

Real-time 3D visual SLAM with a hand-held RGB-D camera

Nikolas Engelhard^a

Felix Endres^a

Jürgen Hess^a

Jürgen Sturm^b

Wolfram Burgard^a

The practical applications of 3D model acquisition are manifold. In this paper, we present our RGB-D SLAM system, i.e., an approach to generate colored 3D models of objects and indoor scenes using the hand-held Microsoft Kinect sensor. Our approach consists of four processing steps as illustrated in Figure 1. First, we extract SURF features from the incoming color images. Then we match these features against features from the previous images. By evaluating the depth images at the locations of these feature points, we obtain a set of point-wise 3D correspondences between any two frames. Based on these correspondences, we estimate the relative transformation between the frames using RANSAC. The third step is to improve this initial estimate using a variant of the ICP algorithm [1]. As the pair-wise pose estimates between frames are not necessarily globally consistent, we optimize the resulting pose graph in the fourth step using a pose graph solver [4]. The output of our algorithm is a globally consistent 3D model of the perceived environment, represented as a colored point cloud. The full source code of our system is available as open source [2]. With an earlier version of our system, we participated in the ROS 3D challenge organized by Willow Garage and won the first prize in the category “most useful”.

Our approach is similar to the recent work of Henry et al [5]. Our approach applies SURF instead of SIFT features. Additionally, our source code is available online.

Figures 2 and 3 illustrate the quality of the resulting 3D models. For both experiments, we slowly moved the Kinect around the object and acquired around 12 RGB-D frames. Computing the model took approximately 2 seconds per frame on an Intel i7 with 2 GHz. We applied our approach also to a large variety of other objects. Videos with more results are available online [3]. We will demo our system during the RGB-D workshop. Further, we plan to evaluate our system using ground truth information in the near future.

Our approach enables a robot to generate 3D models of the objects in the scene. But also applications outside of robotics are possible. For example, our system could be used by interior designers to generate models of flats and to digitally refurbish them and show them to potential customers. At the moment, we do not deal with the problem of automatic view point selection but assume instead that the user is moving the camera through the scene.

^a N. Engelhard, F. Endres, J. Hess, and W. Burgard are with the Autonomous Intelligent Systems Lab, Computer Science Department, University of Freiburg, Germany. {engelhar, endres, hess, burgard}@informatik.uni-freiburg.de

^b J. Sturm is with the Computer Vision and Pattern Recognition Group, Computer Science Department, Technical University of Munich, Germany. sturmju@in.tum.de

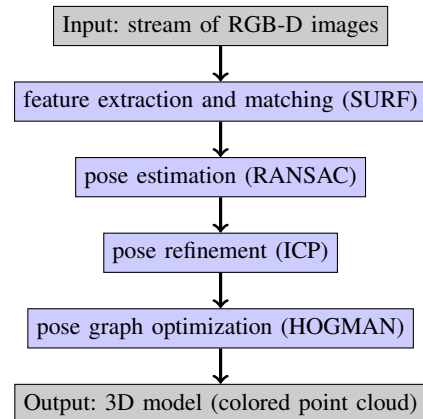


Fig. 1: The four processing steps of our approach. Our approach generates colored 3D environment models from the images acquired with a hand-held Kinect sensor.

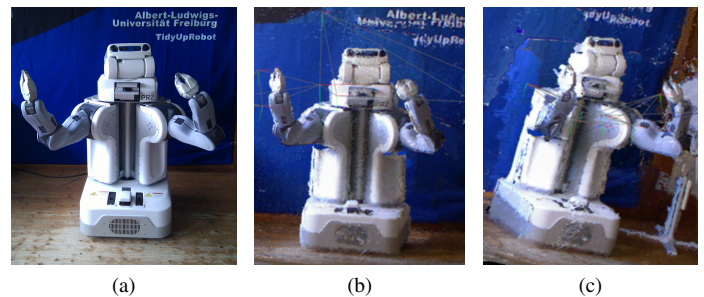


Fig. 2: (a) Image of the PR2 robot in our lab. (b) and (c) Resulting model, visualized from two different perspectives. As can be seen from these images, the individual point clouds have accurately been integrated into the map.

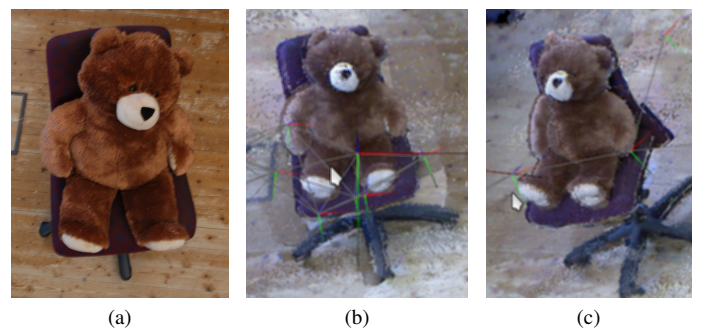


Fig. 3: (a) Image of a teddy bear. (b) and (c) Resulting model, visualized from two different perspectives.

REFERENCES

- [1] A. Segal, D. Haehnel, S. Thrun. Generalized ICP. In *Proc. of Robotics: Science and Systems (RSS)*, 2009.
- [2] F. Endres, J. Hess, N. Engelhard, J. Sturm, and W. Burgard. http://www.ros.org/wiki/openni/Contests/ROS_3D/RGBD-6D-SLAM, Jan. 2011.
- [3] F. Endres, J. Hess, N. Engelhard, J. Sturm, and W. Burgard. <http://www.youtube.com/watch?v=XejNcct2Fcs>, [?v=5qrBEPfEPaY](http://www.youtube.com/watch?v=5qrBEPfEPaY), and [?v=NR-ycTncQu0](http://www.youtube.com/watch?v=NR-ycTncQu0), 2011.
- [4] G. Grisetti, R. Kümmerle, C. Stachniss, U. Frese, and C. Hertzberg. Hierarchical optimization on manifolds for online 2D and 3D mapping. In *Proc. of the IEEE Intl. Conf. on Robotics and Automation (ICRA)*, Anchorage, AK, USA, 2010.
- [5] P. Henry, M. Krainin, E. Herbst, X. Ren, and D. Fox. RGB-D mapping: Using depth cameras for dense 3D modeling of indoor environments. In *Proc. of the Intl. Symp. on Experimental Robotics (ISER)*, Delhi, India, 2010.