GPSlam: Marrying Sparse Geometric and Dense Probabilistic Visual Mapping

Katrin Pirker [1]
kpirker@icg.tugraz.at
Gerald Schweighofer[2]
gerald.schweighofer@joanneum.at
Matthias Rüther[1]
ruether@icg.tugraz.at
Horst Bischof[1]
bischof@icg.tugraz.at

[1] Institute for Computer Graphics and Vision
Graz University of Technology
Graz, Austria

[2] Institute of Digital Image Processing
Joanneum Research GmbH
Graz, Austria

1 Motivation

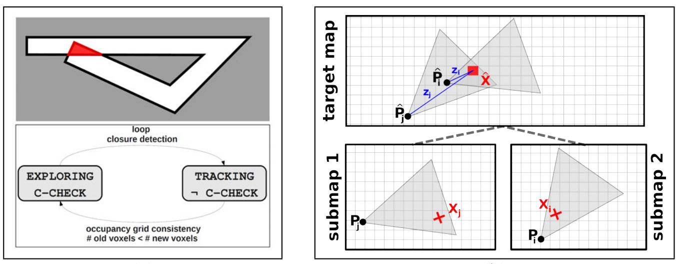
The ultimate goal of a visual Simultaneous Localization and Mapping (SLAM) routine is building a dense three-dimensional representation of the environment which can be used for high-level robotic tasks such as path planning, obstacle avoidance or object manipulation. Previous approaches follow either a pure geometric [3] approach (i.e. incremental Structure from Motion) where triangular meshes or geometric primitives have to be fitted to pointcloud data in order to construct a dense environment. Probabilistic environment modeling instead [2], using occupancy grids, results in dense environment representations at the expense of high memory consumption and long computation time.

We try to combine the positive aspects of geometric and probabilistic SLAM systems using image and depth information delivered from a range image device:

- Pose estimation using a state-of-the-art incremental, keyframe-based Structure from Motion (SfM) algorithm provides accuracy at low computational cost.
- Simultaneous computation of a single dense three-dimensional occupancy grid handles sensor noise, scene dynamics and reduces memory consumption.

2 Sparse geometric SLAM combined with Probabilistic Map Building

Keyframe-based, sparse geometric SLAM is performed by an incremental SfM algorithm. Loop closure detection is handled through a vocabulary tree followed by a pose graph optimization [4] to correct accumulated motion errors. Provided accurate keyframe poses, a single 3D occupancy grid is filled simultaneously using the depth information delivered by the sensor.



(a) (b)

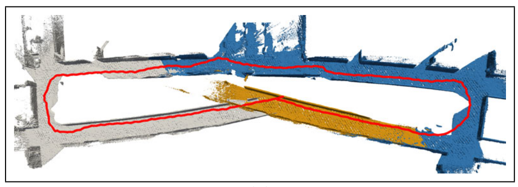
Figure 1: Inconsistency handling and grid morphing. (a) Accumulated motion estimation errors can lead to the destruction of the carried occupancy grid. After detecting this conflict a new submap is created (top). A simple state machine describes the transition between tracking and exploration mode using image based loop closure detection and occupancy information (bottom). (b) For each voxel $\hat{X}$ in the resulting grid we determine the cameras $\hat{P}_i$ it is visible in. Using the known camera poses before P_i and after the loop closing $\hat{P}_i$, we calculate its potential previous positions X_i. Their log-odds weighted with the distances z_i to the cameras $\hat{P}_i$ defines the new occupancy value.

Because pose estimation errors accumulate over time, we may destroy the carried occupancy grid with incoming sensor data (such a conflict is sketched in Figure 1(a)). In order to trigger such conflicts, we diminish between exploration mode (i.e. traveling in unknown regions while expanding the map) and localization mode (i.e. traveling in previously mapped areas while updating the map). Therefore, each voxel stores the traveled distance of the camera it has been most recently updated from. During localization no conflict detection takes place and sensor data can be safely used for map update. In exploration mode instead, a conflict is triggered if the difference between the trajectory length of the camera and the voxels value exceeds a predefined threshold. A state machine describing this behavior is shown in Figure 1(a). On conflicts local submaps are created, which have to be merged afterwards.

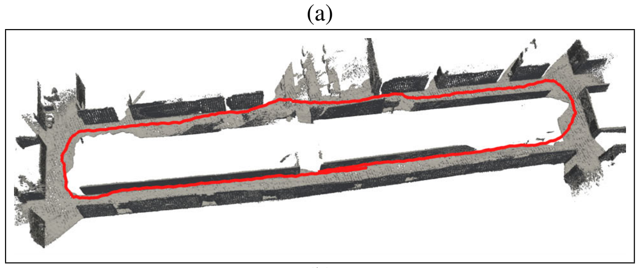
To align occupancy grids to a globally corrected trajectory (i.e. after loop closure) we propose an inverse mapping procedure for each voxel lying in the view cones of the corrected camera poses (compare Figure 1(b)). To retain the probabilistic information included in all submaps a weighted average of their log-odds defines the new occupancy value of the corrected grid.

3 Experimental Results

We acquired two indoor sequences using a calibrated, hand-held Kinect [1] while traveling a loop in an office and a more spacious corridor environment, respectively. As shown in Figure 2(a) we correctly detected inconsistent grid updates during erroneous exploration where three submaps have been created. After geometric loop closure correction occupancy grids are successfully aligned to its sparse counterpart resulting in a consistent representation of the environment (see Figure 2(b)).



(a)



(b)

Figure 2: Corridor sequence. (a) The inconsistency check successfully created three submaps. (b) Resulting environment after loop closure correction and grid morphing.

[1] Microsoft KINECT. URL `www.xbox.com/kinect`.

[2] T. K. Marks, A. Howard, M. Bajracharya, G. W. Cottrell, and L. Matthies. Gamma-SLAM: Using stereo vision and variance grid maps for SLAM in unstructured environments. In *IEEE International Conference on Robotics and Automation*, pages 3717–3724, 2008.

[3] R.A. Newcombe and A.J. Davison. Live dense reconstruction with a single moving camera. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1498 –1505, 2010.

[4] H. Strasdat, J. M. M. Montiel, and A. Davison. Scale drift-aware large scale monocular slam. In *Proceedings of Robotics: Science and Systems*, Zaragoza, Spain, June 2010.