Stereo Vision Odometry

Yulin Yang - yuyang@udel.edu
Di Zhao - zhaodi@udel.edu

Department of Mechanical Engineering
University of Delaware, Delaware, USA

RPNG

Robot Perception and Navigation Group (RPNG)
Last Updated - Dec. 09, 2016

Contents

1 Project Target

1) Using Stereo Camera for 3D Reconstruction and Depth Mapping.
2) Be familiar with the following libraries:
* libviso2 [1]: A small efficient lib for stereo vision odometry
* GTSAM [2]: an efficient lib for nonlinear least-squares optimization
* Pangolin [3]: an lightweight lib for managing OpenGL displaying
* OpenMVG [4]: an lib for Multi-View Structure from Motion
* OpenMVS [5]: an lib for computing dense points cloud, surface etc.
* KITTI Dataset [6]: A novel challenging real-world vision benchmarks suite.

2 Stereo Camera Model

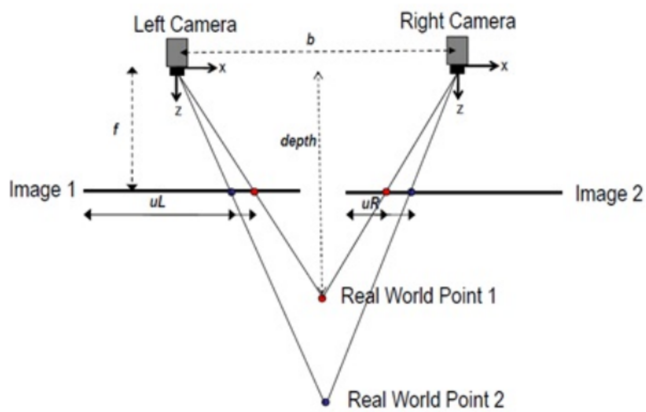

Figure 1: Stereo Vision Model

As shown in the figure 1(pic from Internet), the stereo camera model we are using here is as following:

$$\begin{bmatrix} uL \\ uR \\ vL \end{bmatrix} = \begin{bmatrix} f_x * \frac{P_x}{P_z} + c_x \\ f_x * \frac{P_x - b}{P_z} + c_x \\ f_y * \frac{P_y}{P_z} + c_y \end{bmatrix} \tag{1}$$

where $P(P_x, P_y, P_z)$ represents the real world point. (f_x, f_y) is the focus length, (c_x, c_y) is the image center,b is the baseline for the stereo vision.

From the measurement model, we can solve the world point as:

$$\begin{bmatrix} P_x \\ P_y \\ P_z \end{bmatrix} = \begin{bmatrix} (uL - c_x) \cdot \frac{b}{(uL - uR)} \\ (uR - c_y) \cdot \frac{f_x}{f_y} \cdot \frac{b}{uL - uR} \\ f_x \cdot \frac{b}{uR - uL} \end{bmatrix} \tag{2}$$

As we are using rectified images, the focus length $f = f_x = f_y$, and we denote the disparity as $d = uL - uR$. Hence, we can simplified the above equation as:

$$\begin{bmatrix} P_x \\ P_y \\ P_z \end{bmatrix} = \begin{bmatrix} (uL - c_x) \cdot \frac{b}{d} \\ (uR - c_y) \cdot \frac{b}{d} \\ f \cdot \frac{b}{d} \end{bmatrix} \tag{3}$$

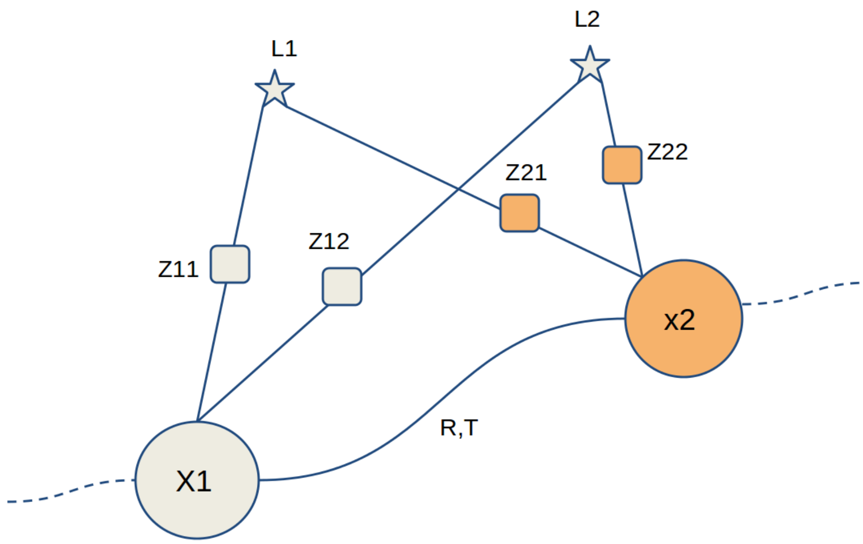

Figure 2: Local Bundle Adjustment

3 Local Optimization

As shown in the figure 2, x_1 and x_2 represents the consecutive camera poses, L_1 and L_2 represents the 2 landmarks that are observed both by x_1 and x_2. z_{ij}, $i = 1, 2, j = 1, 2$ represents the measurements from the camera to the landmarks. Their relationship can be expressed as:

$$z_{ij} = h(x_i, L_j) \tag{4}$$

where $h(\cdot)$ represents the projective transformation from the world 3D points to the stereo measurements. Therefore, we can formulate an optimization function for solving the the poses of the camera and the positions of these landmarks:

$$\sum_i \sum_j ||z_{ij} - h(x_i, L_j)||^2_{(\Sigma_{ij})^{-1}} \tag{5}$$

where Σ_{ij} represents the noises covariance of the measurement z_{ij}.

4 OpenMVG and OpenMVS

General procedures for using OpenMVG and OpenMVS:
1) Image listing;
2) Image description computation;
3) Corresponding images and correspondences computation;
4) SFM solving;
5) OpenMVS for post processing;
6) Meshlab for displaying.

5 Conclusion

As shown in the figure, the red part that are still needed to done in the future.
1) We need to find the feature global feature correspondences to build the 3D point map;
2) We need to implement the interface with OpenM?G and our SLAM system to generate the 3D point cloud.

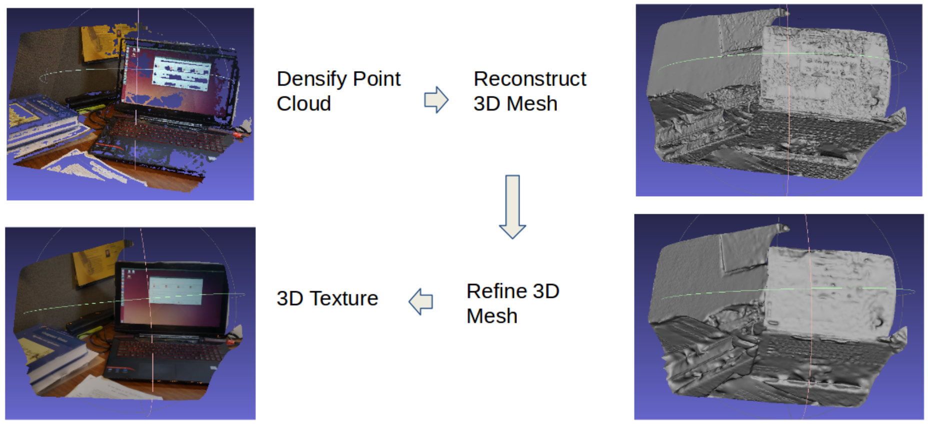

Figure 3: OpenMVS Processing

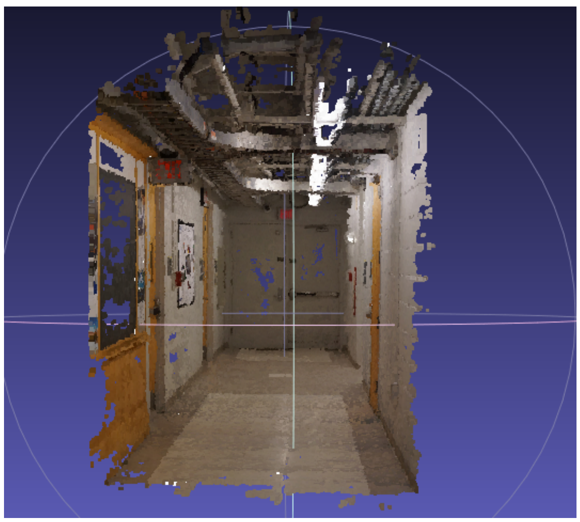

Figure 4: A Complete Demo: Densified Point Cloud

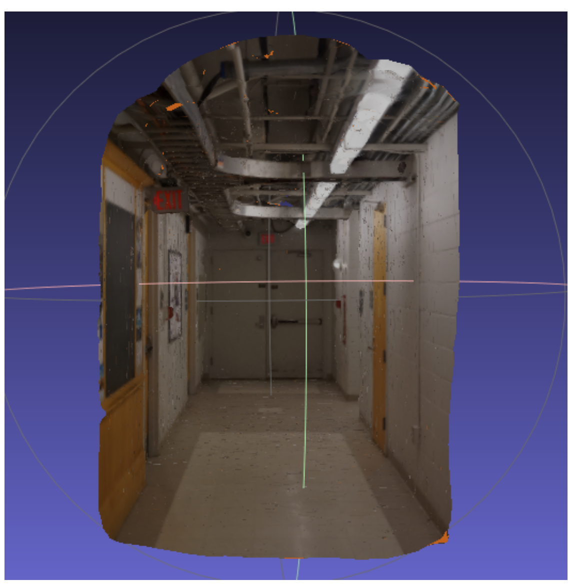

Figure 5: A Complete Demo: Final Textured Model

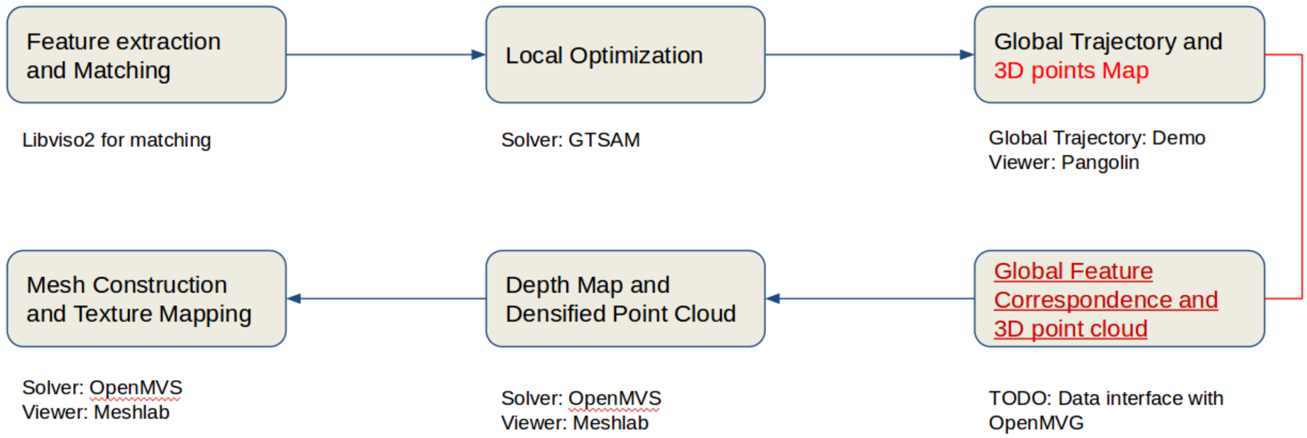

Figure 6: Conclusion

References

[1] A. Geiger, J. Ziegler, and C. Stiller, "Stereoscan: Dense 3d reconstruction in real-time," in *Intelligent Vehicles Symposium (IV)*, 2011.

[2] F. Dellaert, "Factor graphs and gtsam: A hands-on introduction," 2012.

[3] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardós, "ORB-SLAM: a versatile and accurate monocular SLAM system," *IEEE Transactions on Robotics*, vol. abs/1502.00956, 2015.

[4] P. Moulon, P. Monasse, and R. Marlet, "Openmvg (open multiple view geometry)," 2012.

[5] P. Moulon, "Openmvs (open multiple-view stereo)," 2012.

[6] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.