EPSC Abstracts,
Vol. 4, EPSC2009-280, 2009
European Planetary Science Congress,
© Author(s) 2009

Automatic Reconstruction of Mars Artifacts

M. Havlena, A. Torii, M. Jančošek, and **T. Pajdla**

Center for Machine Perception, Department of Cybernetics, Faculty of Elec. Eng., Czech Technical University in Prague, Karlovo náměstí 13, 121 35 Prague 2, Czech Republic {havlem1, torii, jancom1, pajdla}@cmp.felk.cvut.cz

Abstract

We introduce a fast and fully automatic reconstruction pipeline for dense 3D model generation from images acquired on the Mars. Its performance is demonstrated on 9 images captured by the Phoenix Mars Lander.

Introduction

The growing amount of available imagery from the Mars acquired recently by landers and rovers makes the manual image processing techniques used [1] no longer feasible. We introduce a fully automatic reconstruction pipeline for dense 3D model generation using only images themselves and the internal camera calibration as the input. The resulting 3D model can be viewed from an arbitrary viewpoint bringing the user a realistic impression similar to "being in there".

The Pipeline

We demonstrate the performance of our pipeline on data set PHOENIX which consists of 9 images captured by the Phoenix Mars Lander just after performing some digging operation on the Mars surface. The internal camera calibration is obtained from the parameters of the CAHVOR camera model [2] present in the image header.

Determining Image Order

First, we use the image similarity scores to sort the input images into a sequence. Up to thousands of SURF features [3] are detected and described on each of the input images and quantized into visual words according to a visual vocabulary trained on images from the Mars using FLANN [4]. Next, term frequency–inverse document frequency (tf-idf) vectors [5] are computed for each image with more than 50 detected features and finally, pairwise image similarity matrix S_{II} containing cosines of angles between normalized tf-idf vectors t_a, t_b of images I_a, I_b is computed as

$$S_{II}(a,b) = t_a \cdot t_b. \tag{1}$$

Starting from an arbitrary image, the following image is selected as such having the highest similarity score with the previously selected image among the images that have not been selected yet.

Structure-from-Motion Computation

Structure-from-Motion computation recovers the unknown camera poses. First, relative camera poses between consecutive cameras in the sequence are obtained. Different affine covariant feature regions including MSER [6] and SURF [3] are detected in input images. The detected regions are assigned local affine frames (LAF) [7] and described by discrete cosine descriptors [8]. Secondly, tentative feature region matches are constructed from mutually closest descriptors in the feature space using FLANN [4] performing the fast approximate nearest neighbour search based on a hierarchical k-means tree. The 5-point minimal relative pose problem for calibrated cameras [9] is used for generating the camera pose hypotheses and PROSAC [10], an ordered variant of RANSAC, together with voting similar to that used in [11] is used to find the largest subset of the set of tentative matches that is geometrically consistent. Finally, inliers of the geometry test are triangulated into 3D points [12]. Relative camera poses are chained through the sequence resulting in the absolute poses of all cameras.

Dense 3D Model Generation

Knowing the camera poses, we can reconstruct a dense 3D model of the captured scene. We use a Scalable Multi-View Stereo pipeline working with an unordered set of images and corresponding camera poses [13].

The pipeline follows the reconstruction paradigm used in work [14], which can deal with large video sequences working with a few neighbouring frames of each actual frame to compute and fuse the depth maps. We build upon the work of [15]. In particular, we modify the reconstruction process to be scalable by accumulating reconstructed scene and avoiding unnecessary computations and improve the filtering step by using MRF filtering formulation [16]. Three different views of the resulting model are shown in Figure 1.

EPSC Abstracts,
Vol. 4, EPSC2009-280, 2009
European Planetary Science Congress,
© Author(s) 2009

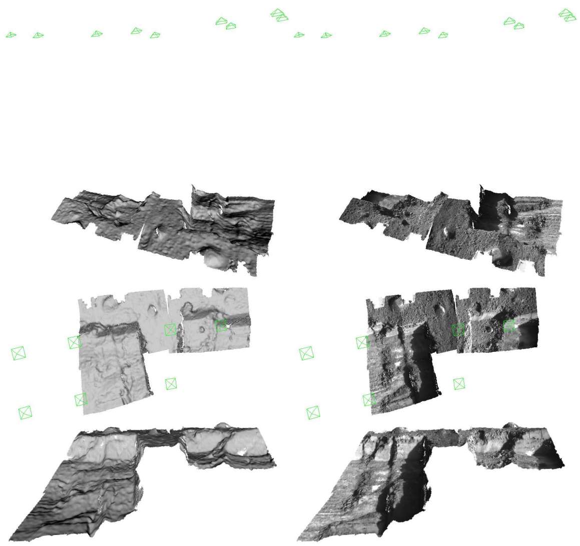

Figure 1: Mars surface reconstruction (PHOENIX data set). Different views of the resulting 3D model without texture (left) and covered by the texture taken from the input images (right). Computed camera poses are denoted by green pyramids.

Conclusions

We introduced a pipeline suitable for a fully automatic reconstruction of Mars artifacts and presented its performance on images acquired by the Phoenix Mars Lander. As the overall computation time on a standard Core2Quad PC was around 3 minutes, our method offers a significant speedup compared to manual techniques currently used.

Acknowledgements

This work was supported by the European IST Programme Project FP7-218814 PRoVisG and by project MSM6840770038 DMCM III. This paper only reflects the authors' views and EC is not liable for any use that may be made of the information contained herein.

References

[1] D. A. Alexander, R. G. Deen, P. M. Andres, P. Zamani, H. B. Mortensen, A. C. Chen, et al., "Processing of Mars exploration rover imagery for science and operations planning," *JGR*, vol. 111, no. E02, February 2006.

[2] K. Di and R. Li, "CAHVOR camera model and its photogrammetric conversion for planetary applications," *JGR*, vol. 109, no. E04, April 2004.

[3] H. Bay, A. Ess, T. Tuytelaars, and L. J. Van Gool, "Speeded-up robust features (SURF)," *CVIU*, vol. 110, no. 3, pp. 346–359, June 2008.

[4] M. Muja and D. Lowe, "Fast approximate nearest neighbors with automatic algorithm configuration," in *VISAPP*, 2009.

[5] J. Sivic and A. Zisserman, "Video google: Efficient visual search of videos," in *CLOR06*, 2006, pp. 127–144.

[6] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions," *IVC*, vol. 22, no. 10, pp. 761–767, September 2004.

[7] Š. Obdržálek and J. Matas, "Object recognition using local affine frames on distinguished regions," in *BMVC'02*, 2002, pp. 113–122.

[8] Š. Obdržálek and J. Matas, "Image retrieval using local compact DCT-based representation," in *DAGM'03*, 2003.

[9] D. Nistér, "An efficient solution to the five-point relative pose problem," *PAMI*, vol. 26, no. 6, pp. 756–770, June 2004.

[10] O. Chum and J. Matas, "Matching with PROSAC – progressive sample consensus," in *CVPR'05*, 2005, pp. 220–226.

[11] H. Li and R. Hartley, "A non-iterative method for correcting lens distortion from nine point correspondences," in *OMNIVIS'05*, 2005.

[12] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, second edition, 2003.

[13] A. Torii, M. Havlena, M. Jančošek, Z. Kúkelová, and T. Pajdla, "Dynamic 3d scene analysis from omni-directional video data," Research Report CTU–CMP–2008–25, CMP Prague, December 2008.

[14] P. Merrell, A. Akbarzadeh, L. Wang, P. Mordohai, J. M. Frahm, R. G. Yang, D. Nister, and M. Pollefeys, "Real-time visibility-based fusion of depth maps," in *ICCV'07*, 2007.

[15] Y. Furukawa and J. Ponce, "Accurate, dense, and robust multi-view stereopsis," in *CVPR'07*, 2007.

[16] N. Campbell, G. Vogiatzis, C. Hernandez, and R. Cipolla, "Using multiple hypotheses to improve depth-maps for multi-view stereo," in *ECCV'08*, 2008.