

Three-Point Direct Stereo Visual Odometry

Jeong-Kyun Lee
leejk@gist.ac.kr
Kuk-Jin Yoon
kjiyoon@gist.ac.kr

Computer Vision Laboratory
Gwangju Institute of Science and
Technology (GIST), South Korea

Existing visual odometry methods are generally classified into two groups; feature-based and direct methods. In the feature-based methods, the measurement errors are defined by the re-projection errors of feature points, and the ego-motion is estimated by minimizing the re-projection errors that are assumed to usually conform to the Laplace distribution [1]. However, since the local appearance of the feature point consecutively varies with time, the location of the tracked feature point tends to drift from its initial location during tracking and this results in error accumulation. In contrast, direct methods minimize the measurement errors measured by intensity differences between consecutive images. Since a large number of pixels are utilized for motion estimation, these methods are superior to feature-based methods in the error accumulation aspect. However, in return, the direct methods have a difficulty in handling outliers and are vulnerable to illumination change.

These methods generally use a maximum consensus set of inlier feature points on the Laplace distribution assumption of measurement errors. However, as shown in Fig. 1, the assumption is often violated in the real-world scene because of lots of outliers and/or biases of measurement errors. In this situation, although a motion candidate may be accurately obtained from a small number of point samples in the RANSAC framework, its final estimate optimized from the set of inliers can be rather erroneous. The work of Chum *et al.* [2] tried to solve the similar problem using an iterative local RANSAC scheme but it is still hard to hold thoroughly uncontaminated inliers only.

In this paper, to solve the problem shown in Fig. 1, we propose a method that estimates ego-motion using only uncontaminated feature points. Since it is difficult to distinguish the uncontaminated features among lots of features, we randomly sample a minimum number of features (*i.e.*, 3 points) and exploit them for the estimation. Thereby, the proposed method maximally excludes the errors caused by inaccurate inliers. However, using fewer features in feature-

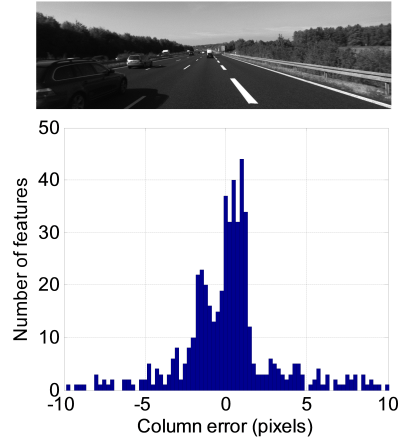


Figure 1: The distributions of re-projection errors for estimated motion in selected frames of the KITTI dataset. The distribution violates the Laplace distribution assumption of the measurement errors because of outliers from moving objects.

based methods can rather degrade the performance because measurement errors caused in feature matching and tracking can be propagated to the ego-motion estimation directly. Therefore, we propose a hybrid method of feature-based and direct methods. It minimizes measurement errors by the ego-motion estimation directly using image intensity values without any intermediate step to search measurements and is more robust than conventional direct methods. Experimental results show the proposed method produces better performance than the existing methods using all inliers.

- [1] Hernan Badino, Akihiro Yamamoto, and Takeo Kanade. Visual odometry by multi-frame feature integration. In *ICCV*, 2013.
- [2] Ondrej Chum, Jiri Matas, and Josef Kittler. Locally optimized ransac. In *Pattern Recognition*, pages 236–243. Springer Berlin Heidelberg, 2003.