Efficient data association for view based SLAM using connected dominating sets

O. Booij, Z. Zivkovic, B. Kröse

Intelligent Systems Lab Amsterdam, University of Amsterdam
1098 SM Amsterdam, The Netherlands

Abstract

Loop closing in vision based SLAM applications is a difficult task. Comparing new image data with all previously acquired image data is practically impossible because of the high computational costs. Most approaches therefore compare new data with only a subset of the old data, for example by sampling the data over time or over space by using a position estimate. In this paper we propose a more natural approach, which dynamically determines a subset of images 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" (CDS) problem which is well studied in the field of graph theory. Application on large indoor datasets results in approximately the same map using only 13% of the computational resources with respect to comparing with all previous images. Also, it outperforms other sampling approaches. The proposed method is particularly beneficial for realistic mapping scenarios including moving objects and persons, motion blur and changing light conditions[1].

Key words: robot vision, visual SLAM, data-association, loop-closing

---

1. Introduction

Data association is a fundamental problem in the field of SLAM (Simultaneous Localization And Mapping), where a global metric map is to be built incrementally from sensor data [2]. In most SLAM approaches it means that the most

Email address: `obooij@science.uva.nl` (O. Booij, Z. Zivkovic, B. Kröse)

[1]This is an extended version of the workshop paper [1]. All datasets and software used in the experiments are available online.

probable associations between sensor measurements and elements of the map have to be found. The problem is that time spent on finding these associations increases as the map is growing.

In view based SLAM [2], the so called "map" consists of a trajectory of robot poses with their corresponding images. In this case perfect data association involves finding for each new image, all the previously acquired images that match it. Two images match if information can be extracted that is useful for updating the map, for example if the relative pose can be determined using epipolar geometry estimation [3]. This is especially the case if loop closing is involved, in which mapping errors are decreased because a new image matches with an image taken some time ago. The effort necessary to find all these corresponding images grows linearly while the map is growing, making perfect data association practically impossible for realistic mapping scenarios. In addition, a second problem is that even with a very robust matching technique, ambiguities will be present due to similar images taken in different parts of the environment. Solving these ambiguities is addressed by Rao Blackwellised Particle Filters [4] and the MCMC based approaches that search in the space of topological maps [5]. In this paper we focus on the first problem, which we call the data association problem: efficiently finding for each new image the matching previous images in a growing map.

Data association for view based SLAM can be performed more efficiently by considering only a selection of previously acquired images for matching new images. The question is how to compare the smallest number of images, while still finding the largest number of matches. A natural approach is to use a hierarchical scheme, in which new images are first compared with a subset of key images that best represents the complete set of images in the global map. The results of these comparisons are then used to search more locally for image matches.

However, it is unclear what constitutes a good representative subset of a collection of images. Parts in the environment where images are harder to match, for example because of bad lighting conditions, should be represented with more images, while parts where a lot of images match each other, need less. One could try to solve this by clustering the set of previously acquired images based on the matches that were found among them. Each cluster is then represented with one image. Because the matching function is usually not a metric, a spectral clustering method should be chosen. However, spectral methods are known to be complex and computational intensive. Also the question remains which images should be chosen to represent each cluster.

In [6] we adopted an approach based on a graph representation of the image set. All previously matched images can be used to form a graph, in which the nodes denote images and a link between two nodes denotes that the two images matched. In graph theory, the subgraph that contains the minimal number of nodes that still covers the complete graph is termed the "Connected Dominating Set" (CDS) [7]. In [8] we have shown that efficient robot localization can be performed by determining such a CDS to find a minimal set of images, that represents the complete set of images in an optimal way.

In this paper we use this CDS method in an incremental hierarchical data association scheme. For a new image first the matching key images are determined, which are then used to search more locally for image matches. For each resulting match a link is added to the graph, which can be used to associate the next image. What results is an incremental mapping framework, that only compares pairs of images which have a high chance of matching. While the map is growing the set of key images dynamically changes by applying the CDS method for each added image.

The question is if the proposed incremental data association scheme results in a comparable map as would result from an exhaustive data association scheme. This depends on the ability of the CDS method to indeed find a subset of images that represents the complete image set. Also, in this paper we investigate how the CDS method compares to other techniques of picking key images.

We have put a focus on view based mapping, because image matching is known to be computationally expensive. Nevertheless, the proposed method can just as well be applied to mapping methods based on other sensors such as laser range scanners or even landmark based approaches.

The rest of the paper is organized as follows. First, in Section 2, related work is discussed. Then, in Section 3, we propose the new data association approach based on the CDS. In Section 4 we briefly explain the image matching technique used in the experiments. In Section 5 the proposed method is evaluated on multiple challenging datasets, acquired in real home environments.

2. Related work

In various large mapping applications or image retrieval tasks efficient data association is achieved by using a fast yet simplistic image comparison method, such as in [9] and [10] where each image is described with a single image feature which can be compared very quickly. In [11] and [12] efficiency is obtained by quantifying local image feature in so called "visual words" and comparing for each image pair the number of corresponding words. Although these methods scale up to a large number of images, the computational time for data association still increases linearly with the number of images. The method proposed in this paper results in less image comparisons, regardless of the method used to compare images. Indeed it can be combined with one of the mentioned efficient methods.

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 [13, 14]. 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. In general the assumption that observations are independent does not hold and this leads to an even more overconfident state estimate. On the other hand, if a SLAM method uses a conservative state estimate, the number of possible images to match is again too large [15].

It is indeed common to ignore the navigation prior when mapping environments containing large loops [16, 12]. In this case, data association for SLAM, is the same as data association for vision based topological mapping [17, 18, 19]. For small scale topological maps it is not necessary to reduce the data set because exhaustive data association is still possible [18, 19]. For mapping larger environments it is common to subsample in time, for example use only one frame per second [17] or to uniformly sample over the space that is being mapped, for example by using odometry measurements [20, 21].

These methods assume that the change in appearance is proportional to change in time or space. In realistic mapping scenarios where light changes occur, humans move, and the driving speed of the robot is not constant this assumption does not hold. Also, in small places such as corridors and door openings, the appearance changes relatively faster than in big convex spaces such as large rooms. In this paper we argue that using the change in appearance which is measured when comparing the already mapped images, results in better data association. If necessary the proposed algorithm can also be combined with such sampling methods, for example by first sampling in space and determining a set of key images given the CDS method from these sampled images.

An approach that does use the matching result of images that are in the map for associating new images applies spectral clustering on the connectivity matrix of the graph of matches [22, 23]. Clustering results in a set of subgraphs each containing images which are visually similar, such as images taken in the same room [24]. Per cluster an image is picked, resulting in a set of key images [22, 23]. However, solving a clustering problem is actually a more difficult and time consuming problem than the problem at hand, especially because the number of clusters is not known. In [25] finding clusters is simplified by only grouping sequentially acquired images. The major drawback is that all images taken from the same place but at a different time are put in a different cluster. Indeed one of the assumptions is that each location is visited only once [25]. In [26] the problem is solved more elegantly by incrementally clustering the graph, resulting in an algorithm which is more closely related to the method we propose.

Recently in [27] a method was proposed that only adds those images to the map that provide most information about the environment. An advantage of the approach is that the information gain is measured directly by inspecting the information matrix of the EIF (Extended Information Filter) based SLAM procedure. A drawback of the method is that once an image is added to the map, it can not be replaced by an even more informative image. The method we propose reconsiders at every iteration each previously acquired image to represent the image set.

In the field of computer vision there is a growing interest in estimating the 3D geometry of famous buildings or touristic sites from large sets of unordered data [28, 29]. The number of images is usually so large that it is not tractable to incorporate the relative poses between all image pairs. This problem is very much related to the SLAM problem discussed in this paper. In [28] an approach is used that is very similar to ours in which the Connected Dominating Set defines a skeletal graph that represents the complete graph. The CDS is computed only once from a graph that is obtained by computing point correspondences between each image pair. In our approach the graph is build incrementally, computing a new CDS for every new image, so we do not have to find correspondences between all image pairs.

3. Incremental data association based on image similarity

The map of the View based SLAM approach consists of the complete set of past robot poses and their matching images. In this section we propose a method to efficiently perform data association for such a mapping approach, comparing as few image pairs as possible. We do this by defining a set of key images using the Connected Dominating Set. Then these key images are used in a practical and efficient incremental data association scheme. We assume that a similarity measure is given that can take two images and computes if they are similar or not. Later in Section 4 we briefly describe the similarity measure used in the experiments.

3.1. The Connected Dominating Set

For now we assume that we already mapped part of the environment and found pairs of robot poses for which the corresponding images matched. The problem is to compute a minimal set of key images that best represents 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 close a loop 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, to find the loop closing event at C, only one of these images, C, D, E or F, has to be marked as key image and the rest can be ignored.

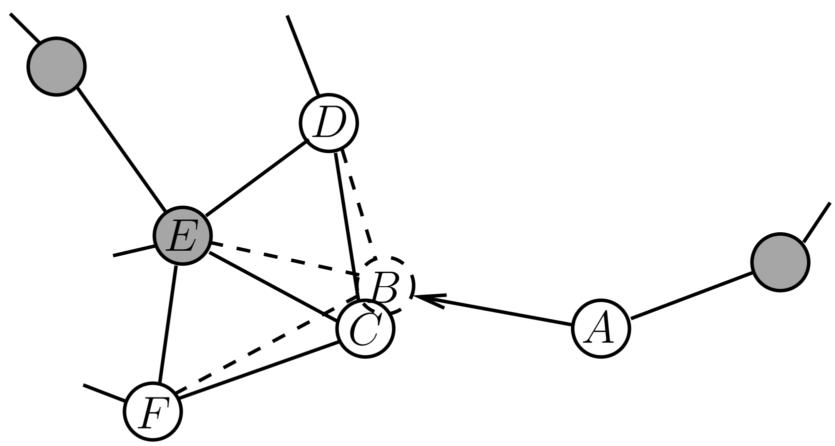


Figure 1: Figure explaining the rationale of using the Connected Dominating Set method. The circles denote robot poses and the links connecting them indicate that the images taken at the robot poses match. Grey circles indicate CDS nodes. The robot moves from robot pose A to a new robot pose B, which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

Of course, a loop closing can occur at all possible previous robot positions. It would suffice to compare the new image with a set of key images which has the property that every image either matched a key image or is a key image itself. 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 [7].

The set of all 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 $(u, v) \in S$ represents that the two images which 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':

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

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 [7]. 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. Below we describe the algorithm used in the experiments.

3.2. Approximation algorithm

Guha and Khuller describe a number of algorithms that find a CDS with close to the minimum number of nodes using computational time in the order of the number of nodes in the graph [7]. We implemented one of these algorithms and modified it slightly so that it can cope with non connected graphs. This modification is needed for example because the graph is in rare occasions not connected, usually caused by a single image that did not match any other image, because the view of the camera was blocked by persons walking near the robot.

The algorithm can be explained as follows, see also Figure 2:

1. Color every node of the graph white (Figure 2(a)).
2. Choose a white node with the highest number of neighbors.
3. Color this node black and color all white neighboring nodes gray (Figure 2(b)).
4. Choose a gray node that has the most links leading to white nodes (Figure 2(c)).
5. If no such gray node exists, goto 2
6. Goto 3 until there is no white node left.
7. The black nodes now compose the Connected Dominating Set (Figure 2(d)).

In [7] it is described how to implement this in $O(\#nodes)$.

An extension of this algorithm described in [7] that sometimes colors two nodes black in one iteration, instead of just one, was also implemented. With this

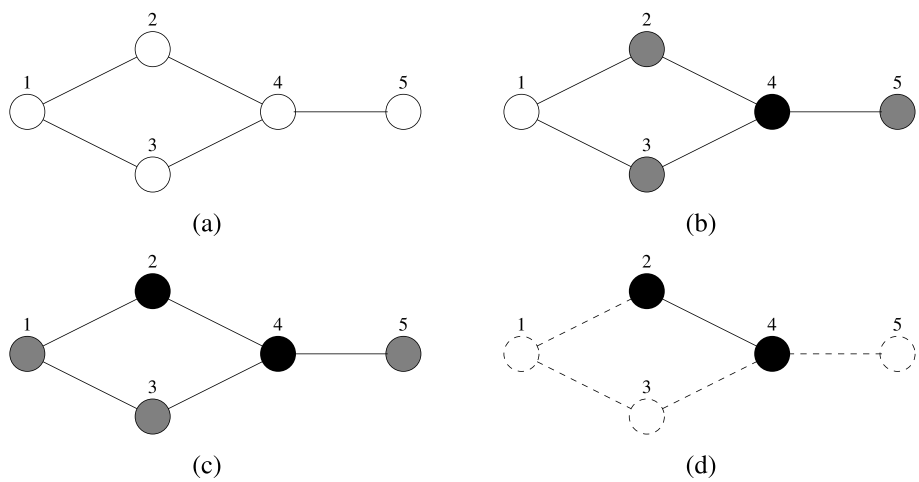


Figure 2: A simple example describing the approximation algorithm.

extension one can prove a nice upper bound on the size of the CDS. However, pilot studies on small mapping problems showed that the resulting CDSs were always larger than the algorithm described above, and the extended version was therefore not used in the experiments.

3.3. Incremental hierarchical data association

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 matched 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, since often the previously acquired image is not in the CDS. Therefore, we do not only compare with neighbors of matching CDS images but also the neighbors of all CDS images that matched the previously taken image. Pilot experiments have shown that this results in an increase of on average 9% of image comparisons. See Algorithm 1 for an overview of the data association scheme.

Algorithm 1 Incremental hierarchical data association scheme

graph $G = (V, S) = (\{\}, \{\})$
repeat
  Take a new image I_c
  Add current node c to graph $V \leftarrow \{V, c\}$
  V' = computeCDS(G)
  for all CDS nodes v' in V' **do**
    if match$(I_{v'}, I_c)$ **then**
      Add link: $S \leftarrow \{S, (v', c)\}$
    end if
  end for
  for all nodes v in V **do**
    if there is a node $v' \in V'$ that links to v: $(v, v') \in S$
        and links to c or p: $(v', c) \in S \vee (v', p) \in S$ **then**
      if match(I_v, I_c) **then**
        Add link: $S \leftarrow \{S, (v, c)\}$
      end if
    end if
  end for
  Current node becomes previous node: $p \leftarrow c$
until end of mapping

4. Comparing images

This section briefly describes the matching technique used to compare two images. For the CDS data association scheme any image comparison technique can be used, for example a fast hierarchical method [11] or a bag-of-words method that use training sets to learn how discriminative image features are [12]. In the experiments in this paper we used a method based on corresponding local image features and imposing the epipolar constraint [30].

Images taken by an omnidirectional vision sensor are first mapped to panoramic images [31], from which feature points are found using the Scale Invariant Feature Transform (SIFT) [32]. Features are described by the standard SIFT descriptor of 128 dimensions.

A set of point correspondences between two images is determined by applying the standard matching scheme as described in [32]. The number of point correspondences could be used to determine if the two images match. 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 [33]. 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 [31, 34].

The 3x3 matrix E is estimated using a variant of the 8-point algorithm [35], for which the constraint is added that the camera moves over a planar surface [36]. This algorithm is used inside the RANSAC robust estimator [37], which estimates the epipolar geometry and at the same time determines the number of fitting correspondences [35, 38]. A point correspondence fits E if it has a small Sampson distance [35, 38] and the corresponding 3D world point has a positive depth in both cameras [39].

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 certain threshold, then the images match. Pilot studies in an office environment with a threshold of 0.1 resulted in a lot of good matches and no false matches.

5. Experiments and results

We evaluated the performance of the proposed CDS data association method on several realistic datasets. In the first experiment we compared our CDS data association method with the straightforward method in which every new image is compared with all previously taken images. In a second experiment the CDS method is compared with data association methods that use sampling over time or over space. Finally we focus on the event of loop closing and traversing a loop in the environment twice.

5.1. Datasets, set-up and evaluation measures

We used four datasets: an "office set" acquired in our university building, two "home sets" taken in real home environments[2] and one "outdoor" set acquired in a typical suburb environment. For all these datasets, including the one taken at the university building, the conditions were far from ideal "lab conditions", including bad lighting conditions, people walking close to the camera and scenes with a low amount of texture. Perhaps more important was the fact that within each dataset the conditions differed for different parts of the environment, see Figure 4.

The office set (1754 images, 4 Hz) includes a hallway that is dark and has only few visual features, compared to the rest of the office environment, compare the images in Figure 4(a)-(b). The set is interesting because it shows how the proposed method copes with a robot traversing a same loop in the building twice.

The first home set (1436 images, 5 Hz) is taken in a relatively feature rich home environment. However, the images were shot in the evening. This resulted in somewhat dark images and more importantly motion blur during sharp turns, because of the higher shutter times needed to capture bright enough images, compare the images in Figure 4(c)-(d).

The second home set (2071 images, 7 Hz) was taken while people walked in close vicinity to the robot. The images were captured during day time with the blinds open, causing some images to be very bright (Figure 4(f)), while others not being in direct view of a window to be quite dark (Figure 4(e)).

The outdoor set (826 images, 1 Hz) was acquired by a car driving through a suburb of Hoofddorp, The Netherlands. The lighting conditions were relatively good and the environment had plenty visual cues. However, the images contained a limited number of useful features, because a large part of the camera view was

[2] The home sets, including images, odometry, sonar and laser range data (all timestamped), are available from `http://www2.science.uva.nl/sites/cogniron/`.

from the overcast sky or the roof of the car where the camera was mounted on (see Figure 4(g)).

To get an idea of the structure of the environments see Figure 3. For a detailed description of the acquisition of the home datasets see [40]. All indoor 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 outdoor dataset was acquired using a car with the same omnidirectional vision system mounted on its roof [41]. The computer vision and data association algorithms were implemented in C++ and were running on a 2Ghz laptop mounted on the robot[3].

For evaluation we need to measure the speed-up of the proposed data association method with respect to an exhaustive data association scheme, as well as the amount of correct matches found. However, we do not want to evaluate the image comparison technique itself. Therefore we treat the results of the exhaustive data association scheme as the ground truth image matching data.

The speed-up is now measured by the number of images comparisons the exhaustive data association performed divided by the number of comparisons performed by the proposed method. In a similar fashion the amount of correct image matches found is measured by the number of images matches found by the proposed method divided by the number of matches found by exhaustive data association.

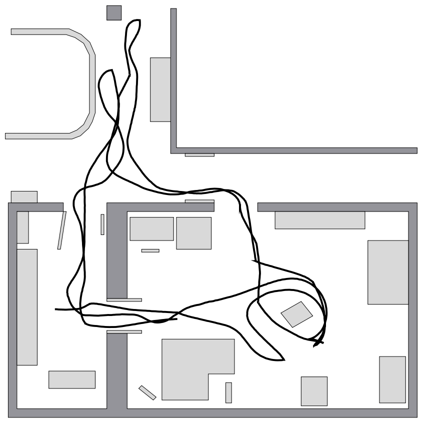
Another interesting evaluation criterion is the percentage of image comparisons that resulted in a match. This value tells something about the quantity of information that is gained per comparison. We call this value "efficiency" and use it for comparing different sampling approaches.

5.2. Comparison to exhaustive data association

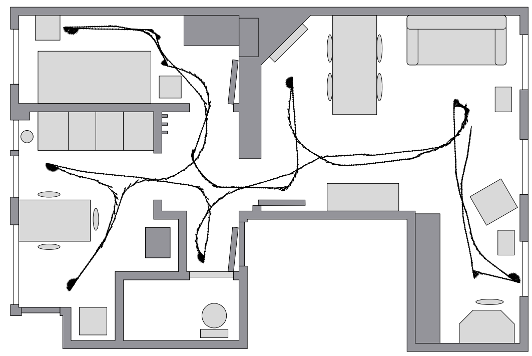
On all four datasets, both the proposed data association method was applied, as well as an exhaustive data association scheme in which each new image was matched with all previously acquired images.

The results of the CDS data association method are visualized in Figure 5 as connectivity graphs, linking the matching images and using hand corrected odometry and GPS information for the position of the image-nodes. Note, however, that

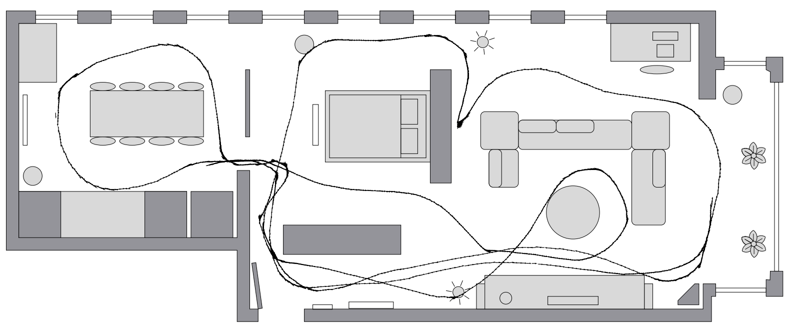
[3] All software used in the experiments is available online at `http://www.science.uva.nl/research/isla/downloads/VisualMapping/index.html`.



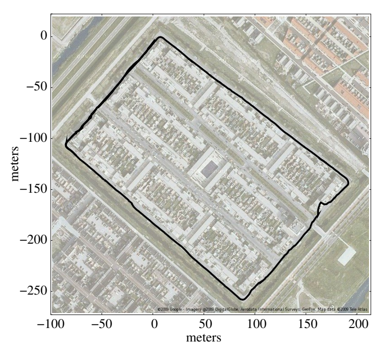
(a) Office



(b) Home 1



(c) Home 2

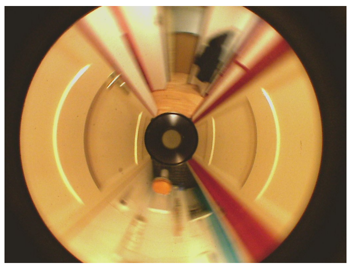


(d) Outdoor

Figure 3: Ground floor maps of the indoor environments and a satellite image of the outdoor environment. The trajectory of the robot is indicated with the black line. The position of the furniture is approximate.

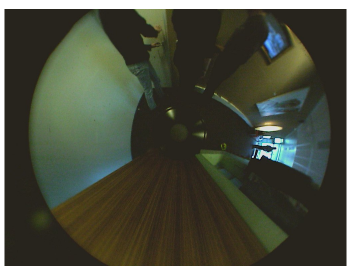
(a) Office, few features

(b) Office many features



(c) Home 1, motion blur

(d) Home 1, feature rich

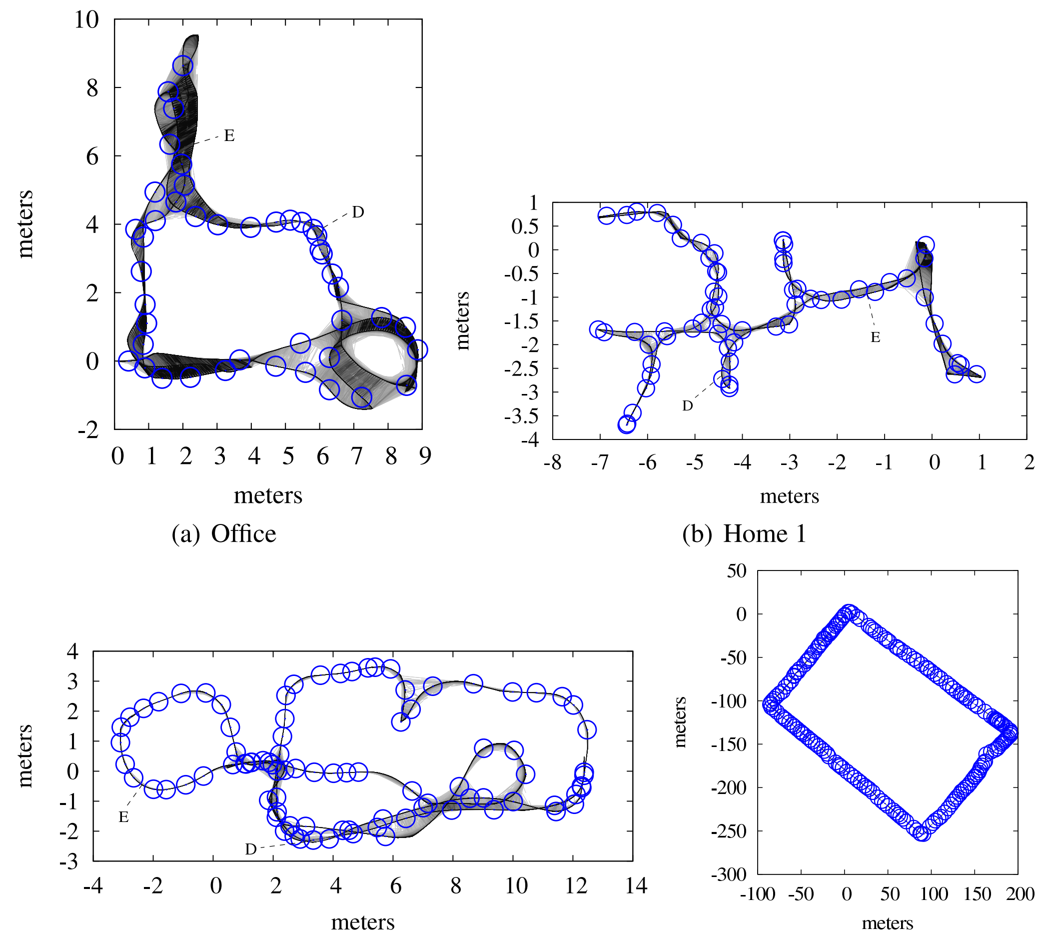


(e) Home 2, dark

(f) Home 2, bright

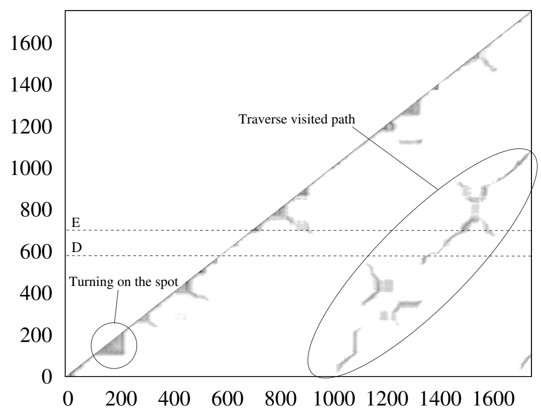
(g) outdoor

Figure 4: Example images taken by the omnidirectional vision system from the datasets. From each indoor dataset there are two images. On the left there are images that are hard to match and on the right images that are easy to match. In Figure 5 the positions of the robot are indicated while taking these images.

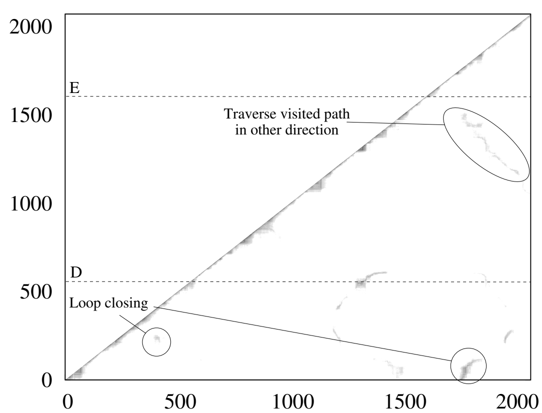


(a) Office (b) Home 1 (c) Home 2 (d) Outdoor

Figure 5: Results of CDS data association method. The graphs were plot using hand-corrected odometry information. 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. The nodes indicated with a “D” correspond to example images that are difficult to match, plotted in the left column of Figure 4. Nodes indicated with an “E” are easy to match and plotted in the right column. 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).



(a) Office



(b) Home 2

Figure 6: Connectivity matrices of the final graphs of the Office set and the Home 2 set built using the CDS method. Image pairs with a higher similarity are represented with darker pixels. The entries on the main diagonals are the result of matching sequential images, while the off-diagonal entries reflect instances of loop-closing.The “D” and “E” again indicate the difficult and easy images shown in Figure 4

Table 1: Comparison of data association based on the CDS and the Full data association, matching every new image with all previous images. The number of images that are compared during map building is much lower for the CDS method than for the Full method, resulting in a speed up of data association (the number of compared by Full divided by the number of compared by CDS). Nevertheless the percentage of matches found by the CDS method is close to 100% (matches found by CDS divided by the matches found by Full).

		Office	Home 1	Home 2	Outdoor
	#images	1754	1436	2071	826
Full	compared	1537381	1030330	2143485	340725
	matched	92831	39631	66618	2219
CDS	compared	202334	108658	214881	151082
	matched	91330	38397	64108	2198
speed up		759%	948%	997%	225%
% matches found		98%	97%	96%	99 %

the odometry information is not used by the proposed method. As can be seen for all datasets a lot of images matched. Figure 6 visualizes some of the resulting graphs found by the CDS method as connectivity matrices, which more clearly shows the loop closing image matches by the off-diagonal non-zero values.

The connectivity graphs and connectivity matrices computed by using exhaustive data association are visually indiscernible from the ones computed using the proposed method and are therefore omitted.

In Table 1 the results obtained with the CDS method are compared with the exhaustive data association. As can be seen the CDS method is on average 7 times faster, determined by dividing the number of image comparisons done by the exhaustive method by the number of comparisons done by the CDS method. Nevertheless, the CDS method finds on average 97.6 % of the matches found by comparing with all images.

In Figure 7 shows a more detailed plot of the number of image comparisons performed while the robot is mapping the office set. The other datasets resulted in similar plots. As can be seen the number of comparisons for the exhaustive data association scheme increases linearly with the number of images in the map. The number of comparisons performed by the CDS method barely increases. The resources used to determine the CDS itself is negligible compared to the time

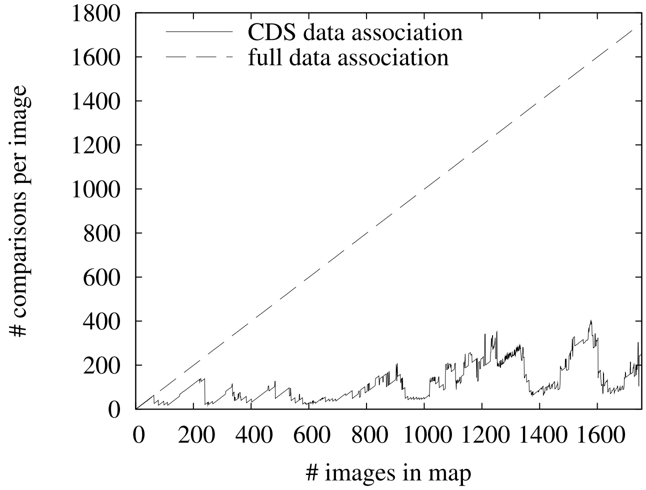


Figure 7: Comparison of the number of images comparisons for each new image of the CDS data association and the conventional brute force method while the map is growing. The fluctuations in the graph of the CDS is mostly caused by the variable speed of the robot. When the robot is moving slowly or moving on the spot, then relatively many images will match, see also Figure 6.

needed for the actual image matching. For all the datasets the computation time was always smaller than 10 ms.

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 characteristics parts of the home environments as depicted in Figure 4 and discussed in Section 5.1. In the connectivity graphs (Figure 5) the robot positions of the example images are visualized with a “D”, for images that are difficult to match, and an “E”, for images that are easy to match. It is clear that in the neighborhood of the difficult images relatively more key images were picked than parts of the environment where good images were acquired.

5.3. Comparison of sampling techniques

The investigation of the distribution of the key image over the tested datasets suggests that picking key images based on previous image matches results in a subset of images that better represents the complete set. 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 compared is more or equal to the number of image-pairs compared by the CDS method. Thus the CDS method will use less or equal the amount of computational time. All methods were used in the same hierarchical incremental data association scheme, described in Section 3.3.

- The first method picks images randomly from the image set. During each iteration a new set is chosen with an average number of images equal to $.063$ times the number of images in the map.

- The second method uses the odometry measurements to sample over displacements of the robot. After each 43 cm an image is added to the set of key images.

- The third approach samples images over time. After each 3.8 seconds a mapped image is added to the set of key images.

In Table 2 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 (efficiency) is highest of all methods, including of course that of a full matching scheme. Although we did not investigate the number of mismatches that were made the percentage of successful comparisons does indicate that the proposed method is more robust against false positives of the image matching technique.

5.4. Revisiting places

In the office set the robot was driven twice over the same loop in the environment. This can be seen clearly in the graph in Figure 5(a) and is also visible in the connectivity matrix in Figure 6(a) by the second diagonal parallel to the main diagonal.

While mapping the environment, more and more images acquired at different positions are added to the dataset and thus the size of the set of key images grows. This is depicted in Figure 8. At image 1020 the robot finished its first loop in the environment and had a CDS size of 39 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 5(a). Because of these links only

Table 2: Comparison of the CDS method to different sampling approaches. The parameters of the different sampling approaches were set to such values that the number of image comparisons was equal to that of the CDS method. The CDS method finds the highest percentage of matches (matches found by CDS divided by the matches found by Full) and, thus, also the highest percentage of image comparisons that result in a match (efficiency).

method	key images	matched	% matches found	efficiency
Full	1436	39,631	100%	4%
Random	92	28,243	71%	26%
Position	82	32,836	83%	29%
Time	77	33,685	85%	31%
CDS	65	38,397	97%	35%

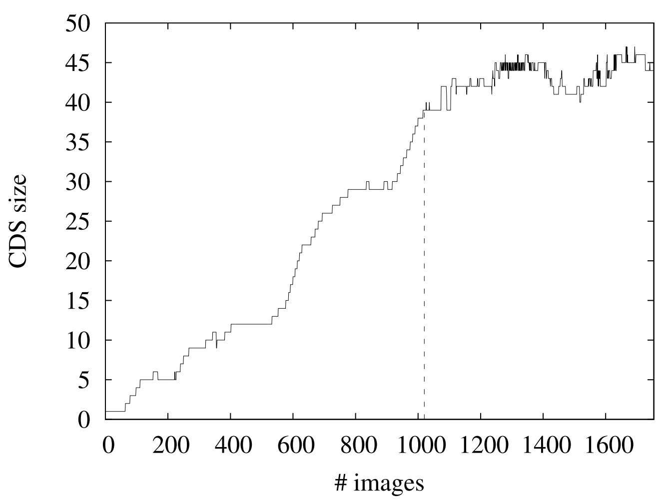


Figure 8: The number of nodes in the CDS, while the map of the office is growing. The vertical dashed line indicates the beginning of the second traversal of the loop.

a few extra nodes were added to the CDS during this second loop resulting in a total of 44 nodes in the final CDS (which are indicated in Figure 5(a)).

Note that the set of 44 key images of the final map are not composed of the 39 key images of the first loop and 5 extra images of the second loop. The optimal set of key images is determined for each new image that is added to the map. Images of the second loop might better represent images taken of a particular part of the environment, making images of the first loop redundant. In the office set the final CDS is composed of 11 images of the first loop and 33 of the second loop.

Figure 8 also shows that in some occasions the number of nodes in the CDS decreases. This happens if a new image is added which matches already mapped images that did not match each other. This indicates that new images can represent an existing set of images better than the previous key images taken from the set.

6. Conclusion

In this paper we proposed an efficient data association method for view based SLAM. Our approach is based on the fact that we consider only a selection of the previously acquired images for matching new images. The selected set of representative images covers the complete set of previously acquired images and we can efficiently detect loops in the trajectory of the robot. We have shown that the problem of finding the minimal number of key images is equivalent to finding the smallest Connected Dominating Set (CDS).

The experimental results show that our method leads to a more efficient distribution of key images. From areas in the environment that are harder to match, for example because of bad lighting conditions, more images are picked. In this way loop closure is much more robust, even if it occurs in such an area.

The CDS method is built in a hierarchical data association scheme that incrementally builds a map without using any prior knowledge about the environment. The set of representative images is dynamic. After each newly acquired image a CDS is determined that best represents the set of images at that moment.

The method is applied on four challenging datasets mostly acquired in real home environments. In all datasets our method finds approximately the same map as is formed in the "full" case that all images are used. However, only 13% of the computational time is used. The efficiency of our method (the number of matches divided by the number of image comparisons) is 35%, which is high compared with the full case (4%).

When comparing our method with other known sampling techniques we found that our method outperforms these method because it results in a smaller set of key images, while it finds much more matching image pairs in the same amount of computational time. Our method finds 97% of the matches that were found with a "full" methods while position and time based methods found less then 85%.

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 [11] or a bag-of-words methods that use training sets to learn how discriminative image features are [12].

In the experiments we used image sets in the order of a few 1000 images. For such dataset sizes, exhaustive data association, used for evaluation, is still possible, though time consuming. Using the CDS method datasets can scale up by a factor 10. In [42] we used a SLAM system to build a map with more than 10,000, implicitly using the CDS method for data association.

References

[1] O. Booij, Z. Zivkovic, and B. Kröse. Sampling in image space for vision based slam. In *Proceedings of the Robotics: Science and Systems workshop Inside Data Association*, June 2008.

[2] H. Durrant-Whyte and T. Bailey. Simultaneous localisation and mapping (slam): Part I The essential algorithms. *Robotics and Automation Magazine*, 13(2):99–108, June 2006.

[3] 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.

[4] 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)*, pages 1151–1156, Acapulco, Mexico, August 2003.

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

[6] Z. Zivkovic, B. Bakker, and B. Kröse. Hierarchical map building using visual landmarks and geometric constraints. In *Proceedings of the International Conference on Intelligent RObots and Systems (IROS)*, Edmonton, Canada, August 2005. IEEE/RSJ.

[7] S. Guha and S. Khuller. Approximation algorithms for connected dominating sets. *Algorithmica*, 20(4):374–387, April 1998.

[8] O. Booij, Z. Zivkovic, and B. Kröse. Sparse appearance based modeling for robot localization. In *Proceedings of the International Conference on Intelligent RObots and Systems (IROS)*, pages 1510–1515, Beijing, China, October 2006. IEEE/RSJ.

[9] Michael Milford and Gordon Wyeth. Mapping a suburb with a single camera using a biologically inspired slam system. *IEEE Transactions on Robotics*, 24(5):1038–1053, 2008.

[10] A.C. Murillo, C. Sags, J.J. Guerrero, T. GoGoedemé, T. Tuytelaars, and L. Van Gool. From omnidirectional images to hierarchical localization. *Robotics and Autonomous Systems*, 55(5):372–382, 2007.

[11] 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 (ICCV)*, volume 2, pages 1470–1477, Nice, France, October 2003. IEEE.

[12] 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.

[13] 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.

[14] A. J. Davison, I. D. Reid, N. Molton, and O. Stasse. MonoSLAM: Real-time single camera SLAM. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(6):1052–1067, 2007.

[15] V. Ila, J. Andrade-Cetto, R. Valencia, and A. Sanfeliu. Vision-based loop closing for delayed state robot mapping. In *Proceedings of the International*

Conference on Intelligent RObots and Systems (IROS), pages 3892–3897, San Diego, USA, November 2007. IEEE/RSJ.

[16] P. Newman, D. Cole, and K. Ho. Outdoor slam using visual appearance and laser ranging. In *Proceedings of the International Conference on Robotics and Automation (ICRA)*, pages 1180–1187, Orlando, Florida, USA, May 2006. IEEE.

[17] I. Ulrich and I. Nourbakhsh. Appearance-based place recognition for topological localization. In *Proceedings of the International Conference on Robotics and Automation (ICRA)*, volume 2, pages 1023 – 1029, San Francisco, California, USA, April 2000.

[18] O. Booij, B. Terwijn, Z. Zivkovic, and B. Kröse. Navigation using an appearance based topological map. In *Proceedings of the International Conference on Robotics and Automation (ICRA)*, pages 3927–3932, Roma, Italy, April 2007. IEEE.

[19] F. Fraundorfer, C. Engels, and D. Nister. Topological mapping, localization and navigation using image collections. In *Proceedings of the International Conference on Intelligent RObots and Systems (IROS)*, pages 3872–3877, San Diego, USA, November 2007. IEEE/RSJ.

[20] R. Sim and G. Dudek. Learning and evaluating visual features for pose estimation. In *Proceedings of the International Conference on Computer Vision (ICCV)*, pages 1217–1222, Kerkyra, Greece, September 1999. IEEE.

[21] 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.

[22] F. Li and J. Kosecká. Probabilistic location recognition using reduced feature set. In *Proceedings of the International Conference on Robotics and Automation (ICRA)*, pages 3405–3410, Orlando, Florida, USA, May 2006. IEEE.

[23] 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 (ICRA)*, pages 4283–4288, Roma, Italy, April 2007. IEEE.

[24] Z. Zivkovic, O. Booij, and B. Kröse. From images to rooms. *Transactions on Robotics*, 55(5):411–418, May 2007.

[25] J. Kosecká, L. Zhou, P. Barber, and Z. Duric. Qualitative image based localization in indoors environments. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 3–10, Madison, WI, USA, June 2003. IEEE.

[26] C. Valgren and A. J. Lilienthal. Incremental spectral clustering and seasons: Appearance-based localization in outdoor environments. In *Proceedings of the International Conference on Robotics and Automation (ICRA)*, pages 1856–1861, Pasadena, California, USA, May 2008. IEEE.

[27] Weizhen Zhou, J.V. Miro, and G. Dissanayake. Information-efficient 3-d visual slam for unstructured domains. *IEEE Transactions on Robotics*, 24(5):1078–1087, 2008.

[28] N. Snavely, S. M. Seitz, and R. Szeliski. Skeletal sets for efficient structure from motion. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, Anchorage, Alaska, USA, June 2008. IEEE.

[29] F. Schaffalitzky and A. Zisserman. Multi-view matching for unordered image sets, or "How do I organize my holiday snaps?". In *Proceedings of the European Conference on Computer Vision (ECCV)*, volume 1, pages 414–431, Copenhagen, Denmark, 2002. Springer.

[30] B. Kröse, O. Booij, and Z. Zivkovic. A geometrically constrained image similarity measure for visual mapping, localization and navigation. In *Proceedings of the 3rd European Conference on Mobile Robots*, pages 168 – 174, Freiburg, Germany, 19-21 September 2007.

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

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

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

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

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

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

[37] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[38] 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.

[39] O. Chum, T. Werner, and J. Matas. Epipolar geometry estimation via ransac benefits from the oriented epipolar constraint. In *Proceedings of the International Conference on Pattern Recognition (ICPR)*, volume 1, pages 112–115, 2004.

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

[41] I. Esteban, O. Booij, J. Dijk, and F. C. A. Groen. On the bending problem for large scale mapping. In *Proceedings of the International Conference on Intelligent RObots and Systems (IROS)*, St. Louis, USA, October 2009. IEEE/RSJ. (To appear).

[42] I. Esteban, O. Booij, Z. Zivkovic, and B. Kröse. Mapping large environments with an omnivideo camera. In *Workshop Proceedings of the Conf. On Simulation, Modeling and Programming for Autonomous Robots*, pages 297–306, Venice, Italy, 2008.