Sampling in image space for vision based SLAM

O. Booij, Z. Zivkovic, B. Kröse
Intelligent Systems Lab Amsterdam, University of Amsterdam
1098 SM Amsterdam, The Netherlands
Email: obooij@science.uva.nl

***Abstract*— Loop closing in vision based SLAM applications is a difficult task. Comparing new image data with all previous image data acquired for the map is practically impossible because of the high computational costs. This problem is part of the bigger problem to acquire local geometric constraints from sensor data for geometric map building termed data association. Commonly the computational costs are kept small by sampling the image data uniformly over time or using a position estimate from a mapping and localization algorithm. In this paper we propose a more natural sampling approach, by determining a subset that best describes the complete image data in the space of all previously seen images. The actual problem of finding such a subset is called the Connected Dominating Set problem which is well studied in field of graph theory. The proposed method is particularly beneficial for realistic mapping scenarios including moving objects and persons, motion blur and changing light conditions. Evaluation on multiple large indoor datasets show that the method performance is very close to that of an exhaustive data association scheme and outperforms other sampling approaches.**

I. Introduction

In the field of SLAM many effort has been devoted to efficiently and consistently computing a global metric map from local geometric constraints [1]. However, most of the theoretical results are based on the assumption that the data association problem to compute these local constraints is solved.

Perfect data association involves finding for each new sensor reading, all the previous sensor readings that were added to the map that correspond to it. However, the effort necessary to find all these corresponding measurements grows linearly while the map is growing, making perfect data association practically impossible for realistic mapping scenarios. And even with a very robust matching technique, ambiguities will always exist due to similar sensor readings in different parts of the environment. Solving these ambiguities is addressed by Rao Blackwellised Particle Filters [2] and the MCMC based approaches that search in the space of topological maps [3]. In this paper we focus on the first problem: efficiently finding for each new sensor reading the corresponding previous sensor readings in a growing map.

In general it is not possible to compare new measurements with each previous measurement of the map because it would cost to much computational time. Instead a choice should be made which measurement of the map to consider for matching. This can be seen as a sampling problem in which we have to pick samples from the map given some distribution. Commonly some simple heuristic is used, for example by looking at which time previous measurements were added to the map and sampling uniformly over them in the time domain. In this paper we argue that by using the knowledge of previously matching measurements, we can sample a set of key measurements that best covers the map in the space of the measurements themselves. This sampling is performed by regarding the previously matching measurements pairs as a graph and use a technique from graph theory termed the Connected Dominating Set (CDS)[4], to find the a minimal set of key measurements that still represents all measurements in the map[5].

In this paper we focus on vision based SLAM, because image matching is known to be computationally expensive. However, the proposed CDS method can just as well be applied to SLAM methods based on other sensors such as laser range scanners.

In particular we focus on view based SLAM [1], in which the so called map consists of a trajectory of robot poses with their corresponding images, or views. In view based SLAM data association involves matching each new image with previously shot images, as opposed to landmark based SLAM, in which features extracted from the new image are matched with 3D landmarks in the map. The vision sensor that is used in the experiments is a single viewpoint catadioptric vision sensor which can capture images with a very wide field of view. However, all methods discussed in this paper could just as well be applied to conventional cameras.

The rest of the paper is organized as follows. First, in Section II, related work is discussed, focussing on how vision based mapping methods choose which parts of the map to match with. Then, in Section III, we propose a new data association approach based on the CDS. In Section IV we briefly explain the lower level image matching technique used in the experiments. In Section V the proposed method is evaluated on multiple challenging datasets, mostly acquired in real home environments.

II. Sampling methods, related work

Sampling always has to be done to bound the computational time spend on data-association. For vision based mapping applications there is an implicit bound given by the frame rate of the camera. Although standard cameras have a frame rate of 30 frames per second. for indoor environments with low brightness this is usually lower due to the higher shutter times needed to capture enough light. For mapping very small environments this reduction in the number of images could be enough to perform full data association in real time. For larger environments it is not uncommon to drop most frames and for example only pick one per second [6]. This is troublesome in applications where robots stand still for long times, as usually is the case for robots interacting with humans such as museum robots.

Another straight-forward sampling method specifically used in view based approaches is to sample over the area that is being mapped [7][8]. This is commonly done by using odometry measurements. This method fails in non static environments, where, for example, light changes occur or objects or humans move, while the robot is standing still. Also, in small places such as corridors and door openings, the appearance changes relatively faster given the movement of the robot than in big convex spaces such as a big room.

A somewhat more sophisticated method is to pick sensor readings given some quality measure. For example, to keep the map small and reducing the risk of adding non-stationary image features, such as reflections and shadows, landmark based approaches usually add just a few landmarks with high distinctiveness per frame to the map [9]. As a result most computational cost of corresponding new feature result from corresponding them with parts of the environment with a lot of distinctive features, while only a little from corresponding to parts of the environment with a relatively low number of distinctive features. This is a quite undesirable effect. Detecting loop closure in texture rich parts of the environment is easy, while loop closing in texture poor parts needs extra effort. In this paper we argue that feature poor parts should be represented with relatively more features instead of less.

In SLAM applications new robot positions in the map can be predicted given the motion model. It is very common to use this so called navigation prior to define parts of the map for data association [10, 9]. There are a few fundamental drawbacks with this approach. Due to linearization errors SLAM methods are usually overconfident. Because of this crucial loop closing observations could be missed, because they are deemed close to impossible by the current map. In general the resulting dependence of the landmark observation on the map is neglected, leading to an even more overconfident state estimate.

There have also been approaches that sample in image space. A common scheme to find a set of key images is to first cluster the complete set of images based on an image similarity measure and then for each cluster choose one representative key image [11, 12]. However, solving a clustering problem is actually a more difficult problem than the problem at hand. In [13] finding clusters is simplified by using the temporal distance with which the images were taken The drawback of grouping images on the basis of temporal distance is that all images taken from the same place but at a different time are put in a different group, producing probably quite similar prototype images. Indeed one of the assumptions in [13] is that each location is visited only once.

Sampling images based on an image similarity measure has however some very nice advantages. Parts of the environment where images have a low similarity will directly be represented by more key images. This is for example the case places with bad or changing lighting conditions, but also in places where their are a lot of moving objects or humans. Images taken by the robot while standing still will be highly similar as opposed to images taken while driving fast. In this paper we describe an algorithm that uses such a sampling method. As opposed to known methods that sample in image space, the proposed method does not use clustering but determines a set of key images directly.

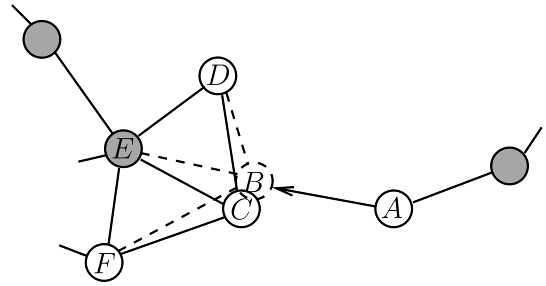


Fig. 1. Figure explaining the rationale of using the Connected Dominating Set method. The circles denote robot poses and the links connecting them indicate that the images taken at the robot poses match. Grey circles indicate CDS nodes. The robot moves from robot pose A to a new robot pose B, which has to be associated with the map.

III. Sampling based on image similarity

The map of the View based SLAM approach consists of the complete set of past robot poses and their corresponding images. In this section we propose a method to efficiently perform data association given such a map and a new image without performing an exhaustive search. We first explain how a set of key images is determined using the Connected Dominating Set. Then these key images are used to define a practical and efficient data association scheme. We assume that a similarity measure is defined that can takes two images and computes if they are similar or not. Later in Section IV we briefly describe the similarity measure used in the experiments

A. The Connected Dominating Set

We assume that we already mapped part of the environment and found pairs of robot poses for which the corresponding images matched. The problem now is to compute a minimal set of key images that best describes the complete image set, given the set of matching image pairs.

See Figure 1 for an example scenario. Suppose the robot moves to a position B close to a previous robot position C. If the world is more or less static, then a newly captured image at B looks a lot like the image taken at C. Thus the new image taken at B probably also matches all the images that matched the image taken at C. To localize oneself in the map it suffices to compare the new image taken at B with only one of these matching images taken at D, E or F. Thus only one of these images has to be marked as key image and the rest can be ignored.

Of course this holds for all possible previous robot positions. It would suffice to compare the new image with a subset of key images which has the property that every image is either a key image or matched a key image. This is exactly the definition of the Connected Dominating Set (CDS), a concept originating from graph theory, which is commonly used for broadcasting in large networks[4].

The set of image pairs in the map can be seen as a graph $G = (V, S)$, in which a node $v \in V$ represents an image and a link $(u, v) \in S$ between represents that the two images that correspond to node u and v match. A Connected Dominating Set V' is defined as follows. The set of nodes in the Dominating Graph V' is a proper subset of the original set V, such that every node u in the original set V is either in the Dominating Set V' or is neighboring a node in V':

$$\forall u \in V : u \in V' \vee \exists v \in V' : (u, v) \in S \tag{1}$$

The problem now is to find a CDS with the minimal number of nodes so to compare as few images as possible. This task is however known to be NP-complete. Fortunately algorithms exist that can find a good approximation in the order of the number of nodes [4]. Most of these algorithms first remove links to make a spanning tree with as many leaves as possible and then define the set of all non-leaves as the CDS.

We implemented an algorithm as proposed by Guha [4] that iteratively finds CDS nodes in a connected graph (see [4] for a complete description of the algorithm). The computation time of the algorithm is negligibly small as compared to the time needed for the actual image matching. For the datasets used for evaluation the computation time was always smaller than 1 ms.

B. Data association scheme

For each new image that is taken by the robot a new CDS is determined. Comparing the newly taken image with the images in the CDS, results in some extra image pairs, but more importantly it indicates where to look for more matching images. To determine as much matching image pairs as possible, the new image is compared with all the images that match matching CDS images. Thus in the example of Figure 1 if the new measurement B matches CDS node E, then B is also matched with C, D and F.

In case the robot always revisits previous locations then the CDS method returns the approximately optimal subset for localization. If the robot, however, drives through a corridor it could happen that it can not match any of the images in the CDS. Therefore, we do not only compare with matching images of matching CDS images but also those of all CDS images that matched the previously taken image.

IV. Comparing images

This section briefly describes the matching technique used to compare two images taken by the omnidirectional vision system used in the experiment section. The method is based on matching local image features and imposing the epipolar constraint [14].

Images taken by the omnidirectional vision sensor are first mapped to panoramic images to be able to apply conventional computer vision techniques [15].Salient image points, or feature points, are found in the images by applying the Scale Invariant Feature Transform (SIFT)[16]. These salient points are described by a standard SIFT descriptor of 128 dimensions. Salient points that have a small Euclidean distance in descriptor space to other salient points in the same image are removed.

A set of point correspondences between the two images is determined by computing the two nearest neighbors of every salient point. Two points match if the ratio between these two neighbors is larger than 0.8. This is the standard matching scheme as described in [16]. The number of point correspondences could be used to determine if the two images depict the same part of the environment. However, the set will also include mismatched image points pointing to different 3D landmarks.

The point correspondences that are the projections of the same 3D point in the environment are constrained by the epipolar geometry[17]. This epipolar geometry is formally described by the Essential matrix E that relates the projections of landmarks as 3D points l_i and r_i on the camera surfaces:

$$\mathbf{l}_i^T E \mathbf{r}_i = 0 \qquad \text{for all } i, \tag{2}$$

For omnidirectional vision l_i and r_i are usually obtained by normalizing the 3D light rays, corresponding to the pixel coordinates, to unit length, effectively projecting them on a sphere[15, 18].

The 3x3 matrix E is estimated using a variant of the 8-point algorithm[19], for which the constraint is added that the camera moves over a planar surface[20]. This algorithm is used inside the RANSAC robust estimator, which estimates the epipolar geometry and at the same time determines the number of fitting point correspondences[19, 21]. A point correspondence fits E if the Sampson distance is below a certain threshold which can be estimated from the data (we used 0.01) [19, 21] and the corresponding 3D world point has a positive depth in both cameras[22].

The number of remaining mismatches, fitting E, is proportional to the total number of features found in the two images. If the number of fitting point correspondences normalized by the lowest number of features found in the two images is larger than a threshold set to 0.10, then the images match.

V. Experiments and results

In order to evaluate the proposed CDS data association method we applied it on several datasets, one ”office set” acquired in our university building and three ”home sets” taken in real home environments [1]. For all these datasets, including the one taken at the university building, the conditions were far from ideal ”lab conditions”.

[1]All the used datasets, including images, odometry, sonar and laser range data (all timestamped), are available from `http://www2.science.uva.nl/sites/cogniron/`.

Fig. 2. Example image from the office data sets.

The office dataset shows how the proposed method copes with a robot traversing the same loop in the building twice. The application of the method on the home environment datasets evaluate the robustness of the method. Although home environments typically have more visual texture resulting in more image features, the dynamic light conditions make image matching hard.

In addition the CDS sampling method is compared with other sampling methods based on time, displacement of the robot and a randomly picking a subset.

Except for the third home set all datasets were acquired using a tele-operated Nomad Scout mobile robot platform, which was equipped with an omnidirectional vision system, consisting of an Accowle convex hyperbolic mirror and a one-megapixel Firewire video camera. The third home set was acquired using Biron (the Bielefeld Robot Companion) equipped with the same omnidirectional vision system. The computer vision and data association algorithms were implemented in C++ and were running on a 2Ghz laptop mounted on the robot.

A. *Office dataset*

The office dataset was kept relatively small so we are still able to apply an exhaustive full data association corresponding each new images with all previously seen images in the map for evaluation purposes. The environment consisted of a small loop through two rooms and a small part of a hallway. The robot was driven over the loop two times, see Figure 3(a), taking in total 877 images. See Figure 2 for and example image taken by the omnidirectional vision system.

Figures 3 visualizes the results of the proposed data association step in two ways: a connectivity graph, linking the matching images and using the odometry information for the position of the image-nodes, and a connectivity matrix, which clearly shows the loop closing observations by the off-diagonal non-zero values.

In total 74,585 image pairs were matched in 303 seconds resulting in 31,199 links. See Table I for an overview of the datasets and their computational time usage. In Figure 4 a more detailed plot of the computational time usage of both the CDS method and the full data association is shown The exhaustive data association scheme resulted in 32,583 links, only 4% more than the efficient CDS method while comparing 5 times more image pairs. See Table II for an overview of the exhaustive data association scheme compared to the proposed method.

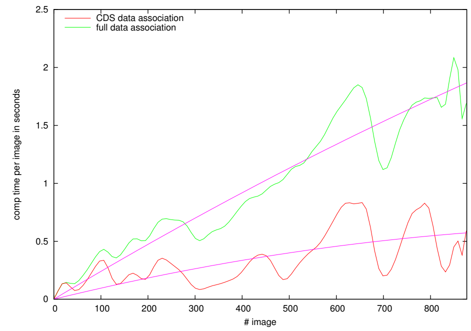


Fig. 4. Comparison of the computational times for each new image of the CDS data association and the conventional brute force method while the map is growing. The large fluctuations in both graphs are caused by the difference in number of image features in the new image. If it has more features, matching to other images takes more time. On both graphs a second order polynomial is fitted indicated by the dashed lines.

TABLE I

COMPUTATIONAL TIME USAGE FOR THE CDS METHOD.

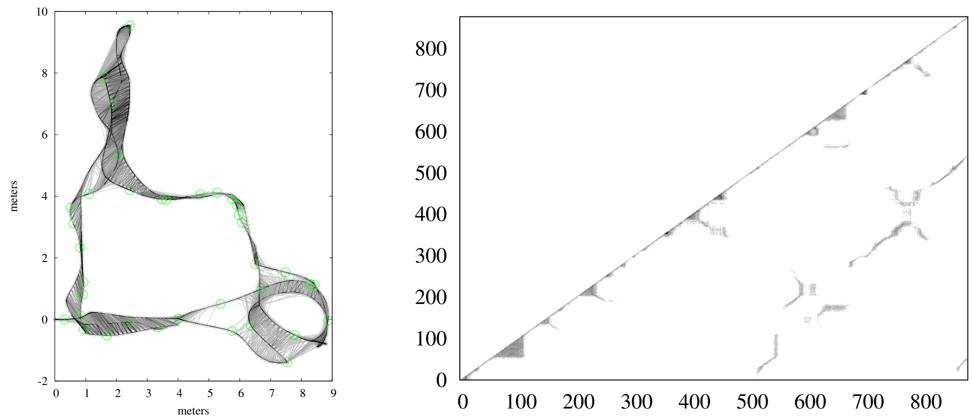
Dataset	**N**	**matched**	**links**	**false**	**total time**	**average(std)**
Office	877	74,585	31,199	0%	303 s	.34 s (.25)
Home 1	1153	93,789	42,529	0%	850 s	.74 s (.52)
Home 2	1845	192,876	72,240	0%	1,359 s	.73 s (.53)
Home 3	1734	267,465	105,843	3%	934 s	.53 s (.45)

While mapping the environment, more and more images acquired at different positions are added to the dataset and thus the size of the CDS grows. This is depicted in Figure 5. At image 507 the robot finished its first loop in the environment and had a CDS size of 30 images. During the second traversal of the loop new images were matched with images taken in the previous loop, creating links between these images as shown in Figure 3(a). Because of these links only a few extra nodes were added to the CDS during this second loop resulting in a total of 37 nodes in the final CDS (which are indicated in Figure 3(a). This is exactly what should be expected if sampling in image space. Each key image represents images taken in both traversals of the loop.

TABLE II

COMPARISON OF EXHAUSTIVE AND CDS DATA ASSOCIATION.

Dataset	**Full**		**CDS**		**CDS/Full**	
	matched	**links**	**matched**	**links**	**matched**	**links**
Office	384,126	32,583	74,585	31,199	19%	96%
Home 1	664,139	43,091	93,789	42,529	14%	99%
Home 2	1,701,108	74,037	192,876	72,240	11%	98%
Home 3	1,502,511	113,555	277,528	106,111	18%	93%



(a) Connectivity Graph

(b) Connectivity Matrix

Fig. 3. Results of applying the CDS data association method on the small dataset. (a) shows the graph, in which the nodes corresponding to the images are denoted by small circles, and the links by lines connecting the circles. The nodes are positioned using the raw odometry readings. Bigger green circles denote the nodes that were in the final CDS. (b) shows the connectivity matrix of the graph, with a darker color for image pairs with a higher similarity. The of-diagonal non-zero entries denote places where loop closure occurred. Note that at image 507 the robot started driving over the loop a second time.

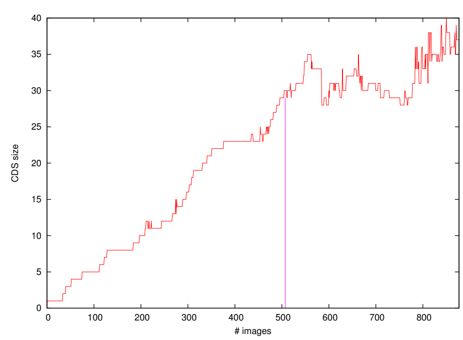


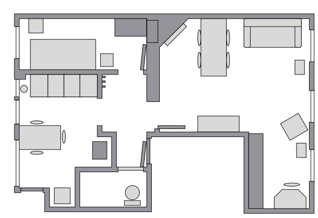
Fig. 5. The number of nodes in the CDS, while the map is growing. The vertical dashed line indicates the beginning of the second traversal of the loop.

Also interesting is the distribution of key images (see Figure 3(a)). Some areas of the environment such as the upper part are represented by only 4 key images, although the trajectory was driven 4 times, while other, much smaller areas, such as at the middle right, are represented by 6 key images. This is a direct result lighting differences between those areas, the second area being much darker than the first area.

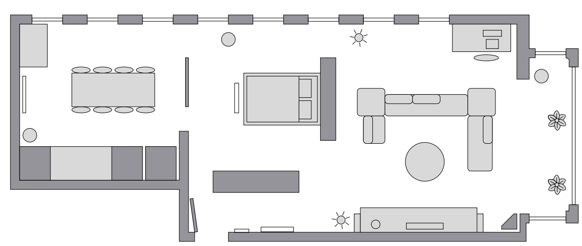
B. Real home datasets

The CDS method was also tested on three challenging datasets acquired in real home environments. To get an idea of the structure of these homes see Figure 6. Some typical images captured by the omnidirectional vision system are shown in Figure 7. For a detailed description of the acquisition of these datasets see [23].

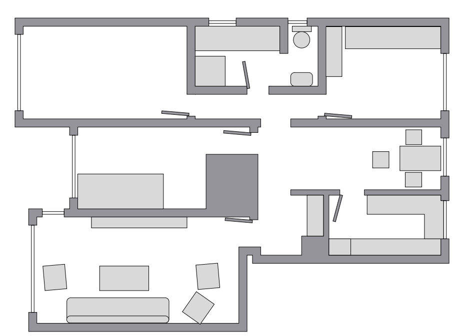
The first home dataset, consisting of 1153 images, is taken in a relatively feature rich home environment. However, the images were shot in the evening, resulting in somewhat dark images and more importantly some motion blur, because of the



(a) Home 1, a student apartment



(b) Home 2, the Unet demo house



(c) Home 3, demo house from the Bielefeld University

Fig. 6. Ground floor maps of the home environments. The position of the furniture is approximate.

higher shutter times needed to capture bright enough images. The second home dataset, which consisted of 1845 images, was taken during day time in a less feature rich home. While the robot was driving around people walked in close vicinity to the robot. Home 3, consisting of 1734 images, is the most difficult. The environment has a relatively low amount of visual texture and the dataset was shot in the evening. Also the robot was heavily loaded, resulting in some additional motion blur. See Figures 7 for some example images of all three datasets.

In Table I and Figures 8 the results of the data association method are given in a similar fashion as for the office set. For all three datasets the computational time of the complete data association was fairly constant during map building at .7 seconds. For Home 3 it was somewhat less due to the lack of features, resulting from the bad lighting conditions. This was also the cause for the relatively high number of false positive image matches which can clearly be seen in Figure 8(e). By visual inspection we found that 3% of the matched images was actually false for Home 3. The other datasets did result in any false match.

As with the office set the results are compared with an exhaustive data association scheme, matching all images with each other. Again the same pattern is visible: the efficient CDS method finds almost all matching images, while doing 5 to 10 times less comparisons, see Table II. The percentage of false negatives is around 4%.

It is interesting to investigate the distribution of the key images over the complete set of images and the position from which they were taken. We highlight some of the specific characteristics of sampling in the space of images by looking into some parts of the home environments.

In Home 1 we see that relatively more key images are located in the middle to left part of the house than at the right side (see Figure 8(a). One of the reasons is that in the left part there are fewer objects and less visual texture. But another important reason is that on the left side the robot made relatively more turns resulting in more motion blur in the images. Compare Figures 7(a) and 7(b) which were taken in the left respectively right side of the house.

The dataset taken in Home 2 was shot during day time and most light was coming in through the windows at the top of Figure 8(c) (see also Figure 6(b)). However, some areas, such as the small entrance hallway at the bottom center were located far away from these windows and were as a result quite dark. Compare the images in Figures 7(c) and 7(d) taken at a dark and a bright area. The implication for the sampling in image space is that relatively few key images are necessary to represent the top left as compared to the bottom center.

In both previous examples it was also the case that the space in which the feature poor images were taken was smaller than the space in which feature rich images were taken. Indeed when landmarks seen in the images are closer to the moving robot the view angle changes more quickly and it is more difficult to match these images. In Home 3 we see a more isolated example of this phenomena. Compare Figures 7(e)

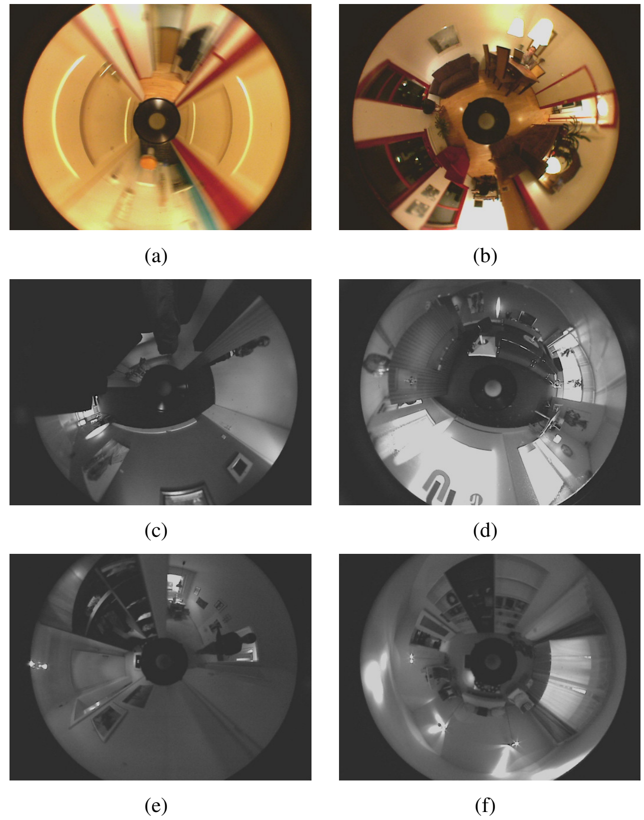


Fig. 7. (c) (d) (a) (b) (e) (f)

and 7(f) which were shot in a small hallway and the big living room respectively. Both spaces are quite feature rich and have comparable lighting. However, 6 key images represent the small hallway while only 3 represent the big living room.

C. Comparison of sampling techniques

The investigation of the distribution of the key image over the tested datasets suggests that sampling in the space of images results in a subset of key images that better represents the complete set, and is thus more suitable for data association tasks. In the following we compare the proposed method with other methods to pick key images of the Home 1 dataset. To make the comparison as fair as possible we set the sampling density for each method such that the number of images-pairs that is evaluated is more or equal to the number of image-pairs evaluated by the CDS method. Thus the CDS method will use less or equal the amount of computational time. Other than the sampling technique the data association method is completely the same.

As a baseline the first method just picks images randomly from the seen images. During each iteration a new set is chosen with an average number of images equal to .07 times the number of images in the map.

The second method uses the odometry measurements to sample over displacements of the robot. After each 54 cm an images is added to the set of key images.

TABLE III

COMPARISON OF DIFFERENT SAMPLING METHODS.

method	key images	links	false neg	links/matched
Full	-	43,091	-	6%
Random	83	32,509	25%	33%
Position	64	37,382	13%	39%
Time	61	38,623	10%	39%
CDS	40	42,529	1%	45%

The third approach samples images over time. After each 5 seconds a mapped image is added to the set of key images.

In Table III the CDS method is compared with these methods and the method of exhaustive data association. As can be seen the proposed CDS method outperforms all these sampling techniques. The set of key images is smallest for the proposed method. More importantly it finds by far the highest number of links, close to the number found by exhaustively searching. As a result the percentage of successful image comparisons (links/match) is highest of all methods, including of course that of a full matching scheme. Although we did investigate the number of mismatches that were made, except for the CDS method, the percentage of successful comparisons does indicate that the proposed method is more robust against false positives.

VI. CONCLUSION

In this paper we proposed an efficient data association approach for view based SLAM that determines a set of representative images by sampling in image space using an image similarity measure. We have shown that the problem of finding the minimal number of key images given a set of previously matched images is equivalent to finding a minimal Connected Dominating Set. The method is applied on four challenging datasets mostly acquired in real home environments. It outperforms known sampling techniques by finding in the same amount of computational time much more matching image pairs, on average 96% of the matches found by an exhaustive search.

Although in the experiments the CDS method was used stand alone, it could just as well be merged with other sampling techniques, for example using the navigation prior of a SLAM method. Also the efficiency of the CDS method could be even further improved by additionally using a more efficient image similarity method like the recently proposed hierarchical methods[24] or the bag-of-words methods that use training sets to learn how discriminative image features are.

In the experiments we used a similarity measure that is based on the epipolar constraint, so it is clear to see how this data association method can be used in a complete view based SLAM system [25]. In [26] an experiment is shown in which the CDS method is used to build a map with more than 10,000 images.

REFERENCES

[1] Hugh Durrant-Whyte and Tim Bailey. Simultaneous localisation and mapping (slam): Part I The essential algorithms. *Robotics and Automation Magazine*, June 2006.

[2] M. Montemerlo, S. Thrun, D. Koller, and B. Wegbreit. FastSLAM 2.0: An improved particle filtering algorithm for simultaneous localization and mapping. In *Eighteenth International Joint Conference on Artificial Intelligence (IJCAI)*, Acapulco, Mexico, August 2003.

[3] A. Ranganathan, E. Menegatti, and F. Dellaert. Bayesian inference in the space of topological maps. *IEEE Transactions on Robotics*, 22(1):92–107, Feb 2006.

[4] S. Guha and S. Khuller. Approximation algorithms for connected dominating sets. *Algorithmica*, 20(4), April 1998.

[5] O. Booij, Z. Zivkovic, and B. Kröse. Sparse appearance based modeling for robot localization. In *IROS*, pages 1510–1515. IEEE, October 2006.

[6] I. Ulrich and I. Nourbakhsh. Appearance-based place recognition for topological localization. In *Proceedings of ICRA 2000*, volume 2, pages 1023 – 1029, April 2000.

[7] R. Sim and G. Dudek. Learning and evaluating visual features for pose estimation. *In Proc. International Conference Computer Vision*, 1999.

[8] M. Jogan and A. Leonardis. Robust localization using an omnidirectional appearance-based subspace model of environment. *Robotics and Autonomous Systems, Elsevier Science*, 45(1):51—72, 2003.

[9] A. J. Davison, I. D. Reid, N. Molton, and O. Stasse. MonoSLAM: Real-time single camera SLAM. *PAMI*, 29(6):1052–1067, 2007.

[10] S. Se, D. G. Lowe, and J. J. Little. Vision-based global localization and mapping for mobile robots. *IEEE Transactions on Robotics*, 21(3):364–375, 2005.

[11] F. Li and J. Kosecka. Probabilistic location recognition using reduced feature set. In *Proceedings of the International Conference on Robotics and Automation*. IEEE, 2006.

[12] C. Valgren, T. Duckett, and A. Lilienthal. Incremental spectral clustering and its application to topological mapping. In *Proceedings of the International Conference on Robotics and Automation*, pages 4283–4288. IEEE, 2007.

[13] J. Kosecká, L. Zhou, P. Barber, and Z. Duric. Qualitative image based localization in indoors environments. In *CVPR (2)*, pages 3–10. IEEE Computer Society, 2003.

[14] B. Kröse, O. Booij, and Z. Zivkovic. A geometrically constrained image similarity measure for visual mapping, localization and navigation. In *ECMR*, pages 168 – 174, Freiburg, Germany, 19-21 September 2007.

[15] R. Bunschoten. *Mapping and Localization from a Panoramic Vision Sensor*. PhD thesis, University of Amsterdam, November 2003.

[16] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. Journal of Computer Vision*, 60(2):91–110, 2004.

[17] H. C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, Sept. 1981.

[18] G.L. Mariottini and D. Prattichizzo. Image-based visual servoing with central catadioptric camera. *International Journal of Robotics Research*, 27:41–57, 2008.

[19] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision, second edition*. Cambridge University Press, 2003.

[20] D. Ortín and J. M. M. Montiel. Indoor robot motion based on monocular images. *Robotica*, 19(3):331–342, 2001.

[21] P. H. S. Torr and D. W. Murray. The development and comparison of robust methods for estimating the fundamental matrix. *Int. Journal of Computer Vision*, 24(3):271–300, 1997.

[22] O. Chum, T. Werner, and J. Matas. Epipolar geometry estimation via ransac benefits from the oriented epipolar constraint. In *Proceedings of the Int. Conf. on Pattern Recognition*, volume 01, pages 112–115, 2004.

[23] Z. Zivkovic, O. Booij, B. Krand E.Topp, and H.I.Christensen. From sensors to human spatial concepts: an annotated dataset. *Transactions on Robotics*, 24(2):501–505, April 2008.

[24] 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*, volume 2, pages 1470–1477. IEEE, October 2003.

[25] R. M. Eustice, H. Singh, and J. J. Leonard. Exactly sparse delayed-state filters for view-based SLAM. *IEEE Transactions on Robotics*, 22(6):1100–1114, December 2006.

[26] I. Esteban, O. Booij, Z. Zivkovic, and B. Kröse. Mapping large environments with an omnivideo camera. In *ASCII*, Heijen, The Netherlands, 2008.

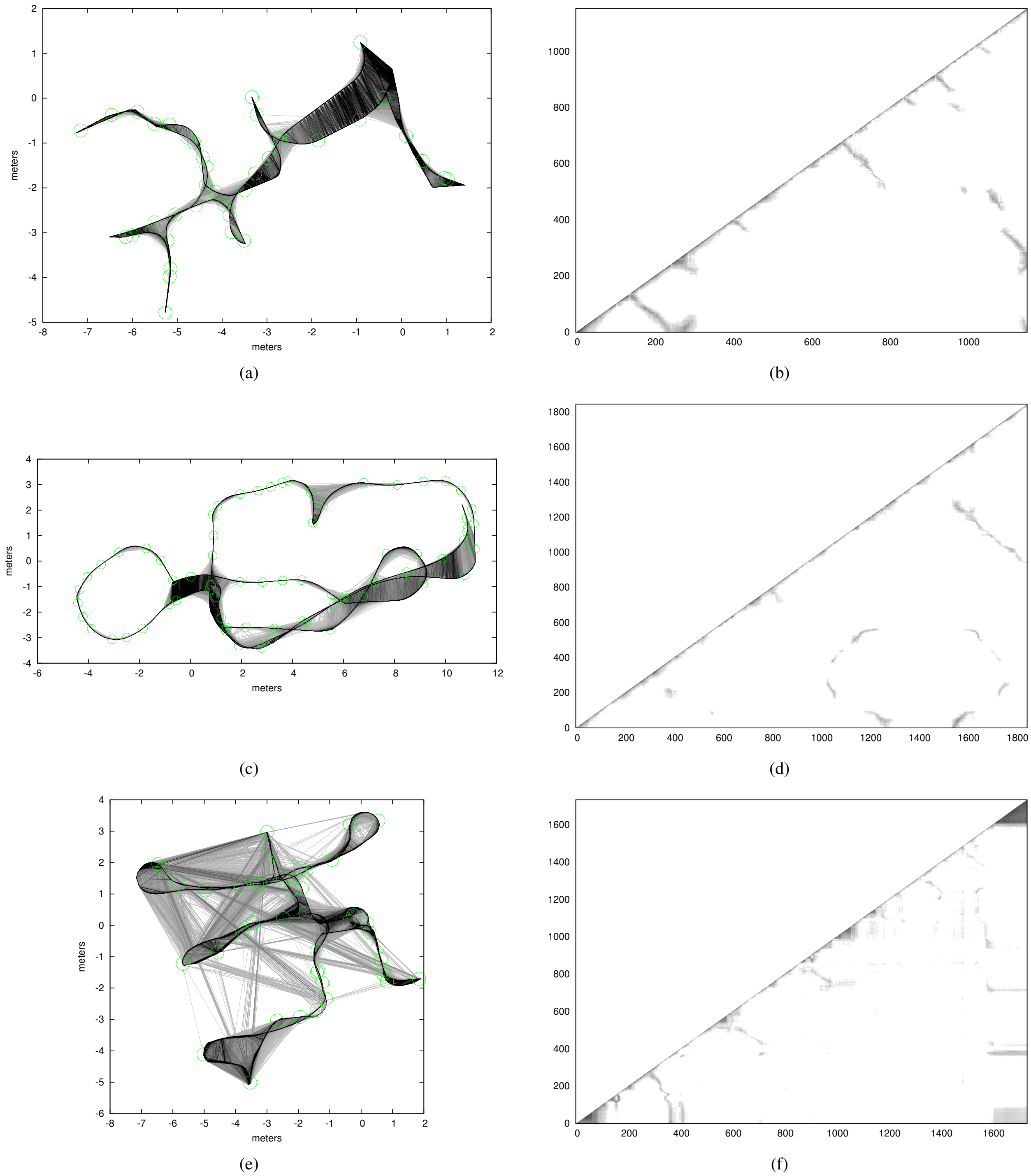


Fig. 8. Results of CDS data association method. (a) and (b) are the resulting graph and the connectivity matrix for the Home 1, (c) and (d) for Home 2 and (e) and (f) for Home 3. The graphs were plot using hand-corrected odometry information for Home 1 an 2 and using the result of Laser-SLAM for Home 3. Green circles denote the images of the final CDS. Lines were drawn between poses to denote that the images corresponding to the two robot poses matched. By using the zoom functions of a PDF reader parts can be magnified to fully respect the number of found image matches, (for Acrobat 8.0 turn off Line Weights, for Acrobat 7.0 turn on Wireframe). The connectivity matrices give another view on the results. The entries on the main diagonals are the result of matching sequential images, while the off-diagonal entries reflect instances of loop-closing.