

# Accurate 3D Railway Track Extraction from Aerial Images

Dong Wei<sup>a,1</sup>, Xiaotong Li<sup>a,1</sup>, Yongjun Zhang<sup>a,\*</sup>, Chang Li<sup>b</sup> and Ziqian Huang<sup>a,1</sup>

<sup>a</sup>*School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, 430072, P.R.China*

<sup>c</sup>*College of Urban and Environmental Science, Central China Normal University, Wuhan, 430072, P.R.China*

## ARTICLE INFO

### Keywords:

3D line segments  
line clustering  
line triangulation  
3D line evaluation

## ABSTRACT

Three-dimensional (3D) lines require further enhancement in both clustering and triangulation. Line clustering assigns multiple image lines to a single 3D line to eliminate redundant 3D lines. Currently, it depends on the fixed and empirical parameter. However, a loose parameter could lead to over-clustering, while a strict one may cause redundant 3D lines. Due to the absence of the ground truth, the assessment of line clustering remains unexplored. Additionally, 3D line triangulation, which determines the 3D line segment in object space, is prone to failure due to its sensitivity to positional and camera errors.

This paper aims to improve the clustering and triangulation of 3D lines and to offer a reliable evaluation method. (1) To achieve accurate clustering, we introduce a probability model, which uses the prior error of the structure from the motion, to determine adaptive thresholds;

## 1. Introduction

Currently, the length of the railway has exceeded 1.3 million kilometers on the earth, for which the maintenance and development of railways have a significant impact on safe operations. As the preliminary stage of extracting 3D railway track (RT) accurately and efficiently, to support engineering design, monitor construction quality, and ensure operational safety, has become one of the basic components in the maintenance of existing railways.

The extraction of RT can be achieved by real-time kinematics, LiDAR, and multiple images. The real-time kinematic is generally mounted on a railway measurement vehicle and obtains the RT by moving along the rail track. In general, it has a satisfactory accuracy while requiring operations on the track, thus demanding the cooperation of railway departments, and there are issues related to both safety and efficiency. LiDAR sensors can be mounted on a drone, which is more convenient and secure than real-time kinematic. Because a further process, like point segmentation or classification, is required for RT extraction, the drone must maintain a low flight altitude to satisfy the standards of the point-cloud density, which would impact the efficiency. A drone with cameras can capture aerial images efficiently with a safe distance from the railway area. But RT extraction is challenging in aerial images: (1) The dense points reconstructed with aerial images are inaccurate around the railway track because of the occlusion and matching problems caused by the parallax variation. (2) Joining image semantics to obtain RT might be workable; However, how to detect the semantics of RT accurately and completely in aerial images remains to be studied.

If we reconstruct the dense points from multiple aerial images and detect the RT from point clouds, the overall method of finding RT is similar to deal with the point clouds that obtained from mobile laser scanning (MLS) or airborne laser scanning (ALS). Generally, the RT can be detected with semantic segmentation, while the significant noise, inaccurate edge localization, and large density variations of point clouds bring about great challenges to the robust semantic segmentation. Thus, most general segmentation algorithm cannot be used directly in RT segmentation; instead, the carefully designed geometric priors was used to guide the segmentation and the grouping of RT: such as constructing the shape features and density data on the basis of railway bed extraction. However, these methods relies heavily on the quality and density of the point cloud, thus requiring the drone to maintain a low flight path to improve point cloud quality and reduce the processing range. Compared with point clouds, images contains rich semantic informations. Thus, several studies exploited the deep learning method that design the network for training and detect the RT from aerial images, which demonstrated the effectiveness of deep learning technology in RT extraction.

\*Corresponding author

weidong@whu.edu.cn (D. Wei); zhangyj@whu.edu.cn (Y. Zhang); lichang@cnnu.edu.cn (C. Li)

<sup>1</sup>Co-first authors.

Moreover, the deep learning method relies heavily on training samples and considering the texture of railway regions varies greatly across the world, it may require an increased number of training samples to obtain a more generalizable detection network. In addition, these methods just deal with single frame and lacked the strategy of processing multiple aerial images.

This paper propose the accurate RT extraction for multiple aerial images, which fully exploits the contexture and geometry informations across multiple images:

- To exploit the geometry constraint of RT across multiple images, we extract the straight line from images as the basic geometry cell, for which we propose the robust clustering and triangulation methods. We first propose the noise-resistant clustering across multiple images to obtain the complete and non-redundant 3D line; Then, we propose the novel and accurate triangulation algorithm to refine the 3D line position of the RT.
- To exploit the rich texture information in images, our clustering method exploit the deep features of existing network trianed from millions of images, rather than a new network specifically designed for RT extraction; thus requiring non pre-training, which generally needs expensive samples.

Compared to LiDAR based methods, we use more affordable imaging drones to conduct an efficient and safer railway aera maping than ALS drones or MLS equipments, and the rich contexture is exploited to compensate for issues caused by point cloud quality. Compared to the former image-based methods, we propose the complete clustering and reconstruction strategies that obtain the accurate and non-redundant 3D RT from multiple aerial images; and non pre-training is required due to the utilization of geometry guidance in multiple images.

## 2. Related works

## 3. Methodology

The flow of our methodology is presented in figure 1. We take the aerial images of the railway aera as the input, for which SfM and point clouds are employed in advanced with existing softwares. We first extract the image line and convert it to space line with the locally optimal plane of the point cloud. Then, we cluster the single 3D line to RT candidates with the frame work derived from DBSCAN, during which the contexture information of multiple images extracted from ResNet is used as one of the inlier distance, and the locally SAM are exploited to reduce the seed of DBSCAN to control the false positive. Having obtained the RT cluster, we then extract the vector-based RT with RANSAC and refine it with our triangulation methods, which fully exploit the RT structure to resolve the uncertainty caused by initial image line segment extraction and the point cloud error.

### 3.1. Initial 3D line reconstruction

The initial 3D line  $L$  is the basic cell for the following method. For each 2D line  $\mathbf{l} = [\mathbf{p}_1]$  detected in each image, the two endpoint 3D candidate line segment must be on the we randomly sample three space points around  $\mathbf{l}$  and obtain the space plane  $P$  that  $\mathbf{l}$  may lie on; then, the 3D candidate line segment of  $\mathbf{l}$  corresponding to  $P$  must be on the two rays is obtained by the ray-to-plane intersection:

$$\bar{L} \in R^{3 \times 2} = Inter(P, C + M^{-1} [t_1 \mathbf{p}_1, t_2 \mathbf{p}_2]) \quad (1)$$

and generate  $n$  space planes  $\{P_i\}_{i=1}^n$  and generate the initial 3D line via projecting it into the local 3D plane, where  $P$  is calculated with sampled points around the endpoints of image line. Due to the uncertainty of point cloud quality, it is hard to use a fixed distance threshold to define the inliers

## References