data is available stating which image matches are valid, each image match has to be visually verified by the person evaluating the map. Due to the simplicity of an image match verification task to humans, we believe at this point no subjectivity is introduced into the evaluation. This way false positives are eliminated. In order to also account for false negatives, a human should verify if a query image has actually a corresponding image in the map that the image retrieval algorithm could not locate. However, this task is not trivial and unlike comparing just two images to detect a false positive, it requires to go through all the images in the map. Hence, we choose the best possible image retrieval algorithm and think it as a black-box that behaves like an independent external source introducing this noise in terms of false negative image matches. Due to the fact that this algorithm affects both of the maps being evaluated, no bias is introduced into quality evaluation.

From each test run $n = 10$ images are chosen randomly. Each image is fed as the query image into the image retrieval algorithm and localization in the map is declared if the algorithm can match that image within the map and the user visually verifies it. For each successful localization, coverage ratio is increased and the accuracy of the match is calculated based on equation 2. Some examples of successful localizations are shown in figure 3. This procedure is repeated $m = 10$ times and the average scores are returned as the result of the evaluation of appearance maps in terms of localization task which are presented in table 1. As expected, m_1 with 148 vertices incurred in lower accuracy error when compared to m_2 with 99 vertices due to the higher probability of containing a similar image to a random image. On the other hand, in terms of localization coverage, m_1 scored almost the same in the first two test runs whereas m_2 scored almost twice as m_1 in the third test run. As it can be seen in Fig. 2, the two paths followed to collect images to build maps are almost identical except the small loop in the second map. The images in the third run were collected while the robot was traversing in the middle of the lab fol-
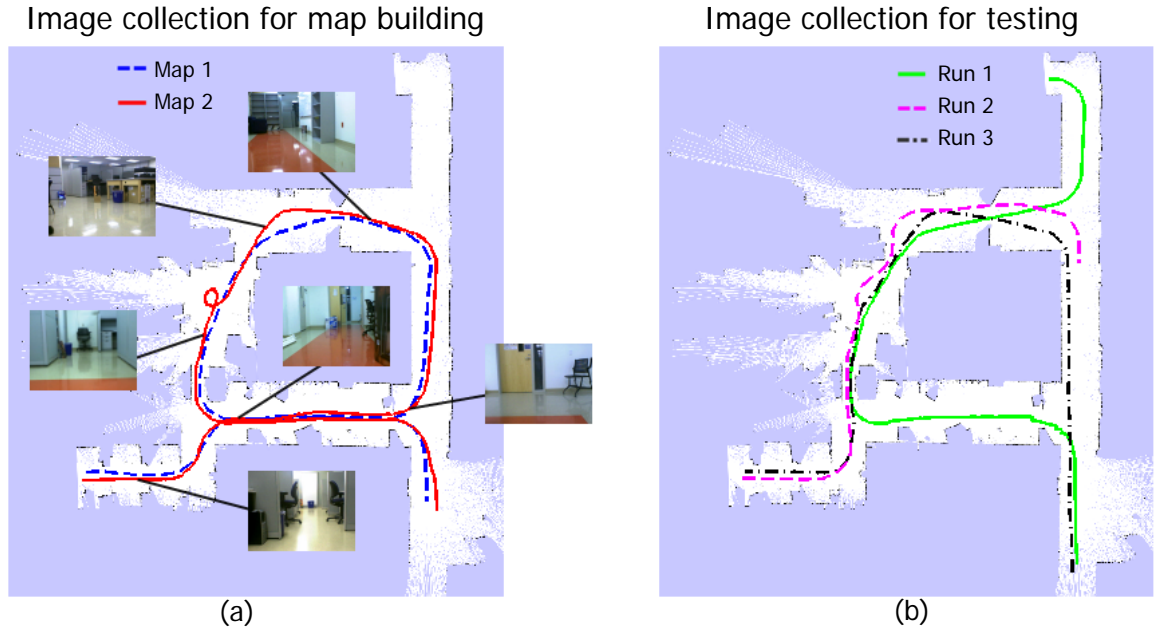
Figure 2: a) The paths followed by the robot during the map building processes are shown. b) Three paths followed to collect test images are shown.

lowing the opposite direction of the map generation runs. Therefore, as expected, images pointing the opposite direction did not match to any of the images in m_1, whereas they matched the images captured while the robot was making a full turn.

Table 1: Localization quality

	Run1		Run2		Run3	
	cov	err_{acc}	cov	err_{acc}	cov	err_{acc}
m_1	0.52	27.02	0.78	34.48	0.22	21.74
m_2	0.50	43.47	0.77	33.32	0.37	37.03

3.2   Planning

As presented above, the first map benefits from having more vertices and obtained a higher score in localization accuracy. Theoretically, a map with infinite number of images should have the utmost accuracy. On the other hand, this map will gain high scores from one side, but will lose performance in planning tasks due to high number of vertices and associated features. Table 2 summarizes the elements of the maps and the amount of data required to store them.

Table 2: Amount of data captured within maps

	$\|V\|$	$\|E\|$	$\|D\|$	dataSize(MB)
m_1	148	1396	24487	2.43
m_2	99	408	15824	1.56

In order to measure the exactness and completeness of the set of paths which can be generated within a map, $n = 10$ image pairs are randomly selected from the map. Then, Dijkstra's algorithm is utilized to find a path connecting these two images where the number of matching features are used as costs. Repeating this procedure $m = 10$ times,

the total number of false negatives are counted. The ratio of the number of false negatives to the total number of path computing requests shows the connectivity of the map. Instead of counting the number of disconnected components in the map, this score considers the size of each disconnected component. For instance, failing to create an edge between second and third vertex has much less effect than failing it between two vertices from two large disconnected graphs.

The ratio of false negatives are measured for two maps being evaluated. The second map m_2 is fully connected and therefore gets completeness score of 1. On the other hand, m_1 has multiple connected components and therefore gets non-zero false negatives. In order to demonstrate how this metric behaves in the presence of only few edges, the parameter TS is used to set the minimum number of feature matches required to declare two images as a match. In different tests the value of TS varies and the performance is evaluated as a function of TS. The increase in the threshold results in a map with edges only between very similar images. Due to this decrease in connectivity the false negative ratio increases as shown in Fig. 4. The sudden jump between $TS = 30$ and $TS = 40$ is the result of losing connectivity in the middle of the path causing the map to split into two components with large number of vertices. Therefore, the probability of randomly selecting both vertices from the same connected subgraph decreases drastically.

The measure of exactness, on the other hand, captures the idea of detecting false positives, i.e. paths that do not exist in the real environment. In other words, the map should contain images captured in disconnected parts of the environment. However, a map generated from a single run will not have such images since the robot will only be able to travel between physically connected places. Hence, it will get an exactness score of 1.

The value of this metric will become apparent only when

Figure 3: Sample image matches are shown. The left column contains query images whereas the retrieved images are shown in the right column. Localization accuracy errors computed for these matches are: 19.13, 50.26, 72.01 (top-down)
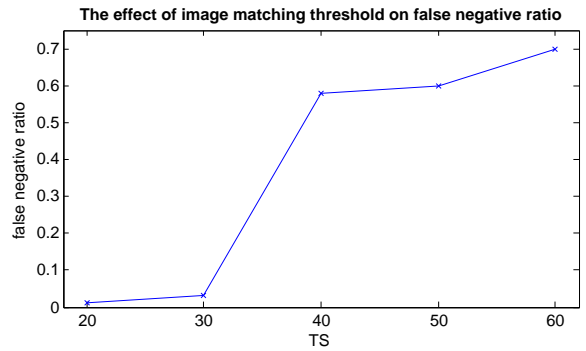

Figure 4: The effect of the number of feature correspondences to declare an image match, TS, on the number of false negatives of path validity test is shown.
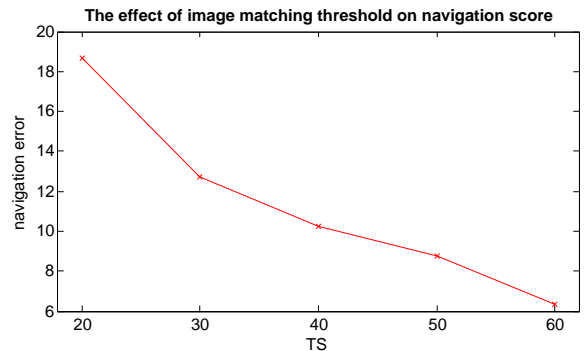

Figure 5: The effect of the number of feature correspondences to declare an image match, TS, on the average navigation error of randomly selected paths is shown.

evaluating maps that are the results of a multi-robot map merging process. In a scenario where robots create local maps of different floors of a building, they can exchange information wirelessly without the necessity to share the same physical environment. The merged map, on the other hand, may contain links between similar images, even though they might belong to physically disconnected parts of the environment. This metric, measuring the ratio of such paths, i.e. false positives, will evaluate the exactness of merged maps. Hence, we leave the validation of this metric for future work, since visual map merging algorithms are not currently available.

3.3 Navigation

The navigability score associated with an appearance map is defined as the average quality of the fundamental matrix between consecutive images of a random image sequence. By design this metric favors maps which use a more conservative approach by only creating edges between very similar images. In order to show this effect, navigability metric is applied on the maps built in the previous section by changing the threshold responsible for determining when an edge is created. Fig. 5 presents the trend of this error as the edge selection procedure gets more selective. As expected, the more similarity required to create an edge, the lower navigation error is obtained for that map.

4. DISCUSSION AND CONCLUSIONS

The proposed metrics measure the performance of appearance based maps with respect to different tasks. These met-

rics focusing on different aspects of the map may not agree on one final map being the best for all tasks, but will measure the map's specific task-based performance. However, these objective measures should be repeatable so that other researchers can reproduce the same experiments to compare published results with their own algorithms. Since the real environment is used as ground truth and providing access to it is not feasible, the data that is used in the computation of the performance metrics should be shared in an online public repository. First of all, images that are captured during the navigation of the robot are needed in order to generate a map of the environment. Different map creation algorithms can be tested on this image set and then evaluated by the task based measures proposed in this paper. In addition to this image set, the computation of the localization metric needs images captured at random locations in the environment. The planning metric, on the other hand, requires the information whether there is actually a physical path in the real environment between the locations encoded in the randomly selected images from the map. This information can be captured in a binary *connectivity matrix* which stores all possible image combinations. Each element in this matrix stores true if two images corresponding to this element are physically connected, and false otherwise. Similarly, in order to count false negatives to calculate the planning metric, we

need to know whether random images from the environment are physically connected. In summary, in order to make visual map generation and benchmarking possible even without the privilege of accessing the real environment, the following elements should be made available to third parties: 1) the image set collected by the robot along its path; 2) random test images captured in the same environment; 3) their connectivity matrices. Only then the experiments can be repeated and the performance of map generation algorithms can be compared.

5. REFERENCES

[1] 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, October 2008.

[2] B. Balaguer, S. Balakirsky, S. Carpin, and A. Visser. Evaluating maps produced by urban search and rescue robots: Lessons learned from robocup. *Autonomous Robots*, 27(4):449–464, 2009.

[3] B. Boufama and R. Mohr. Epipole and fundamental matrix estimation using the virtual parallax property. In *Proceedings of International Conference on Computer Vision*, 1995.

[4] G. Chesi, G. L. Mariottini, D. Prattichizzo, and A. Vicino. Epipole-based visual servoing for mobile robots. *Advanced Robotics*, 20(2):225–280, 2006.

[5] T. Collins, J. Collins, and C. Ryan. Occupancy grid mapping: An emprical evaluation. In *Proceedings of the Mediterranean Conference on Control and Automation*, 2007.

[6] A. Danesi, D. Fontanelli, and A. Bicchi. Visual servoing on image maps. In *Proceedings of IEEE International Symposium on Experimental Robotics (ISER06)*, 2006.

[7] G. Erinc and S. Carpin. Image-based mapping and navigation with heterogeneous robots. In *Proceedings of the 2009 IEEE International Conference on Intelligent Robots and Systems*, pages 5807–5814, 2009.

[8] D. Filliat. A visual bag of words method for interactive qualitative localization and mapping. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2007.

[9] F. Fraundorfer, C. Engels, and D. Nister. Topological mapping, localization and navigation using image collections. In *Proceedings of International Conference on Intelligent Robots and Systems (IROS)*, 2007.

[10] Y. Kalantidis, G. Tolias, E. Spyrou, P. Mylonas, and Y. Avrithis. Visual image retrieval and localization. In *International Workshop on Content-Based Multimedia Indexing (CBMI)*, 2009.

[11] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60:91–110, 2004.

[12] L. Maohai, H. Bingrong, and L. Ronghua. Novel method for monocular vision based mobile robot localization. In *Proceedings of International Conference on Computational Intelligence and Security*, 2006.

[13] G. L. Mariottini, G. Oriolo, and D. Prattichizzo. Image-based visual servoing for nonholonomic mobile robots using epipolar geometry. *IEEE Transactions on Robotics*, 23(1):87–100, 2 2007.

[14] E. Motard, B. Raducanu, V. Cadenat, and J. Vitria. Incremental on-line topological map learning for a visual homing application. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2007.

[15] S. Segvic, A. Remazeilles, A. Diosi, and F. Chaumette. A mapping and localization framework for scalable appearance-based navigation. *Computer Vision and Image Understanding*, 2008.

[16] A. Shademan and F. Janabi-Sharifi. Using scale-invariant feature points in visual servoing. In *Proceedings of International Society for Optical Engineering*, 2004.

[17] J. Sivic and A. Zisserman. Video google: A text retrieval approach to object matching in videos. In *Proceedings of the International Conference on Computer Vision*, 2003.

[18] S. Thrun. Robotic mapping: A survey. In G. Lakemeyer and B. Nebel, editors, *Exploring Artificial Intelligence in the New Millenium*. Morgan Kaufmann, 2002.

[19] Z. Wu, Q. Ke, J. Sun, and H.-Y. Shum. A multi-sample, multi-tree approach to bag-of-words image representation for image retrieval. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2009.

[20] Z. Zivkovic, O. Booij, and B. Krose. From images to rooms. *Robotics and Autonomous Systems (RAS)*, 55:411–418, 5 2007.