New approach in solving the kidnapped robot problem

Eng., Andras Majdik, Technical University of Cluj-Napoca, Romania
Eng., Mircea Popa, Technical University of Cluj-Napoca, Romania
Eng., Levente Tamas, Technical University of Cluj-Napoca, Romania
Eng., Istvan Szoke, Technical University of Cluj-Napoca, Romania
Prof. Dr. Eng., Gheorghe Lazea, Technical University of Cluj-Napoca, Romania

Summary

We present a solution for the global localization of a mobile robot which disposes of a prior map and is waked up in an unknown position. This case is known in robotics as the kidnapped robot problem. The algorithm is based on the adaptation of the 2D SURF (Speeded Up Robust Features) image features to 3D landmarks of the environment. In this way the prior map is built, containing only 3D Surf landmarks. When the robot wakes up in an unknown location the robot starts to search after these landmarks, which are compared to the priory known ones based on the SURF descriptors. To estimate the displacement between the online and offline landmarks the RANSAC random sampling method is used. This method also has the advantage that the falsely detected and associated landmarks are eliminated. The experiments were carried out in an indoor office environment.

1 Introduction

This paper deals with the problem of the kidnapped robot, which is a special global localization issue in the field of mobile robot localization and navigation. Our solution to solve this place recognition task is based on the stereo visual system.

The paper structure is as follows. First the problem is appointed and the state of the art is described in the area of robotic visual perception. In the second chapter our solution is presented and a brief description is given about the theoretical aspects on which our solution relies.

The results of the developed system and the experiments that were carried out in an indoor scenario of an office building are shown in the third chapter. In the end of the paper the conclusions and some ideas about the further work are formulated.

1.1 Problem definition

The kidnapped robot problem is a special global localization issue in mobile robotics. It is a place recognition problem in which the robot disposes of a prior (known) map. The challenge is that the robot has to determine his correct position when it is kidnapped at one location on the map and is waked up in another unknown position. To be able to accomplish this task the robot has to be able to recognize the unknown position where it is waked up in the prior map.

1.2 The state of the art

In the recent years many works were published in the field of robotics concerning the navigation and mapping based on visual systems. The vision system is one of the most popular sensors in mobile robotic applications because at relatively low price the vision system provides an abundance of information about the environment. But this large amount of information needs to be processed and organized in an efficient and useful way to be meaningful data for the robots perception.

Related work regarding place recognition using visual features is reported in [1] where the problem of appearance-only simultaneous localization and mapping (SLAM) for very large scale outdoor scenario is addressed.

Another interesting application is described in [2], which deals with the exploration and mapping of underwater ocean floors. These applications can be categorized as been pose-based approaches to the localization problem in which the pose is recognized by detecting the same features on images taken at the same robot locations at different time. In this paper we are experimenting with the use of the image features as 3D landmarks, similar approaches which are using the scale-invariant feature transform invariant can be found [3], [4]. In [5] is given a sub mapping technique to manage and limit the number of features to perform robust and scalable SLAM.

ISBN 978-3-8007-3273-9

2 Theoretical background

In this chapter the proposed algorithm and the theoretical aspects are described that were used to develop the method, respectively the experiments presented in this paper.

2.1 Proposed algorithm

In **Figure 1** is shown the diagram of the proposed algorithm, which solves the kidnapped robot problem. In the first step the algorithm needs a prior map of the environment. The offline, prior map is a 3D feature map containing all the 3D image features collected by the robot while it navigates throughout the environment. For the visual perception of the mobile robot we used the SURF image feature detector and descriptor. By detecting identical features on both of the images of the stereo pairs 3D features can be computed. Details about the SURF features and the stereo geometry of the visual system are given later on. If the robot disposes of this prior map and is waked up in an unknown position it starts to capture stereo image pairs on which the features are computed similarly to the offline case. At every position the online 3D feature point cloud is searched based on the SURF descriptor in the saved map. For the features found a point clouds is selected from the offline map. To estimate the displacement between these two feature clouds a variant of the RANSAC algorithm is used. The features that are considered to be outliers are excluded from the online point cloud. The algorithm is repeated iteratively.

2.2 SURF for interest point detection

To solve the kidnapped robot problem the recognition of the landmarks of the environment is a key issue. Our algorithm uses image features detector and descriptor called SURF. SURF is an abbreviation from Speeded Up Robust Features; it is a method, known in the field of computer vision to detect interest points on images.

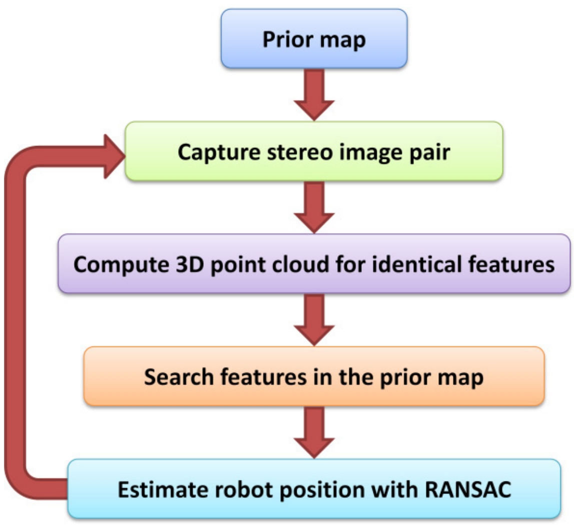

Figure 1 The flow diagram of the proposed algorithm.

SURF was first introduced by [6], the advantage of this algorithm is the computational speed compared with other performing image feature detectors like Scale-Invariant Feature Transform (SIFT) presented by [7], maintaining approximately the same or even better performances regarding: repeatability, distinctiveness and robustness. In [8] is concluded that SURF is one of the most suitable descriptor for visual simultaneous localization and mapping applications.

The algorithm and the performances are largely discussed in [9] and [6], whereas hereby only a short overview is presented. The main reason why the algorithm decreases the computational time is because SURF uses integral images [10] as intermediate representation, in this way the sum of intensities over any upright, rectangular region is calculated with only four additions.

The SURF algorithm is based on the Fast-Hessian detector, which computes the determinant of the Hessian matrix:

$$\mathrm{H}(p,\sigma) = \begin{bmatrix} \mathrm{L_{xx}}(p,\sigma) & \mathrm{L_{xy}}(p,\sigma) \\ \mathrm{L_{xy}}(p,\sigma) & \mathrm{L_{yy}}(p,\sigma) \end{bmatrix} \qquad (1)$$

Where with $\mathrm{L_{xx}}(p,\sigma)$ is denoted the convolution of the second order Gaussian derivative $\partial^2 g(\sigma)/\partial x^2$ of the image at point $p(x,y)$ and scale σ in the x direction. $\mathrm{L}_{xy}(p,\sigma)$ and $\mathrm{L}_{yy}(p,\sigma)$ are calculated similarly, these derivatives are known as Laplacian of Gaussian.

Another reason why the SURF algorithm performs fast is that it approximates the Laplacian of Gaussian with a box filter representation. The box filter allows a performance increase in time when they are computed on integral images.

2.3 Stereo depth estimation

This section focuses on the recovering of the 3D information from 2D images in order to estimate distances of the features from the robot. To compute the depth of an image feature a stereo vision camera system is used. A simple representation of the stereo sensor geometry of the system is given on **Figure 2**, where $\mathrm{C_l}$ and $\mathrm{C_r}$ are the centers of the

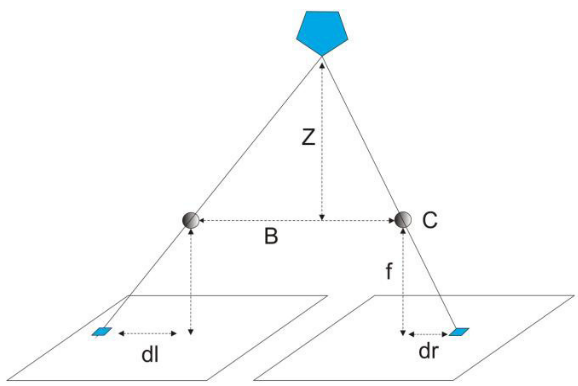

Figure 2 Stereo geometry of the vision system.

ISBN 978-3-8007-3273-9

lenses for the right and left camera, f is the focal length, B is the baseline and Z is depth in the coordinate system. If the same feature associated to an object (pentagon in the figure) is detected in the left and right images by simple geometrical deduction the depth information Z can be obtained in the following way:

$$\mathrm{Z} = \frac{\mathrm{f} \times \mathrm{B}}{\mathrm{d}} \tag{2}$$

$$\mathrm{d} = |\mathrm{dl} - \mathrm{dr}| \tag{3}$$

Where d represents the disparity, the disparity can be defined as the difference between the coordinates (d_l and d_r) of the same feature in the left and right image.

The focal length and the baseline are structural constants of the vision system. Once the depth is determined, to compute the real world X and Y coordinate of each feature is a trivial task.

2.4 RANSAC for robot position estimation

The RANSAC (RANdom Sample And Consensus) is a popular algorithm for parameter estimation from data affected by noise and corrupted by outliers, introduced by [11]. By outliers are meant the points which do not fit the model instantiated by a small set of points called minimal sample set (MSS). The methods based on this technique can deal with data sets that are more than 50 % of outlier contaminated [12].

In a mathematical form the problem that we tried to solve is to find the rigid transformation: rotation and translation $g(R,T)$ between the two selected feature point clouds using the RANSAC estimation technique. We consider the model point set M is to be the features selected from the prior map and the scene point set S is the features collected by the robot after it is waked up in an unknown position (online).

The adopted RANSAC based algorithm includes the following steps, adopted from [13]:

a) Select the scene control points:
   - randomly select a couple of points s_1 and s_2 in the scene point set S and compute the distance

$$d_S = \|s_1 - s_2\| \tag{4}$$

b) Select the model control points:
   - randomly select (m_1, m_2) from the model point set M with the constraint $d_M = d(m_1, m_2) \cong d_S$, such that:

$$|d_S - d_M| \leq d_{th} \tag{5}$$

   where d_{th} is a preset error threshold and:

$$\varepsilon = |d_S - d_M| \tag{6}$$

c) Estimate the model parameters:
   - find the transformation parameters T and R
   - apply the transformation to the scene set

$$S' = g(R,T)S \tag{7}$$

d) Verify the model:
   - count the number N_{in} of points S' consistent with the model M (inliers)
   - if the maximum iteration number is not exceeded, go back to the a)
   - finally select the hypothesis with the largest number of inliers and solve the least mean squares problem only for inliers:

$$(R,T)_f = \arg min_{(R,T)} \sum_{i=1}^{N_{im}} \|m_i - g(R,T)s_i\|^2 \tag{8}$$

In [14] it is pointed out that if the error threshold d_{th} is to high, the algorithm can return poor results.

$$C = \sum_i \rho(\varepsilon_i) \tag{9}$$

$$\rho(\varepsilon) = \begin{cases} 0, & inlier \\ constant, & outlier \end{cases} \tag{10}$$

Starting from this idea, the RANSAC can be seen as an optimization algorithm for the cost function, the original algorithm is redefined by considering the estimator as part of the M-estimator family, being known as MSAC [14].

$$\rho_m(\varepsilon) = \begin{cases} \varepsilon, & inlier \\ d_{th}, & outlier \end{cases} \tag{11}$$

This way the algorithm provides better performances as shown in [15]. In the development of our algorithm we use this variant of RANSAC.

In the literature were reported novel variants of the algorithm by [16]. In the paper a new optimal sequential strategy for randomized evaluation of model quality is introduced, based on Wald's theory of sequential decision making.

In [17] a new approach is presented to choose the most likely hypothesis set, when prior probabilities can be estimated. The aim of these novel techniques is to speed up the computational time and improve the algorithm performances. In our work this field of advanced RANSAC algorithms is subject of further development.

3 Experiments

In this segment are presented the experiments, which were carried out in the laboratory and are based on the theory presented above. For all experiments a Pioneer 3 All-Terrain mo-bile robot, made by Mobile Robots Inc. and a Bumblebee2 stereo camera, made by Point Grey Research Inc were used.

ISBN 978-3-8007-3973-9

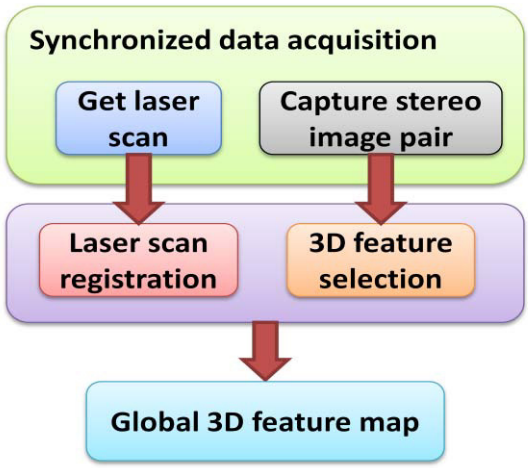

Figure 3 Diagram of the offline mapping algorithm.

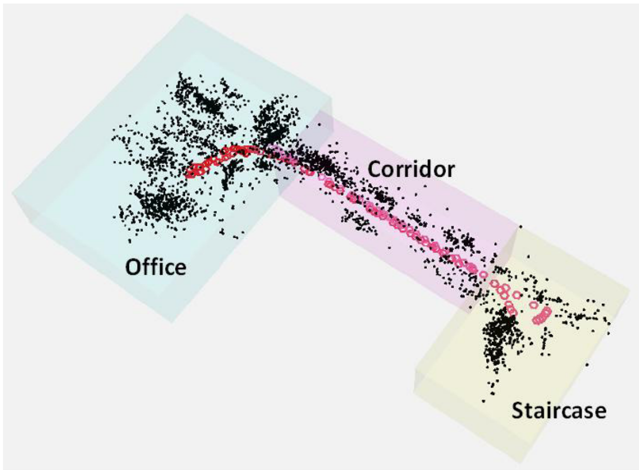

Figure 5 Offline 3D feature map

3.1 Building the offline map

In **Figure 3** is shown the diagram of the offline mapping algorithm. The visual data and the laser scans need to be taken synchronously.

After the data acquisition is done form the displacement of the consecutive laser scans the robots displacement (translation and rotation) is computed with a standard Iterative Closest Point algorithm [18] between two consecutive scans. Because the test environment is relatively small a more complex scan registration method is not needed like in [19], where the problem of mapping a large mine is addressed.

In the mean time another task is to detect the 2D SURF features on both images of the stereo pair and calculate the local 3D feature map using the stereo geometry of the system presented in the previous chapter.

In every robotic application it is desirable to limit the amount of data to decrease computational time as much as possible. For this reason the prior map shouldn't be overpopulate, but in the mean time the distinctiveness of the data is important.

Therefore the 3D image features that are already in the global map or the similar ones to these are filtered out based on the SURF descriptor.

In the final step the local feature point cloud coordinate system is updated with the global position of the robot computed by the laser scan registration algorithm and the new distinctive features are recorded in the global 3D feature map.

The offline, prior map is built iteratively with the presented algorithm by navigating the robot in the test environment of an indoor office building.

A selected image sequence can be seen on **Figure 4** from the 155 stereo pair image sequences that were taken in the office, corridor and staircase of our laboratory.

The obtained 3D feature map is plotted on **Figure 5**, where with black spheres are represent the 3D feature landmarks obtained in the whole test environment, divided in the three main rooms. With the red disk are denoted the positions of the robot, where the stereo image pares were taken. The origin of the map, the first robot location is the dark red one. The fade color of the disks shows the path travelled by the robot.

Figure 4 Selected images from the offline data set.

Figure 6 Images taken online at the four robot position

ISBN 978-3-8007-3273-9

3.2 Solving the kidnapped robot problem

The diagram and the flow of the proposed algorithm were described in the second chapter. The results of our tests are shown on **Figure 7**. With magenta is represented the 3D landmark point cloud, detected online from the same SURF features of the stereo image pairs captured in four consecutive robot positions after the robot is kidnapped and it is wake up. These positions are represented with the magenta triangles and the right images of the stereo pairs captured are shown in **Figure 6**. The robot initially suspects that its position is in the origin of the map.

For the landmarks gathered online the corresponding 3D SURF features (red dots) are selected from the offline map (black dots) based on the SURF descriptors.

In the next step of our method by applying the MSAC algorithm between these two selected point clouds the falsely associated landmarks are eliminated and the absolute position (blue triangles) of the robot is computed by calculating the rotation and translation between the online and offline data sets according to the theory presented.

In Figure 7 is shown that the algorithm works well and the robot recognizes his global position which is in the end of the corridor and not in the office (origin of the map), so the landmarks detected online (blue dots) are in facts those ones from the end of the corridor in the offline map as can be seen from the images in Figure 6. Thus the robot found its own correct position.

Even if the robots navigation field is disturbed by the presents of a dynamic objects (person in the current case) after the robot is kidnapped the algorithm manages to recognize the robots position. Hence we can conclude that the algorithm developed is a robust one.

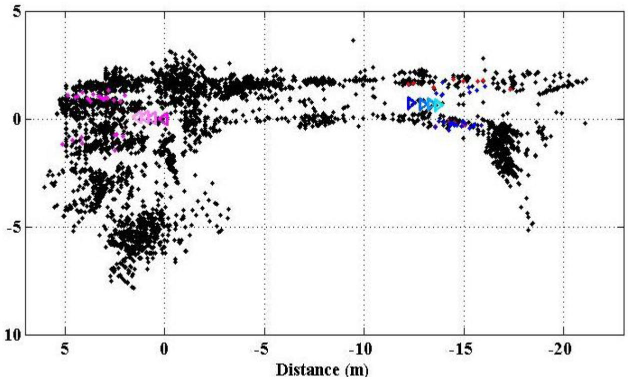

Figure 7 Result of the proposed algorithm after four steps (top view)

4 Conclusions and further work

This paper presents an algorithm which solves the kidnapped robot problem in an indoor office environment. Our solution relies on the SURF image features descriptor and detector. The image features are transformed to be 3D landmarks of the environment using the geometry of a stereo vision system. After the robot is kidnapped and it is waked up in an unknown position, starts to search 3D feature landmark. The found landmarks are searched in the prior map to recognize its position. The robot location is computed with the use of the MSAC variant of the random sample and consensus algorithm.

The proposed algorithm is robust because even if the robot navigation field is disturbed by the presents of dynamic objects the algorithm manages to recognize the robots position. The features detected online associated to dynamic objects are not found in the prior map. If a falsely association is done this is eliminated by the use of the MSAC step because these features are considered to be outliers and are eliminated from the point cloud gather online. The system is fast and can be computed during the robots navigation because instead of using dens stereo point clouds to recognize the robots position only the best points are selected from the point of view of repeatability, distinctiveness and robustness by the use of the image features.

We intend to develop a system to perform simultaneous localization and mapping, based on the presented feature landmarks. In this way the robot will not need the prior mapping step for its self localization.

Another interest of our research is focused on the possibility to extend the system to a larger outdoor urban environment in order to solve more complex tasks in the field of urban city explorer robots. To be able to accomplish this we intend to develop a more complex localization algorithm based on multisensory information.

5 Acknowledgment

This work was done within the PRODOC (Project of Doctoral Studies Development in Advanced Technologies) project at the Technical University of Cluj-Napoca, financially supported by European Social Fund, contract No.: POSDRU/6/1.5/S/5. The publication of the paper is supported by Eurotrans foundation, contract No.: 09.V.2.0124.

6 Literature

[1] Cummins, M., Newman, P.: Highly scalable appearance-only SLAM – FAB - MAP 2.0, Proceedings of Robotics: Science and Systems, 2009

[2] Mahon, I., et al.: efficient view-based SLAM using visual loop closures, IEEE Transactions on Robotics, Vol. 24, pp. 1002-1014, 2008

[3] Se, S., Lowe, D. G., Little, J.: Global localization using distinctive visual features, International Conference on Intelligent Robots and Systems, pp. 226-231, 2002

[4] Se, S., Lowe, D. G., Little, J.: Vision-based global localization and mapping for mobile robots, IEEE Transactions on Robotics, Vol. 21, pp. 364-375, 2005

[5] Paz, L. M., Pinies, P., Tardos, J. D., Neira, J.: Large scale 6DOF SLAM with stereo-in-hand, IEEE Transactions on Robotics, Vol. 24, pp. 946-957, 2008

ISBN 978-3-8007-3273-9

[6] Bay, H., Ess, A., Tuytelaars, T., Van Gool, A.: SURF: Speeded up robust feature, Computer Vision and Image Understanding, Vol. 110, pp. 346—359, 2008

[7] Lowe, D. G.: Object recognition from local scale-invariant features, International Conference on Computer Vision, pp. 1150-1157, 1999

[8] Gil, A., et al.: A comparative evaluation of interest point detectors and local descriptors for visual SLAM, Machine Vision and Applications, Springer, 2009

[9] Evans, C.: Notes on the OpenSURF library, University of Bristol, Technical Report: CSTR-09-00, 2009

[10] Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features, Computer Vision and Pattern Recognition, 2001

[11] Fischler, M. A., Bolles, R. C.: Random Sample Consensus: a paradigm for model fitting with applications to image analysis and automated cartography, Vol. 24, pp. 381-395, Communications of the ACM, 1981

[12] Zuliani, M., Kenney, C.S., Manjunath, B.S.: The multiransac algorithm and its application to detect planar homographies, pp. 153-156, 2005

[13] Fontanelli, D., Ricciato, L., Soatto, S.: A fast RANSAC - based registration algorithm for accurate localization in unknown environments using LIDAR measurements, In Proc. IEEE Int. Conf. on Automation Science and Engineering, 2007

[14] Torr, P., Zisserman, A.: MLESAC: A New Robust Estimator with Application to Estimating Image Geometry, Computer Vision and Image Understanding, Vol. 78, 2000

[15] Tamas, L., et al.; Laser based localization techniques for indoor mobile robots, Advanced Technologies for Enhanced Quality of Life, pp. 169 – 170, 2009

[16] Chum, O., Matas, J.: Optimal Randomized RANSAC, IEEE Transactions on Pattern Analysis & Machine Intelligence, Vol. 30, pp. 1472-1482, 2008

[17] Botterill, T., Mills, S., Green, R.: New Conditional Sampling Strategies for Speeded-Up RANSAC, Proceedings of the British Machine Vision Conference, 2009

[18] Besl, P. J., McKay, N. D.: A Method for Registration of 3-D Shapes, IEEE Transactions on Pattern Analysis & Machine Intelligence, Vol. 14, pp. 239-256, 1992

[19] Thrun, S., et al.: A System for Volumetric Robotic Mapping of Abandoned Mines, IEEE International Conference on Robotics and Automation, pp. 4270-4275, 2003

ISBN 978-3-8007-3273-9