Bearing-Only SLAM in everyday environments using Omnidirectional Vision

Siegfried Hochdorfer, Matthias Lutz and Christian Schlegel
University of Applied Sciences Ulm
Computer Science Department, Prittwitzstr. 10, 89075 Ulm, Germany
email: {hochdorfer, lutz, schlegel}@hs-ulm.de

***Abstract*— SLAM (simultaneous localization and mapping) mechanisms are a key component towards advanced service robotics applications since knowledge about the own pose and representations of the environment are needed for a series of high level applications. A futher challenge is to design service robots for life-long and robust operation in dynamic environments. To achieve this, two approaches are combined in this work.**

First, an approach from our previous work [1] is used to handle the problem of the ever growing amount of landmarks over time. Typically, SLAM approaches just accumulate features over time and do not discard them anymore. Therefore, the required resources in terms of memory and processing power are growing over time. In the presented approach, the absolute number of landmarks can be restricted by an upper bound since we introduce a method to specifically select and replace landmarks once the upper bound has been reached. The second approach [2] is related to improving the robustness of the landmark assignment problem in case of image based features as needed with natural landmarks.

Real-world experiments are used to demonstrate the performance of our approach. These experiments are performed on a P3DX-platform with an omnidirectional vision based bearing-only SLAM approach.

I. INTRODUCTION

Localization and mapping are fundamental problems in service robotics. Knowledge about the own pose and representations of the environment are needed for a series of high level applications. For example, acting in a goal oriented manner as required in fetch-and-carry tasks is simplified significantly.

Omnidirectional cameras are beneficial for SLAM because the large field of view of 360 degree allows the observation of landmarks in every direction. Therefore, the orientation of the robot for landmark recognition is not important. With this property, it makes no difference in which direction the robot travels along a path. For robots with standard camera systems, it is difficult to reobserve landmarks when they travel a known path in the opposite direction. As long as no calibrated System is needed, omnidirectional cameras are cheap and small and thus perfect for service robotics applications. As drawback, one does not get range information to landmarks as needed with many algorithms. However, bearing-only SLAM mechanisms can take advantage out of an omnicam image since these only require angles to landmarks. To determine angles to landmarks, even no methods to correct image distortion or to correct perspective have to be applied.

Although many SLAM solutions already exist, most of them still do not address core requirements of service robotics. Service robots should be designed for life-long and robust operation in dynamic environments.

Life-long operation raises the question of limited resources. Typically, SLAM approaches just accumulate features over time and do not discard them anymore. For example, dynamic objects can introduce lots of new features that are never again removed but that are of use for a short period of time only. Therefore, the required resources in terms of memory and processing power are growing over time.

Robustness in everyday environments is mostly related to the landmark association problem. In service robotics applications this is demanding since one cannot rely on artificial landmarks. In contrast, one has to rely on natural landmarks. These are often identified on recurring structures like door and window frames. Thus, simple feature matching is not sufficient and would result in false assignments. In particular, EKF-based (Extended Kalman Filter) SLAM approaches are seriously affected by false assignments.

The contribution of this paper is twofold. First, we describe an approach to handle the landmark assignment problem in case of image based features. We exploit that feature vectors can be sorted based on an Euclidean distance measure. Thus, the set of landmark candidates that are similar in terms of their descriptor can be retrieved efficiently. However, this similarity does not yet take into account the probability of expecting a certain landmark candidate at a certain observation position. Thus, the set of landmark candidates is further shortened based on the Mahalanobis distance. This approach combines efficient feature retrieval with spatial plausibility, resulting in a suitable trade-off between efficiency and robustness.

Second, an approach to address the ever growing number of landmarks in life-long operation is presented. The absolute number of landmarks is restricted by an upper bound. This requires a method to specifically select and replace landmarks once the upper number of landmarks has been reached. Our approach is to evaluate landmarks based on their utility for localization purposes which is different from just replacing the most uncertain landmark.

Proceedings of the 2nd. Workshop on Omnidirectional Robot Vision
A workshop of the 2010 IEEE International Conference on Robotics and Automation (ICRA2010)
Anchorage, Alaska, USA, May 7, 2010, ISBN 978-88-95872-02-5
pp. 26-31

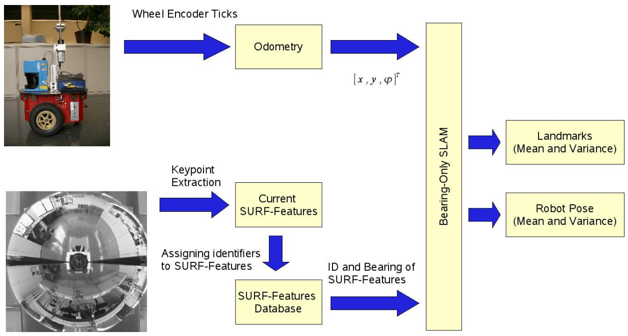

Fig. 1. The overall bearing-only SLAM system based on SURF features [3] as visual landmarks

II. RELATED WORK

The general approach of using an Extended Kalman Filter (EKF) for bearing-only SLAM based on artificial landmarks has been described and evaluated in [4] and [5]. The bearing-only SLAM approach was then extended to work with SIFT features (Scale-Invariant Feature Transform) as visual landmarks [6].

A visual bearing-only SLAM system based on a Rao-Blackwellized particle filter is presented by Strasdat [7]. It is based on SURF [3]. The feature matching is improved by using a cost function taking into account the expected landmark position in terms of image coordinates.

Andreasson proposes in [8] an efficient graph-based visual SLAM approach called MiniSLAM. The approach uses odometry data and image similarity of the omnidirectional images to determine the maximum likelihood estimate of the image poses.

Landmark rating needs a measure for determining the benefit of a landmark for localization purposes. In [9], the observation region of a landmark together with the landmark pose uncertainty is used for defining a measure for the benefit of a landmark. K-means clustering [10], [11] is used to identify regions in the environment with a high landmark density. The k-means algorithm separates the landmark representatives into n clusters. The number of clusters is set proportional to the number of landmark representatives and is selected empirically.

A quality measure to compute the best landmark out of a set of landmarks is also used by Dissanayake [12]. First, all landmarks are collected whose state changes in the current step from visible to invisible. From this set only the highest quality landmark is kept and all others are discarded. Thus, the selected highest quality landmark is a single representative for the set of previously visible landmarks. However, selection of landmark representatives is based on a local set of landmarks and thus depends on the exploration path and the resulting visibility sequence. There is still no global measure of landmark quality. Nevertheless, this is one of the rare approaches addressing landmark deletion with respect to a landmark's use in terms of observability.

A fundamentally different approach is proposed by Strasdat [13]. The presented approach uses Monte-Carlo Reinforcement Learning to learn landmark selection policies that optimize the navigation task. He demonstrates his approach in two scenarios. The first is a single goal navigation task. The second is a round-trip navigation task where subgoals are visited more than once. Due to the complexity of the learning algorithm and the number of training episodes, it is not feasible to learn these policies during real-world experimentation. Therefore, Strasdat recommends to learn the policies in simulation.

III. LANDMARK ASSIGNMENT

Robust assignment of identifiers to SURF-Features is one of the major problems in EKF based Visual SLAM approaches. This is due to the brittleness of the EKF in case of false assignments. Common data association approaches as for instance Joint Compatibility Branch and Bound (JCBB) [14] assume indistinguishable landmarks. SURF features used in our approach has distinctive descriptors. A strong hypothesis of data association is obtained after the descriptor comparison. Therefore an expensive branch and bound search $O(1.53^n)$[14] as in JCBB is not needed.

Our approach is to combine efficient feature retrieval with spatial plausibility. The latter is based on the Mahalanobis distance. This reduces the number of *false positives* and thus increases robustness. For performance reasons, a memory-efficient and computationally-efficient kd-tree [15] is used to store the SURF features. Its search complexity is $O(n * log(n))$ with n the number of features in the kd-tree.

The comparison of SURF as well as SIFT descriptors can be based on different methods. The Euclidean distance ratio is recommended and used by Lowe [16]. Even though this approach is widely used, it is sensitive to the number of similar feature descriptors. A different approach interprets descriptor vectors as bins. The comparison of two descriptors is then equivalent to the comparison of two distributions. Appropriate standard methods are then provided in statistics. A χ^2 test for comparing multidimensional vectors is recommended by Kiang [17]. Since the sample size is too small, he also states that the ANOVA test (analysis of variance) cannot be used for descriptor comparison. In first approaches, we used the correlation coefficient since it can be calculated easily. This is not a suitable approach since it checks for similarity only but ignores the distance between the descriptor vectors.

The first step (*A*) in the assignment process is to find the nearest neighbor (Euclidean distance) in the kd-tree. We then compare this descriptor from the database with the descriptor of the observed SURF feature by a χ^2 Test. If the result of the χ^2 Test is above a threshold, the match failed and the feature gets a new identifier. Otherwise, the observed feature is a reobservation of an already known landmark (either initialized or uninitialized) and the feature is assigned with the identifier from the database.

In case we successfully matched an initialized landmark (*B*), this so far is based solely on the similarity of the descriptors of the landmarks. There is no validation of the spatial plausibility of the pose of the matched landmark done

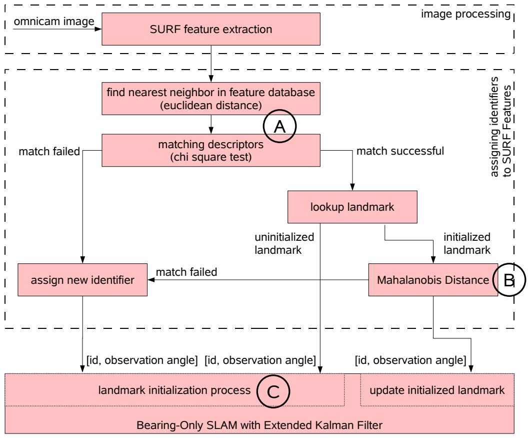

Fig. 2. The decision tree behind the identifier assignment procedure

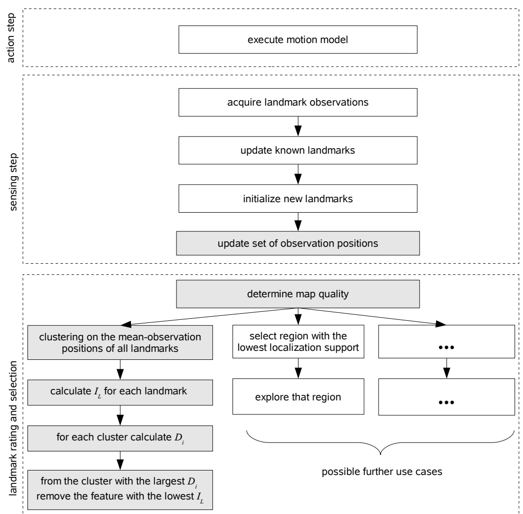

Fig. 3. Extension of the standard SLAM approach by mechanisms for landmark rating and selection.

until now. However, the EKF provides additional information about expected observation poses. Thus, we can apply a validation gate based on the Mahalanobis distance to cross check the plausibility of feature matches before this match is forwarded to the EKF.

An initialization of a new landmark (*C*) is performed only if there exists a *well-conditioned* pair of measurements for a non-initialized landmark. As described in detail by Bailey [4], a new feature is considered *well-conditioned* if the true probability density function of its location closely resembles the Gaussian approximation obtained from a Jacobian-based linearized transform. After initialization of a landmark, the remaining measurements of it are applied in a batch update. In the batch update, the position and the variance of the landmark is known. Therefore, it is possible to again use the Mahalanobis distance as validation gate to reject wrong assignments (*false positives*).

IV. Landmark rating and selection

The principal steps of a SLAM mechanism are illustrated in Figure 3. The highlighted steps denote the functions for landmark rating and selection presented in [9]. The sensing step keeps track of the set of robot poses from which a landmark has been observed so far. This provides the basis for describing from where in the environment a certain landmark is observable. The observability is then used to evaluate the benefit of a landmark for localization purposes. At arbitrary points in time, one can determine the benefits of landmarks for localization purposes. One approach would be to determine the landmark with the lowest impact on reducing the robot pose uncertainty while still ensuring coverage of the operational area. This landmark might then be removed from the SLAM representation of the environment.

The viewpoint location of every landmark is the key to cover the operational area of the robot with landmarks. Therefore, this work focuses on the improvement of the landmark distribution. A future approach would determine regions that currently have a too low localization support. These regions are either not covered by observable landmarks or the observable landmarks do not reduce the pose uncertainty below the desired uncertainty. One could then specifically explore such regions.

To estimate regions of viewpoints locations at all, we intent to cluster them into local groups. In the previous work [9] K-means clustering as one kind of the partitional clustering techniques is used. The DBSCAN clustering algorithm [18] is a density-based approach with only two parameters, $MinPts$ and Eps. Removing a landmark from a region with high localization support results in a small degradation of robot localization quality only. Clustering algorithms are used to identify regions in the environment with a high landmark density. DBSCAN clustering is especially suitable to determine those regions. The algorithm typically constructs clusters around local dense maxima, separated by regions of low density. Further advantages of DBSCAN clustering are that the algorithm does not need to know the number of clusters in advance and the efficiency $(0(n * log(n)))$ of the algorithm is equal to K-means [18] [19]. The difference of information content [12] D_i of the landmarks within each cluster C_i is calculated. The cluster with the maximum difference is determined by $max(D_i)$. From this cluster the landmark with the lowest information content is then removed. Removing landmarks within low density regions is critical. All landmark representatives with a distance greater Eps to the Eps-neighborhood are considered as outliers. Outliers are not part of a cluster and are thus never removed.

V. Results

In this section, the results of the experiments are discussed in detail. The performance is demonstrated within a standard indoor environment by means of real-world experiments. We ran the same experiment three times with the same real-world dataset. The runs are different only with respect to their landmark reduction mechanism. Due to using the same real-world dataset, the results of the runs can be directly compared. The experiment has been performed in our lab, the adjacent hallway and a neighbored room (see figure V).

Proceedings of the 2nd. Workshop on Omnidirectional Robot Vision
A workshop of the 2010 IEEE International Conference on Robotics and Automation (ICRA2010)
Anchorage, Alaska, USA, May 7, 2010, ISBN 978-88-95872-02-5
pp. 26-31

Fig. 4. The real world environment and the robot used in our experiments. The upper left image show the ZAFH laboratory, the lower left image the adjacent hallway and the right image our Pioneer 3DX robot equipped with Microspace PC and omnidirectional camera.

We have not taken any precautions to avoid direct sunlight, specular reflections and differences in brightness. The low height of the robot leads to occlusions of landmarks by tables and chairs. These experiments are performed on a P3DX-platform (see figure V) with a bearing-only SLAM approach [4] [5] with SURF Features [3] as landmarks.

The basic configuration is the same in all three runs. The path length of a run is approximately 150m and results in 510 observation positions. The travel distance between two observation positions is approximately 0.3m. During the experiment, two major loops with a length of approximately 14m have to be closed. Each of the three experimental runs however, has an individual configuration with respect to the landmark reduction mechanism and the clustering algorithm. The first run has no landmark limitation: as many landmarks as possible (579) are initialized. The other two runs are limited by an upper bound of 150 initialized landmarks. They differ in the way how viewpoint location clusters are built, either K-means or DBSCAN clustering techniques are used. Using K-means clustering, the number of clusters is set dynamically to 1/4 of the number of currently known landmarks. The parameters of DBSCAN are set to $\epsilon = 0.5$ and the minimum number of landmarks in each cluster has to be two. The uncertainty of the observation angle measurement is set to ${\sigma_\alpha}^2 = 0.2727deg^2$. This value is derived from a 1-pixel jitter in the omnidirectional image. The parameters ${\sigma_d}^2$ and ${\sigma_\phi}^2$ of the action model of the robot are determined according to $\lambda_d = 0.001m^2/m$ and $\lambda_\phi = 4deg^2/360deg$. Due to the lack of GPS in indoor environments, it is quite difficult to get the ground truth position of the robot. We solve the problem of determining the ground truth position by manually measuring the distance from the robot to two a priori known coordinates in the environment with a Bosch Digital Laser Rangefinder (DLE 150). The 2σ value of our ground truth poses is [$0.06m^2$; $0.04m^2$].

In the first part of the experiments, we introduced methods in terms of localization quality of the robot in the environment. An ever growing number of landmarks does not automatically result in an improvement of the localization quality. In addition to that we test the capability of the methods to cover the operational area with landmarks. A good landmark coverage is pivotal for the localization of a service-robot. The method should select those landmarks that best cover the operational area taking into account the benefit for localization purposes.

A. Localization quality

The localization quality directly depends on the quality of the landmarks and their distribution over the operational area. Our approach shows that if the landmark reduction is done in a correct way, the localization quality does not suffer from the reduction. Figure 5 illustrates the localization error against ground truth measurements. The error value plotted there is the Euclidean distance from the ground truth measurement to the mean of the robot's state estimation. Together with the summary of the estimated trajectory (fig. 6), one can see that the localization error is within the same range no matter whether using landmark reduction or not. The progression of the robot pose uncertainty (eigenvalues from the covariance matrix of the x and the y component) is compared in figure 7. The uncertainties in all three runs are within the same range. This verifies that the localization quality does not necessarily suffer from the landmark reduction.

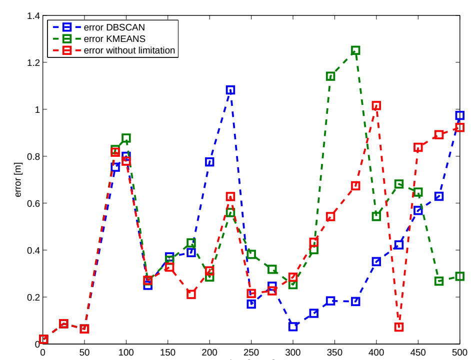

Fig. 5. Comparison of the localization error against ground truth.

B. Landmark coverage quality

The landmark coverage of the environment is, as previously mentioned, a crucial factor for the relocalization capability. If the landmark reduction algorithm removes all landmarks from one area, a later relocalization within this area would be difficult. The two tested clustering algorithms, K-means and DBSCAN, lead to different results in landmark coverage. The landmark reduction using DBSCAN leads to a well distributed landmark coverage over the operational area, as illustrated in figure 8. In contrast thereto, the landmark reduction based on K-means clustering results in a much more dense representation around the later timesteps as illustrated in figure 9. Both figures show the clusters of viewpoint locations. A viewpoint location is the position

Proceedings of the 2nd. Workshop on Omnidirectional Robot Vision
A workshop of the 2010 IEEE International Conference on Robotics and Automation (ICRA2010)
Anchorage, Alaska, USA, May 7, 2010, ISBN 978-88-95872-02-5
pp. 26-31

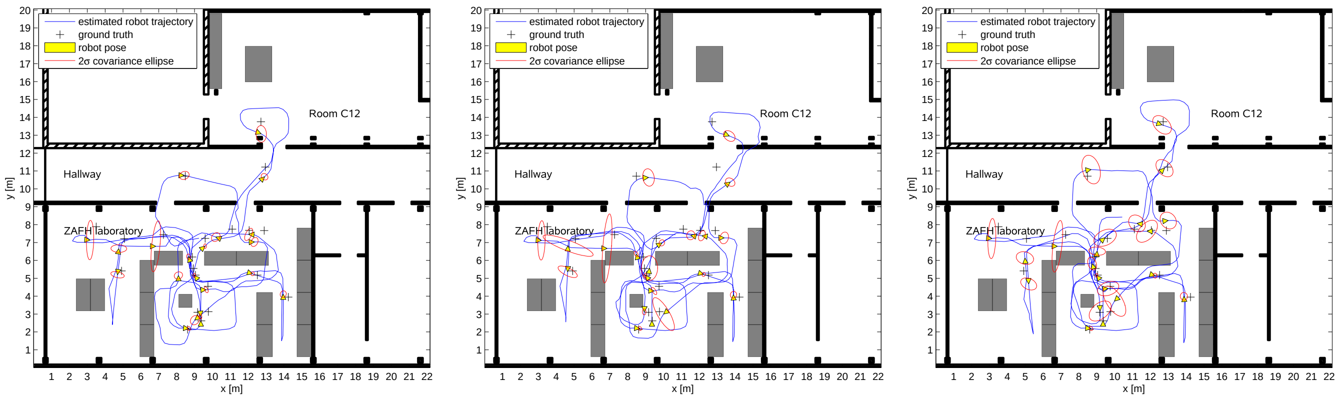

Fig. 6. Estimated trajectory without any landmark limitations (left), K-means clustering with maximum 150 landmarks (middle) and DBSCAN clustering with maximum 150 landmarks (right).

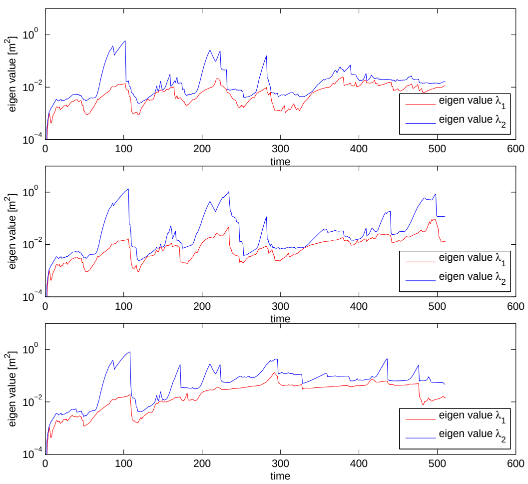

Fig. 7. Eigenvalues of the robot position covariance matrix during the run with restricted number of landmarks. The y-axis is log-scaled. The top figure shows the robot uncertainty without landmark reduction. The other figures illustrate the robot uncertainty during runs with landmark reduction mechanisms (middle figure K-means based mechanism, bottom figure DBSCAN based mechanism).

from where a landmark can be observed. All representatives belonging to a cluster (spatial closeness) are drawn with the same color.

C. Robust Data Association

A robust data association is one of the main components of a robust Visual SLAM approach. In an additional experiment [2] we can show that the presented method, results in a robust data association. During validation of reobservations we rejected approximately 120 wrong associations. In the other case, during the landmark initialization process, we identify approximately 45 *false positive*. For clarification figure 10 shows one example of a *false positive*. Table I shows the results for the evaluated comparison methods. The first method is the test described by Lowe [16]. For Lowe Test we search in the kd-tree these two features where the descriptor vector has the smallest Euclidean distance to the observed feature descriptor. The observed SURF feature of the current image is considered as not matching a known feature if the ratio of the smallest and the second smallest distance value is above a given threshold (value 0.6). This is similar to the SIFT feature descriptor comparison by Lowe [16]. The Euclidean distance to the nearest neighbor and the second nearest neighbor in the database is 0.1487 respectively 0.2872. Thus, the Lowe test results in a match ($0.1487 < 0.2872 * 0.6 \Rightarrow matching = true$).

For the correlation coefficient test we search that feature in the kd-tree where the descriptor vector has the least Euclidean distance to the observed feature descriptor. Then we calculate the correlation coefficient of those descriptors. If the correlation coefficient is above a threshold ($corrCoeff > 0.9$), we assume a match.

In the next test the observed feature descriptor is compared with the descriptor from the kd-tree with the least Euclidean distance by a χ^2 Test. The threshold for the χ^2 Test is set to 0.15.

During our experiment we demonstrate that decisions based solely on information of the image processing step cannot solve the landmark assignment problem. The overall result of all considered methods (see table I) is that the new feature in the right image matches with the previously observed feature with the ID 63 in the left image. However, as can be seen in figure 10, this assignment is wrong. The validation gate is the only mean to reject this assignment. It exploits additional information like the position and the uncertainty of the landmark as available from the EKF. We now can detect with high probability that the assignment in this example is wrong. In all our experiments, the threshold used with the Mahalanobis distance test is set to 0.1015.

VI. CONCLUSIONS

The experimental setting included varying lighting conditions and repeating structures. Despite that, the approach successfully solved the SLAM task in everyday environments even with limited system resources. Thus, the proposed approach successfully addresses the aspect of suitability for daily use as mandatory in service robotics.

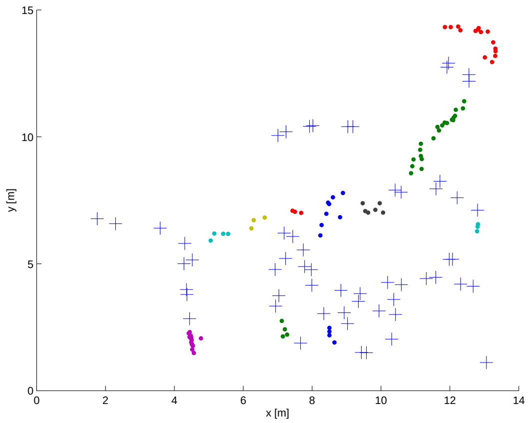

Fig. 8. Landmark clustering using DBSCAN after 509 timesteps.

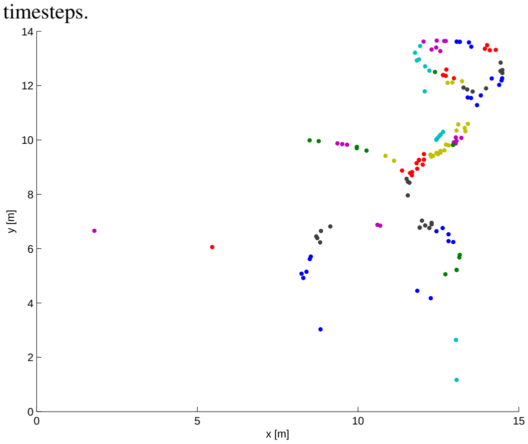

Fig. 9. Landmark clustering using K-means after 509 timesteps.

TABLE I

RESULTS OF SURF FEATURE DESCRIPTOR COMPARISON METHODS

Comparison Method	**Value**	**Threshold**	**Classification**
Lowe	$0.15 < 0.6 * 0.29$	$d_1 < 0.6 * d_2$	matching
Correlation Coefficient	0.9882	> 0.90	matching
χ^2 Test	0.079	< 0.15	matching
Mahalanobis Distance	656.781	< 0.1015	not matching

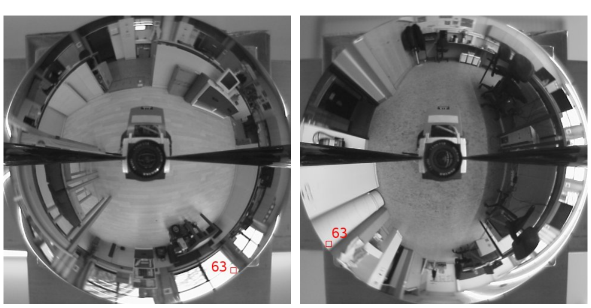

Fig. 10. Based solely on a descriptor comparison we could not distinguish the feature in the left image (timestep 7) from that in the right image. Therefore the feature in the right image (timestep 24) gets the same ID (63) as the feature in the left image.

VII. ACKNOWLEDGMENTS

This work has been conducted within the *ZAFH Servicerobotik* (http://www.zafh-servicerobotik.de/en). The authors gratefully acknowledge the research grants of state of Baden-Württemberg and the European Union.

REFERENCES

[1] S. Hochdorfer, M. Lutz, and C. Schlegel, "Lifelong Localization of a Mobile Service-Robot in Everyday Indoor Environments Using Omnidirectional Vision," in *IEEE International Conference on Technologies for Practical Robot Applications (TePRA)*, Nov. 2009, pp. 161 –166.

[2] S. Hochdorfer and C. Schlegel, "Towards a robust Visual SLAM Approach: Addressing the Challenge of life-long Operation," in *IEEE International Conference on Advanced Robotics (ICAR 2009)*, 2009.

[3] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded up robust features," in *9th European Conference on Computer Vision*, Graz Austria, May 2006.

[4] T. Bailey, "Constrained initialisation for bearing-only SLAM," *IEEE International Conference on Robotics and Automation (ICRA)*, vol. 2, pp. 1966–1971, Sept. 2003.

[5] C. Schlegel and S. Hochdorfer, "Bearing-only slam with an omnicam - an experimental evaluation for service robotics applications," in *Autonome Mobile Systeme (AMS) 2005, 19. Fachgespräch, Stuttgart, 2005*, ser. Springer, Informatik Aktuell. Springer, 2005, pp. 99–106.

[6] S. Hochdorfer and C. Schlegel, "Bearing-Only SLAM with an Omnicam: Robust Selection of SIFT Features for Service Robots," in *Autonome Mobile Systeme (AMS) 2007, 20. Fachgespräch, Kaiserslautern, 2007*, ser. Springer, Informatik Aktuell. Springer, 2007, pp. 8–14.

[7] H. Strasdat, C. Stachniss, M. Bennewitz, and W. Burgard, "Visual Bearing-Only Simultaneous Localization and Mapping with Improved Feature Matching," in *AMS*, ser. Informatik Aktuell, K. Berns and T. Luksch, Eds. Springer, 2007, pp. 15–21.

[8] H. Andreasson, T. Duckett, and A. J. Lilienthal, "Mini-slam: Minimalistic visual slam in large-scale environments based on a new interpretation of image similarity," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2007, pp. 4096–4101.

[9] S. Hochdorfer and C. Schlegel, "Landmark rating and selection according to localization coverage: Addressing the challenge of lifelong operation of SLAM in service robots," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, no. 0, 2009.

[10] G. A. F. Seber, *Multivariate Observations*. New York: Wiley, 1984.

[11] H. Spath, *Cluster Dissection and Analysis: Theory, FORTRAN Programs, Examples*. Halsted Press,New York, 226 pp., 1985.

[12] G. Dissanayake, H. F. Durrant-Whyte, and T. Bailey, "A Computationally Efficient Solution to the Simultaneous Localisation and Map Building (SLAM) Problem," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2000, pp. 1009–1014.

[13] H. Strasdat, C. Stachniss, and W. Burgard, "Which landmark is useful? learning selection policies for navigation in unknown environments," in *IEEE International Conference on Robotics and Automation (ICRA), Kobe, Japan*, 2009, to appear.

[14] J. Neira and J. Tardós, "Data association in stochastic mapping using the joint compatibility test," *IEEE Transactions on Robotics and Automation*, vol. Vol. 17, no. 6, pp. pp. 890 – 897, December 2001.

[15] J. L. Bentley, "Multidimensional binary search trees used for associative searching," *Commun. ACM*, vol. 18, no. 9, pp. 509–517, 1975.

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

[17] K.-M. Kiang, R. Willgoss, and A. Blair, "Distinctness analysis on natural landmark descriptors," in *Field and Service Robotics*, ser. Springer Tracts in Advanced Robotics, P. I. Corke and S. Sukkarieh, Eds., vol. 25. Springer, 2006, pp. 67–78.

[18] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," in *Proc. 2nd Int. Conf. on Knowledge Discovery and Data Mining (KDD'96)*, E. Simoudis, J. Han, and U. Fayyad, Eds. AAAI Press, 1996, pp. 226–231.

[19] T. N. Tran, R. Wehrens, and L. M. Buydens, "Clustering multispectral images: a tutorial," *Chemometrics and Intelligent Laboratory Systems*, vol. 77, pp. 3–17, 2005.