A map for mobile robots consisting of a 3D model with augmented salient image features

Friedrich Fraundorfer

Institute for Computer Graphics and Vision, Graz University of Technology
fraunfri@icg.tu-graz.ac.at

Abstract:
In this work we propose a map for mobile robots which is suited for simultaneous localization and map building (SLAM). A 3D model with augmented salient appearance based image features alleviates the deficiencies of a bare 3D model in localization. The augmented appearance based features allow to distinguish between similar locations which may show little difference in the 3D model, eg. similar doors in a hallway. A new method to find distinctive landmarks is introduced as a key feature of the proposed map.

1 Introduction

Simultaneous localization and map building (SLAM) is a key feature for mobile robots. While SLAM has already been suitable developed for ultrasonic sensors or laser range finders, this task has not been completed for vision systems. The SLAM-concept seems to be very natural, because localization and map building depend strongly on each other. Exploring an unknown environment and building a map of the environment needs the ability of self-localization. On the other hand the methods for localization strongly depend on the kind of map which is built from the environment. In addition, it has to be considered how well the map supports the robot in navigation and route planning. To facilitate navigation and route planning we use a reconstructed 3D model of the environment as a map for a mobile robot. To reduce ambiguities and allow fast access the bare 3D model is augmented with texture information. In the following sections this idea will be described in more detail.

2 Overview of world representations and localization methods

One of the first proposed maps suitable for mobile robots was the certainty grid (or occupancy map) by Moravec and Elfes [9]. Such a map provides a floor plan divided into grid cells. Each cell gets the status occupied or free. A certainty value for each cell states how certain the robot is about the state of the cell. Such an occupancy map was used by Murray and Little in [10]. Another approach is the use of 3D models. These vary from simple line models where only vertical lines are used (as

in [1] and [6]) up to complex CAD models (as in [2]). Building a map from landmarks also contains 3D world coordinates but often they do not give a dense description of the environment. Such an approach using SIFT-features is described by Se et al. in [12]. Appearance based maps of the environment simply store all images in a compressed representation. Localization can be done by matching the images by means of an appearance based method like PCA. Such an approach is given by Matsumoto et al. in [8] or by Jogan and Leonardis in [4]. Localization using 3D models (e.g. line or CAD models) can be done by matching a generated view from the map with the view which is visible by the robot. The generated view consists usually of horizontal and vertical lines only (or some other primitive geometric objects) and the view of the robot is represented by lines extracted by an edge detection algorithm. This method seems to have some deficiencies in an environment were similar structures occur, for instance a hallway with lots of similar doors. Or if you go into a bigger scale, a city with lots of similar crossings. In this case the line information is not enough. Figure 1a shows a 3D model of a part of the hallway of our institute. There are lots of similar doors. One may consider a robot located in-front of a door and viewing only the door itself. Using the outline of the door only (gained by edge detection) the robot cannot localize itself because of the similarity of the doors. Some additional information should be used in addition to 3D model data. Figures 1b,c show two different doors like seen from a robot. In the 3D model they do not differ, but one of them has a poster sticked to it. When extracting line features to match the image with the geometric model of the door, this poster information gets lost. The robot cannot distinguish the two doors.

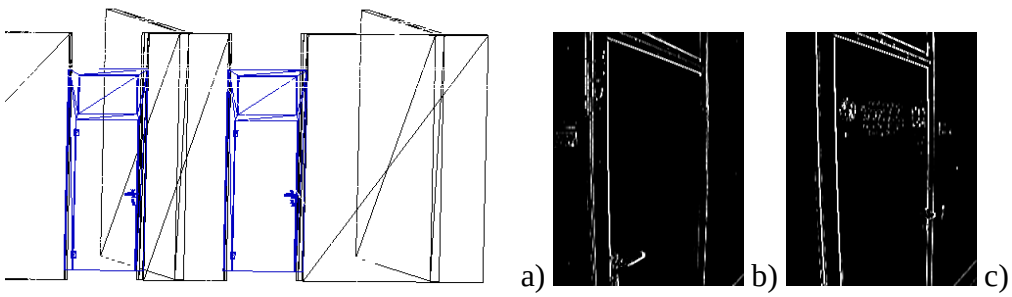

Figure 1: a) Visualization of the 3D model which is used as world representation. It shows a part of the institute's hallway. b)c) Images of the two doors after edge detection. The images differ only by a poster sticked to the right door. Using only the outline of the door it is impossible to distinguish them.

Appearance based maps store all the images of the robots environment. On localization the robot compares the new image with the images in the map. If a similar image is found the robot then knows the location and the heading from the previous run. One obvious deficiency of this method is the accuracy of the computed robot position. The position of the robot will be interpolated from the nearest positions the robot has visited so far. The accuracy depends on the density of the stored images and this is limited by the available memory in the mobile robot. Therefore highly accurate positions have to be computed by other means.

3 A map consisting of a 3D model with augmented salient image features

The deficiencies of a world representation using a 3D model and of an appearance based map could be eliminated by combining both methods. This will lead to a coarse map built of appearance based features which allows to get an accurate position using 3D model matching. The appearance based features should be regions which are distinctive and invariant against viewpoint and scale change. In detail the suggested map should consist of a 3D world model described by vertical planes, which allows the representation of walls. Surface patches which were found to be distinctive and invariant against viewpoint and scale change are used to give the appearance based information and are integrated into the model as a kind of texture. There is no need for a complete texture description because only distinctive and invariant surface patches are useful. Such distinctive or salient surface patches should be detected automatically. Figure 2 shows the visualization of the proposed map. Surface patches integrated into the 3D model allow to distinguish two doors which are similar if only the 3D model is considered. For localization in the proposed map we have to solve three problems:

- Detection of distinctive regions for use as landmarks
- Matching of image regions using an invariant description
- Landmark learning and localization

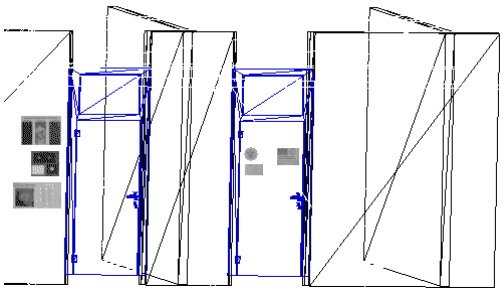

Figure 2: The proposed map. A 3D model as world representation with augmented surface patches which allow appearance based coarse localization.

3.1 Detection of distinctive regions

An example for landmarks are SIFT features as described by Se et al. in [12] which are invariant to image translation, scaling and rotation. Image regions with characteristic local maxima of edge density can also be used as landmarks which is shown by Sim and Dudek in [13]. Ayala et al. (in [16]) use planar posters for robot localization. Their approach is focused on rectangular posters which were learned manually and were found afterwards by edge extraction. Our approach finds salient surface patches of general structure as landmarks. For this an algorithm proposed by Kadir

and Brady (in [5]) is used. Their algorithm detects salient regions by calculating the entropy value of a support region over multiple scales. Salient regions are characterized by a peaked entropy value over a narrow range of scales and not over the whole scale range. Experiments by Kadir and Brady show that the algorithm finds the same salient regions in images taken from different viewpoints, under rotation or re-scaling.

3.2 Matching of salient image regions

For localization in the proposed map a reliable method to match image regions is needed. A new salient region is compared with a set of learned salient regions (with known 3D coordinates) for localization. If a match is found the actual coarse position is known. Some promising methods for matching regions are listed in the following. Histogram matching is used by Ulrich and Nourbakhsh [15] for localizing a mobile robot. Six color histograms are used to match an acquired image with a set of learned omnidirectional images. While this approach seems to work quite well in their application, our tests with histogram matching have not been satisfying with small salient regions. Furthermore we would like to use grey-scale images only. Another approach was proposed by Schmid and Mohr [11]. They describe the neighborhood of an image point by the set of its derivatives. From this so-called "local jets" differential invariants are computed to generate a feature vector which describes the region around the image point. This method was used for image retrieval. In [3] Mindru et al. propose invariant features for the recognition of planar color patterns. The features are based on moments of powers of the intensities in the individual color bands and combinations thereof. These features are also used by Tuytelaars and Van Gool in [14] for wide baseline stereo matching. A work from Sim and Dudek [13] about mobile robot localization describes an approach using principal component analysis. In a learning step prototypes are built using sets of small image regions (landmarks) viewed from different positions. Localization then works by selecting the prototype with least Euclidean distance in the subspace from the actual taken image. In a work from Manjunath and Ma [7] Gabor wavelet features are used for image retrieval. A Gabor filter bank with different scales and orientations extracts texture features from image regions which are used to find similar images in a database. In this work the Gabor wavelet features proposed in [7] are used for matching the salient regions. The Gabor filter bank uses 4 scales and 6 orientations. To represent a region the mean value μ_{mn} and the standard deviation σ_{mn} of the magnitude of the transform coefficients are used. A feature vector is constructed using μ_{mn} and σ_{mn} as feature components. With 4 scales and 6 orientations a feature vector consist of 48 entries. To define the distance between two image regions $d(i,j)$ in the feature space following distance measure is used.

$$d(i,j) = \sum_m \sum_n \left| \frac{\mu_{mn}^{(i)} - \mu_{mn}^{(j)}}{\alpha(\mu_{mn})} \right| + \left| \frac{\sigma_{mn}^{(i)} - \sigma_{mn}^{(j)}}{\alpha(\sigma_{mn})} \right|$$

$\alpha(\mu_{mn})$ and $\alpha(\sigma_{mn})$ are the standard deviations of the entries of the feature vector and are used to normalize the individual feature components.

3.3 Learning and localizing

Only really distinctive and reliable re-recognizable regions should be learned as landmarks. For this the image regions have to fulfill two criteria. The first criteria is that salient regions should be learned only if they are distinctive against all other regions found in the image. For this the regions found in the image have to be matched against themselves and the ones which show only a small distance in feature space are discarded. This step will remove all salient regions with similar texture properties and therefore similar feature vectors. The second criteria is for obtaining reliable landmarks. This means a landmark should be learned only if it can be found in images from other viewpoints too. Therefore a landmark is learned only if it is visible in three subsequent images of the image sequences of the moving robot. These measures lead to a set of landmarks which can be used as prototypes for minimum distance classification which utilizes localization. For localization at first salient regions are calculated in the actual image. Then similar regions are discarded to gain only distinctive regions. The resulting regions are then classified using a minimum distance algorithm on the learned set of landmarks. The region with the minimum distance to a landmark in feature space will be used to obtain a coarse position of the robot.

4 Experiments

Experiments are carried out on an image sequence of the institutes hallway. The image sequence contains 38 grey-scale images. To find salient regions a coarse resolution of 240x320 pixels is used. Though this resolution is low it is sufficient because we are interested in rather big image regions. The image patches are extracted from higher resolution images of 480x640 pixels so that they show enough detail for Gabor filter matching. It will be shown that using the proposed augmented salient features it is possible to distinguish two similar doors differing only by the posters on them.

4.1 Finding salient regions

To find salient regions in an image the entropy is calculated for 5 scales with a diameter of 20, 26, 32, 38 and 44 pixels. Only regions which are supported by 30 surrounding peaks in entropy will be selected. Figure 3 shows salient regions found in images of the doors scene.

4.2 Distinctive Regions

Figure 4 shows salient regions extracted from an image of the door sequence. One may notice several very similar image patches which are not suited to be used as landmarks. Figure 5 shows the salient regions from figure 4 which are distinctive enough to be used as landmarks. The similar regions show between themselves an average distance in feature space of 17.0 while the set of distinctive regions shows an average distance of 58.3.

Figure 3: a,b) Two different parts of the doors scene image sequence. The bigger images show the detected salient regions and the small image patches are the extracted landmarks.

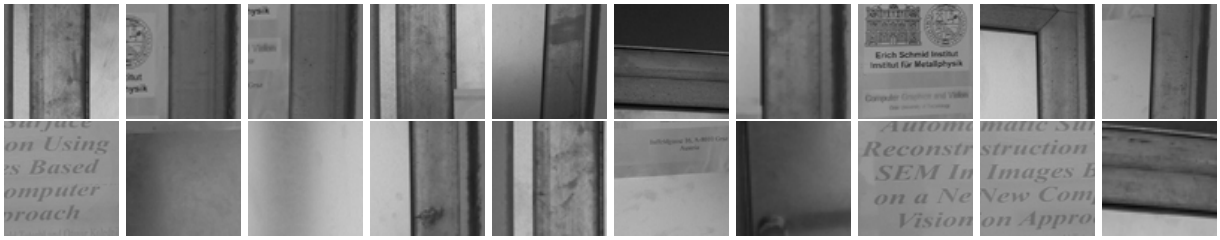

Figure 4: All extracted salient regions of an image of the door sequence. Some of the image patches look very similar which makes them unsuitable as landmarks.

4.3 Reliable landmarks

To provide stable and reliable landmarks image patches will be learned only if the occur in three subsequent views. Landmarks which only appear in a single view are not useful because it is unlikely that the same position will be reached again. Figure 3 shows the image patches extracted from three subsequent images of the doors scene which fulfill all criteria to act as an landmark. The average distance of matching image patches in feature space is 12.5 (max. 19.98). This can be separated well from the non-matches with an average distance of 135.2 (min. 78.2).

4.4 Localization

From three subsequent images landmarks based on the previous given methods were extracted. Figure 3a,b) shows the used images and the extracted landmarks for the two doors. In the case of this

Figure 5: Distinctive patches of the set from figure 4. They are suitable as landmarks.

Figure 6: a,b) Images for which the location should be determined. The smaller images are the extracted distinctive regions. The location can be determined if one of the distinctive regions can be matched to one of the learned landmarks.

scenes from about 10-20 salient regions per image only one region has been found suitable as a landmark. Figure 6a,b) shows the images and the distinctive salient regions for which the location should be determined. A minimum distance algorithm finds the smallest distance between the third landmark of 3a) and the third distinctive region of 6a). This classification is correct. For the localization of 6b) the minimum distance is between the first landmark of 3b) and the first distinctive region of 6b) which is also a correct classification.

5 Conclusion

In this paper a map for mobile robots consisting of a 3D model with augmented salient image features has been proposed which allows for simultaneous localization and map building (SLAM). The proposed map was motivated by difficulties in localization with a bare 3D model which were discussed in section 2. Experiments showed that a coarse localization is possible using landmarks extracted on the basis of salient regions. In future work the robustness of this method will be investigated. We also want to develop a new method for image region matching based on salient regions

which should be invariant against rotation and scaling. The idea is to use geometric relations of small salient regions inside of image patches to describe the region. We also want to improve the landmark extraction. In addition to a small distance in feature space based on texture feature only geometric relations should be checked. For this we intend to estimate the trifocal geometry of three subsequent views and check if landmark candidates fulfill the trifocal constraints.

References

[1] N.J. Ayache and O.D. Faugeras. Maintaining representations of the environment of a mobile robot. *IEEE Transactions on Robotics and Automation*, 5(6):804–819, December 1989.

[2] M. Ayromlou, C. Beltran, A. Gasteratos, O. Madsen, W. Ponweiser, and M. Vincze. Robvision - vision based navigation for mobile robots. In *Proc. OAGM 2001*, pages 25–32, 2001.

[3] T. Moons F. Mindru and L. Van Gool. Recognizing color patterns irrespective of viewpoint and illumination. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 368–373, 1999.

[4] M. Jogan and A. Leonardis. Robust localization using panoramic view-based recognition. In *Proc. ICPR00*, volume 4, pages 136–139, 2000.

[5] T. Kadir and M. Brady. Saliency, scale and image description. *International Journal of Computer Vision*, 45(2):83–105, November 2001.

[6] A. Kosaka and A.C. Kak. Fast vision-guided mobile robot navigation using model-based reasoning and prediction of uncertainties. *CVGIP*, 56(3):271–329, November 1992.

[7] B.S. Manjunath and W.Y. Ma. Texture features for browsing and retrieval of image data. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 18(8):837–842, August 1996.

[8] Yoshio Matsumoto, Kazunori Ikeda, Masayuki Inaba, and Hirochika Inoue. Visual navigation using omnidirectional view sequence. In *Proc. of IEEE Int. Conf. on Intelligent Robots and Systems*, pages 317–322, 1999.

[9] H.P. Moravec and A. Elfes. High resolution maps from wide angle sonar. In *Proc. IEEE International Conference on Intelligent Robots and Systems*, pages 116–121, 1985.

[10] Don Murray and James J. Little. Using real-time stereo vision for mobile robot navigation. *Autonomous Robots*, 8(2):161–171, 2000.

[11] Cordelia Schmid and Roger Mohr. Local grayvalue invariants for image retrieval. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(5):530–535, May 1997.

[12] Stephen Se, David Lowe, and Jim Little. Local and global localization for mobile robots using visual landmarks. In *Proceedings of IEEE International Conference on Intelligent Robots and Systems*, pages 414–420, October 2001.

[13] Robert Sim and Gregory Dudek. Learning and evaluating visual features for pose estimation. In *ICCV*, pages 1217–1222, 1999.

[14] T. Tuytelaars and L. Van Gool. Wide baseline stereo matching based on local, affinely invariant regions. In *British Machine Vision Conference BMVC'2000*, September 2000.

[15] Iwan Ulrich and Illah Nourbakhsh. Appearance-based place recognition for topological localization. In *Proc. IEEE International Conference on Robotics and Automation*, pages 1023–1029, April 2000.

[16] V.Ayala, J.B.Hayet, F.Lerasle, and M.Devy. Visual localization of a mobile robot in indoor environments using planar landmarks. In *IEEE Int. Conf. on Intelligent Robots and Systems*, pages 275–280, November 2000.