

Title

Rail Corridor Object Detection and Positioning Improvement from a Railway Mobile Mapping System

Author: [Xu, Zelin](#)

Year

2023

Abstract

There are multiple types of objects distributed along the rail trackside to support the operation of railroads. Damage to those objects will result in delay and cancellation of trains. Therefore, the maintenance of the objects is important to ensure the safety of railway operations, and the key to maintenance is to know the accurate position of the objects. RILA - a railway mobile mapping system developed by Fugro - provides continuous monitoring of the rail environment without interrupting the daily operation of rail tracks. The RILA dataset contains different types of data including video frames, point clouds, and measurement trajectories from the Global Navigation Satellite System and Inertial Measurement Unit (GNSS-IMU). This enables a remote and automatic object positioning process based on object recognition in video frames and positioning using photogrammetric techniques.

Most of the rail trackside objects are parts of the railway electrification system and the railway signaling system. The available ground truth data contains more information about signaling objects, so this thesis mainly focuses on the positioning of signaling objects. Three typical components: signal lights, railway cabinets, and railway side markers are selected to make case studies and develop the object detection and localization pipeline for them.

There are two main issues that existed in former research: One is object recognition only gives a 2D pixel position for an object, but we want to know its 3D position in the world; the other is the contextual information between frames is ignored in object positioning. Based on that, this thesis developed a new workflow: firstly, this thesis proposes the use of an existing 3D object detection model, Single-Stage Monocular 3D Object Detection via Keypoint Estimation (SMOKE), to recognize objects of interest from the RILA video frames. This model can not only predict an object's 2D pixel position on the image, but also its 3D position with respect to the camera. This enables the localization of an object in each of the frames it is detected in, which provides redundancy to define its final position. Secondly, this thesis developed a pipeline to clean the detection output of SMOKE so that most false detections can be removed. In addition, this thesis developed a method to use the contextual information in image sequences to position objects: first, extract the sequences by classifying all the predictions in the detection output through the Euclidean distances, then for each sequence, analyze the varying trend of predicted positions to select reliable positions. The mean of the selected positions is the final improved position of that object. By making tests on a part of the railway near Ely, UK, with 2983 frames, the workflow detected and positioned all of the 4 signal lights, all of the 10 markers, and 13 of the 14 cabinets in the testing dataset. Most signal lights and cabinets can be positioned within 1 meter to the ground truth, and the mean positioning offset for the railway side markers can be controlled around 2 meters. Complex scenarios could be solved, when up to three cabinets were located close together.