# DIRAC Training - Report

Michal Havlena
Center for Machine Perception
Department of Cybernetics
Faculty of Electrical Engineering
Czech Technical University in Prague
Karlovo nám. 13, 121 35 Prague 2, Czech Republic
havlem1@cmp.felk.cvut.cz

January 7, 2007

Image-based navigation is very popular these past few years. Typical systems usually rely on two or more perspective cameras. As perspective cameras are used, some problems arise. Occlusions and sharp camera turns can cause good feature tracks to be lost easily, camera poses which are computed from features far away in 3D from the camera are not precise enough and the limited field-of-view might not be adequate for good texture generation. All of these problems can be avoided if omni-directional cameras (e.g. a fish-eye lens converter [3]) are used.

During my three month visit to KULeuven we succeeded in adapting the structure-from-motion system implemented by Nico and Kurt Cornelis [1] to use omni-directional cameras. The major changes involved putting constraints on feature detection and significant changes in geometry as rays represented by unit vectors had to be used instead of image pixels in all parts of the framework.

The hypothesize-and-test loop changed completely. Using the RANSAC [2] stopping condition, together with a reduction of the sample size from 6 to 3, led to a significant acceleration of retrieving the camera pose with the biggest support of putative matches. The average number of iterations needed when running on our test sequence became 12 compared to 1700 iterations performed when running the original framework which used a fixed number of iterations.

The reduction of the sample size was possible by using Nister's algorithm providing the minimal solution to the generalized 3-point pose problem [5]. The main advantage in using this algorithm, designed for non-central projection cameras, lies in the fact that the rays do not need to be concurrent and therefore rays going through both the left and the right cameras can be combined together in a single sample. Because Sturm sequences and bisection with a fixed number of iterations were used to solve the resulting 8-degree polynomial [4], the demands on computing time are minimized.

When looking for a high quality lens to use in the experiments of the modified framework, we had chosen the Nikon FC-E9 offering 183 degrees field-of-view, mounted on a Kyocera Finecam M410R providing 3fps of high resolution images with the radius of the captured view field approximately 800 pixels with nearly excellent image quality.

I will continue to work together with KULeuven to further improve the framework. Speeding up the framework will be our main goal for the near future. We believe that using an Intel-optimized compiler, rewriting some frequently called functions, and transferring function parameters into global variables, together with a reduced input image size enabled by a fisheye-oriented extension which describes features on a locally unwarped image, can bring us closer to a real-time performance.

We will also extend the structure-from-stereo algorithm with a part that combines tracks describing the same corners which were accidentally lost in one or more of the frames due to occlusions or fast camera movements. This should overcome the gradual loosing of the scale and the drifting of the camera poses still present now.

Another big portion of future work lies in adapting the other parts of the framework to using omni-directional cameras - starting with the iterative bundle adjustment routine through dense reconstruction, topological map generation, and texture generation. Mostly the latter part could benefit a lot from omni-directional vision.

A paper describing the modifications in detail together with the results of the experiments has already been submitted to Computer Vision Winter Workshop 2007 (CVWW07) and another paper based on current work is being prepared to The 5th International Symposium on Mobile Mapping Technology (MMT07).

# Acknowledgements

We are grateful to Kurt Cornelis for helping us to understand the original framework and would also like to acknowledge Wim Moreau for the construction of the stereo rig which we used.

# References

[1] N. Cornelis, K. Cornelis, and L. Van Gool. Fast compact city modeling for navigation pre-visualization. In *CVPR 2006*, volume 2, pages 1339–1344, 2006.

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

[3] B. Mičušík and T. Pajdla. Structure from motion with wide circular field of view cameras. *IEEE Trans. PAMI*, 28(7):1135–1149, July 2006.

[4] D. Nistér. An efficient solution to the five-point relative pose problem. *IEEE Trans. PAMI*, 26(6):756–770, June 2004.

[5] D. Nistér. A minimal solution to the generalized 3-point pose problem. In *CVPR 2004*, volume 1, pages 560–567, 2004.