DIRAC Training Report
Robot self-localization through matching between an incoming image and a textured 3D model

Jan Knopp
Center for Machine Perception
Department of Cybernetics, CTU Prague
Karlovo nam. 13, 121 35 Prague 2, Czech Republic
knoppj1@cmp.felk.cvut.cz

September 24, 2007

For audio processing techniques it is important to acquire information about the environment where the signal was obtained. The environment is more easily recognized from the video input than the audio. The goal of my work during my three month stay at KU Leuven was to determine the exact position from the image.

All papers published on self-localization through a topological map use omnidirectional cameras with 360 ° view, unlike my approach. We face the challenge of using a camera with 180 ° view, which is more useful for mobile localization in the terrain. This results in many differences in map building and the self-localization process compared to reference papers.

Firstly, the topological line of the world [3] was made from a group of images, although the framework had been designed for a topological map, in which case it is necessary to use the exact position of each image [4]. Significant images were determined which represent a group of images with similar properties and similar position. This map building is the off-line part of the algorithm, which means that it is not time critical. In order to find similarities the fast method SURF [1] is used. All feature points and their description were precomputed as well for each image in this off-line part.

Images that are obtained online are given as input to the final self-localization algorithm. The process is divided into many levels to increase the efficiency of our algorithm. The first level is for the first few input images. On this set only Bayesian filtering [2] is aplied with only a few feature points and only for the precomputed significant images from the data set. On the next set of input images the first level algorithm is applied, which uses knowledge about the omnidirectional images represented by the significant images with high probability. The last level makes use of the same information as the first two levels and additionally the epipolar geometry to obtain the accurate position of the input image. Nister's five points algorithm [6] to obtain the essential matrix is used. This allows us to compute the vector to the final position of the image [5]. Also the work done by Michal Havlena [4] during his DIRAC intership at KU Leuven is used to determine the exact position of an image from the dataset.

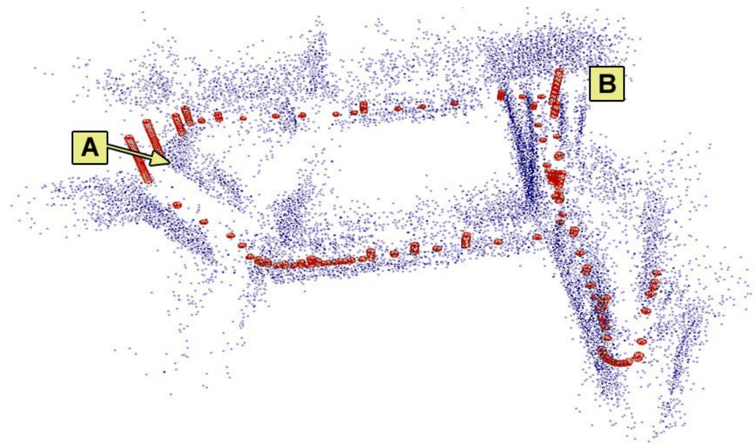

Figure 1: *Reconstructed city. The probability of each significant image is indicated by a cylinder (the height of the cylinder represents the probability). The right position is indicated by "A". The similar environment, indicated by "B" in the above figure, will be eliminated by using the epipolar geometry.*

The speed of the algoritm varied between 0.4s (for the first level) and 10s (for accurate position localization). The speed is sufficient for practical use, but the slowest part, matching, could be implemented more efficiently. In spite of the framework design for the topological map, the algorithm was only tested for the topological line. However we plan to test it for the topological map as well in the near future. The algorithm seems to be useful for recognizing the position of images despite the occurence of similar environments in data set images. We tested our algorithm on three different scenarios which includes 1500 images in total.

Acknowledgements

We are grateful to Kurt Cornelis and Mario Ausseloos to supervise my work. I would also like to acknowledge Maarten Vergauwen for helping me with the epipolar geometry. Also so many thanks to all VISICS staff for their friendliness and support.

References

[1] H. Bay, T. Tuytelaars, and L. Van Gool, *"SURF: Speeded Up Robust Features"*, European Conference on Computer Vision, Lecture notes in computer science, vol. 3951, pp. 404-417, 2006.

[2] T. Goedem, M. Nuttin, T. Tuytelaars, and L. Van Gool, *"Markerless Computer Vision Based Localization using Automatically Generated Topological Maps"*, Proceedings European Navigation Conference GNSS, 2004.

[3] T. Goedem, T. Tuytelaars, and L. Van Gool, *"Visual topological map building in self-similar environments"*, International Conference on Informatics in Control, Automation and Robotics, ICINCO 2006.

[4] M. Havlena, K. Cornelis, and T. Pajdla, *"Towards City Modeling from Omnidirectional Video"*, CVWW 2007, St. Lambrecht, Austria.

[5] R. Hartley, A. Zisserman, *"Multiple View Geometry in computer vision"*, Cambridge University Press 2000.

[6] D. Nistr, *"An Efficient Solution to the Five-Point Relative Pose Problem"*, Proc. Computer Vision and Pattern Recognition,pp. 195-202, 2003.