Fast Line Description for Line-based SLAM

Keisuke Hirose
hikeisuke@hvrl.ics.keio.ac.jp

Hideo Saito
http://hvrl.ics.keio.ac.jp/saito/

Graduate School of Science and Technology,
Keio University

Simultaneous localization and mapping (SLAM) is a technique to simultaneously perform mapping of environments and localization of a camera in real-time. Vision based SLAM is used for real applications such as augmented reality. Most existing monocular vision based SLAM techniques employ point features as landmarks[3].

Our approach uses line segments rather than points as landmarks, since there are some advantages in using line segments. Images of artificial environments with little texture contain many line segments, whereas few point features can be localized in such a scene. Moreover, line segment matching is more robust than point matching with respect to partial occlusion and view-point changes. With regard to the SLAM based on line segments, existing line based SLAM systems[1, 4] don't have any descriptions of line segments. In order to establish 2D and 3D correspondences, they simply use the distance between line segments in the image space. Therefore, wrong correspondences often occur in complicated scenes that include many line segments.

We propose here a real-time SLAM system that uses line segments as landmarks, and a fast line descriptor (LEHF:Line-based Eight-directional Histogram Feature) in order to establish correct 2D and 3D correspondences.

In localization, line segments are detected by the line segment detector (LSD) method[2] at every frame. LEHF that is our new line descriptor is computed for each detected line segment. We based the development of LEHF on the mean standard deviation line descriptor (MSLD)[5], which uses a SIFT-like strategy. Fig.1 shows how LEHF is computed.

For each point that is uniformly taken around the line segment, the gradient vector is computed based on differential values (dx, dy). As shown in the figure, we obtain 14 eight-directional gradient histograms by summing 45 gradient vectors along the line segment. 14 computed eight-directional gradient histograms are merged to obtain a line descriptor referred to as LEHF. However, if we simply merge all the histograms, computed LEHFs are not matched between the images that one image is rotated 180 degrees since the directions of the eight-directional gradient histogram are not matched. Therefore 14 histograms are merged symmetrically.

In order to estimate a camera pose, 2D and 3D line correspondences are established. First 3D line segments are projected into the image space by a previous camera pose. In existing methods, the projected 3D line segment simply corresponds to a detected 2D line segment that is nearest from the projected line segment in the image space. This often results in wrong correspondences being detected. We compute LEHF distances between the projected 3D line segment and some detected 2D line segments. Then the detected 2D line segment that has the minimum Euclidean distance between LEHFs is chosen.

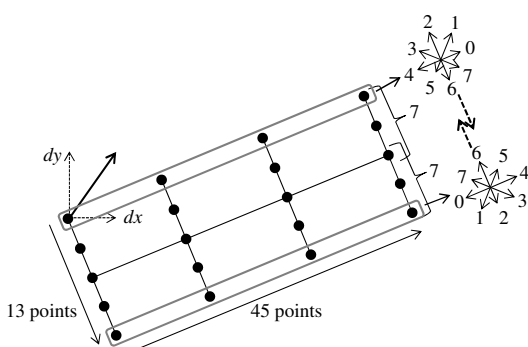The RANSAC algorithm and the line-based orthogonal iteration algorithm (LBOI)[6] estimate a camera pose from 2D and 3D correspondences. All 3D line segments are re-projected by the estimated camera pose to compute re-projection errors. Each 3D line segment is determined inlier or outlier based on the computed re-projection error.

In the mapping, line segments that are not used for localization are tracked between frames as new line segments. In order to track line segments correctly, computed LEHF is used. Tracked line segments in a number of frames are reconstructed for mapping 3D line segments.

We conducted experiments with our SLAM system. The experimental results of demonstrating our SLAM system in a desktop environment are shown in Fig. 2.
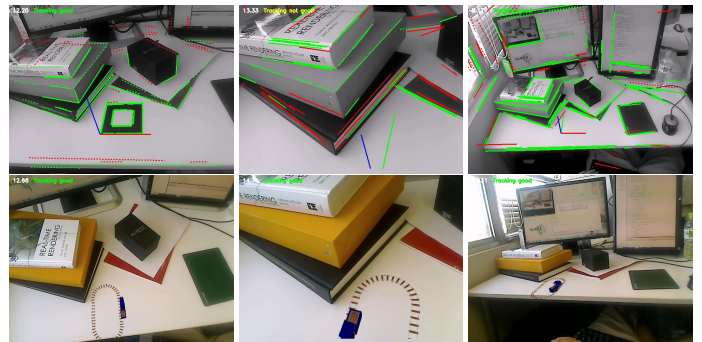

Figure 2: Results of demonstrating our SLAM system in a desktop environment.

The green line segments are inlier line segments and red line segments are outlier line segments. The mean processing time of 1087 frames was 87.78ms. Moreover, we evaluated our SLAM system by using synthetic data. In the synthetic data experiment, we compared our SLAM system with the existing approach using the nearest neighbor search. We show that the use of LEHF provides better accuracy of estimated camera poses.

We have presented a real-time SLAM system based on a line feature called a LEHF. By using LEHF, 2D and 3D correspondences are established correctly and the camera poses are robustly estimated. Our SLAM was demonstrated for augmented reality in a desktop environment and evaluated by using synthetic data.

[1] P.G. Andrew and W. Mayol-Cuevas. Real-time model-based slam using line segments. In *2nd International Symposium on Visual Computing*, 2006.

[2] R. Grompone von Gioi, J. Jakubowicz, J.M. Morel, and G. Randall. Lsd: A fast line segment detector with a false detection control. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(4): 722–732, 2010.

[3] G. Klein and D. Murray. Parallel tracking and mapping for small ar workspaces. In *Mixed and Augmented Reality, 2007. ISMAR 2007. 6th IEEE and ACM International Symposium on*, pages 225–234. IEEE, 2007.

[4] P. Smith, I. Reid, and A. Davison. Real-time monocular slam with straight lines. In *British Machine Vision Conference*, volume 1, pages 17–26, 2006.

[5] Z. Wang, F. Wu, and Z. Hu. Msld: A robust descriptor for line matching. *Pattern Recognition*, 42(5):941–953, 2009.

[6] X. Zhang, K. Wang, Z. Zhang, and Q. Yu. A new line-based orthogonal iteration pose estimation algorithm. In *Information Engineering and Computer Science, 2009. ICIECS 2009. International Conference on*, pages 1–4. IEEE, 2009.

Figure 1: Overview of LEHF