2021 European Conference on Computing in Construction
Online eConference
July 26-28, 2021

GENERATING RAILWAY GEOMETRIC DIGITAL TWINS FROM AIRBORNE LIDAR DATA

M.R.M.F. Ariyachandra and Ioannis Brilakis
Department of Engineering, University of Cambridge, Cambridge, UK

ABSTRACT

The cost of the railway digital twinning process counteracts the expected benefits of the resulting model. State-of-the-art methods yielded promising results, yet they could not offer large-scale digital twinning required over kilometres without forfeiting precision and manual cost. The proposed framework exploits the potential of railway topology to perform better when detecting and modelling the geometry of railway elements in railway point clouds with varying geometric patterns. Experiments on 18 km railway datasets illustrate that the framework improves the current cost and benefit ratio by reducing the overall twinning time by 90% without using any prior information.

INTRODUCTION

A Digital Twin (DT) is a digital replica of a real-world asset such as a building, a bridge, a railway or any other man-made asset of the built environment. A DT is based on massive, cumulative, real-time, real-world data measurements in multiple dimensions (Jones et al., 2020) and uses the information of a digital model across the entire lifecycle of infrastructure (Kaewunruen and Lian, 2019). The fundamental feature of DTs is their geometry, without which many DT applications do not exist. The authors use the adjective 'geometric' to highlight the DT with only geometry data, i.e. GDT. A GDT is generated using raw spatial data, such as Point Cloud Datasets (PCD)s as the initial input. This is useful for rail inspection maintenance and practices, which currently needs extensive costs and timescales.

The UK has the fastest-growing railway network in Europe, with an increase in passenger numbers of 40% expected by 2040 (Office of Rail and Road, 2020). However, railways are complex, safety-critical systems (Wilson et al., 2007) which unfortunately faces catastrophic risks such as derailments and collisions (European Railway Agency, 2020). While these incidents are considered to be rare, the total costs of railway accidents, including derailments, are estimated at £3.4 billion in 2018 (European Railway Agency, 2020). Maintenance, safety management and retrofitting are therefore vital operations in the life-cycle of existing rail infrastructure.

Yet, European and UK rail industries are partly built on antiquated legacy systems, becoming more difficult to maintain. The railway system in the UK is the oldest in the world (Lee, 1945) and a patchwork of overlapping designs built at different times (RailEngineer, 2020). Current maintenance processes can no longer cope with the increasing complexity of modern complex socio-technical systems (Zio, 2018) due to the absence of Information and Communication Technology (ICT) sector-level data management. This explains why there is a huge market demand for less labour-intensive railway maintenance techniques that can efficiently boost railway operations and productivity. Industry experts believe that the wider adoption of DTs will unlock 15-25% savings to the global infrastructure market by 2025 (Gerbert et al., 2016). For instance, the proposed DT for Maharashtra Metro in India is expected to provide real-time data for predictive maintenance strategies that are expected to save at least $222 million over 25 years of the railway's operational life (Davis, 2019).

DTs are of four levels and differ depending on their relationship with the physical asset's life cycle and the DTs' operators.

Level 1: Digital Twin Prototype (DTP). Design engineers produce DTP that describes the prototypical artefact for a new asset (Grieves and Vickers, 2016). Hence, DTP exists before there is a physical asset. This model contains design attributes such as initial designs, analyses, and processes generated by designers, subcontractors, and suppliers (Madni et al., 2019; Tao et al., 2017).

Level 2: Digital Twin Instances (DTI)s. The project management and facilities management teams continuously produce individual instances of the physical asset known as DTIs. These DTIs represent different virtual twin variants throughout the physical asset's life cycle once the asset has been built (Grieves and Vickers, 2016). Hence, DTI defines the physical asset's specific correspondences at any given point in time and uses it to explore the physical asset's behaviour under various what-if scenarios (Madni et al., 2019). Data capturing sensors (i.e. laser scanners, drones, photogrammetry) often update the DTI during alternative instances (Gardner et al., 2020).

Level 3 and 4: Digital Twin Environment (DTE)s known as Adaptive DT and Intelligent DT.

Adaptive DT is a high-level DT that offers an adaptive user interface to the physical and virtual twins (Madni et al., 2019). This user interface is sensitive to the end-users' preferences and priorities; hence, it can learn and prioritise the end-users' preferences for different instances (Alexopoulos et al., 2020; Bolender et al., 2021) with supervised machine learning techniques (Erkoyuncu et al., 2020). **Intelligent DT** is developed with both supervised and unsupervised machine learning techniques. It defines assets and patterns encountered in the operational environment by itself (Ashtari Talkhestani et al., 2019). It automatically updates itself and provides benefits and abilities beyond the explicitly defined information in the existing DT versions.

This paper only focuses on generating level 2 DTs and their most basic form representing DTIs' geometry. Hence, Geometric DT (GDT) corresponds to the DTI's geometric representation at all instances, including quantities, sizes, shapes, locations, and orientations of the asset's components in an Industry Foundation Classes (IFC) format. Soni (2016) reported that the total time to reconstruct such DTI of 0.5 m length track section using PCDs was between 20-40 minutes. Every DT generation hour saved can prevent critical failures or accidents so that continuous operations of railways can be achieved without impeding the national economy (Rail Delivery Group, 2014).

Defining Rail Infrastructure

Rail infrastructure refers to typical railway elements which constitute railway track structure, superstructure and masts (Figure 1). These structures include the most important and critical elements in railways (Dvořák et al., 2017; Urbancová and Sventeková, 2019). Also, the methods for generating the GDTs of the rest of the elements, including bridges, tunnels, platforms, signalling, etc. are beyond the scope of this work and have been studied separately by others (Kaewunruen et al., 2020; Lu and Brilakis, 2020; Tomar et al., 2020).

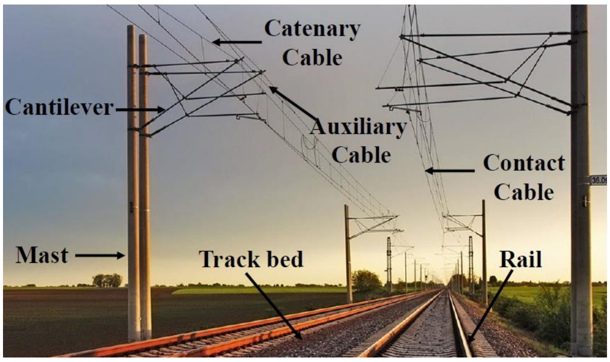

Figure 1: Railway elements

The authors outline the end-user requirements (EURs) of DTs and then provide a brief review of existing software solutions to check their degree of automation regarding the EURs.

End-User Requirements (EURs)

Developing detailed EURs for each instance of DTI is outside the scope of this study. The EURs define the expected information requirements and deliverables that the end-users will request DTs such as engineers, operators, decision-makers and owners. Still, the nature of the EURs depends on the complexity of the project, the experience, and the requirements of the end-users. Based on the interviews with several railway engineering consulting companies (i.e. Network Rail, HS2), railway asset management companies (i.e. Bentley Systems) and railway survey companies (i.e. Fugro), authors deduce the following EURs of a railway DT.:

- EUR 1: Component-level digital representation which includes the main structural component types of a sensed asset with a component-level resolution (Sacks et al., 2017).
- EUR 2: Component's explicit geometry representation (in as-is condition) and property sets (Borrmann and Berkhahn, 2018).
- EUR 3: Component's taxonomy. The components should be labelled by their element types (Koch and Konig, 2018).
- EUR 4: Component's implicit information such as structural relationships, material, cost, schedule etc. (Sacks et al., 2018).
- EUR 5: Component's damage information, including damage types, location, and orientation, along with the texture data (Hüthwohl et al., 2018).
- EUR 6: All above-listed EURs should be presented in a platform-neutral data format, such as Industry Foundation Classes (IFC) (Koch and Konig, 2018).

Current Practice of Railway Digital Twinning

Leading software vendors such as Autodesk, Bentley, Trimble, AVEVA and ClearEdge3D provide advanced commercial twinning solutions in two categories. The 1st category is BIM authoring tools (BATs), which are currently semi-automated at best. This includes but is not limited to Autodesk Revit, SketchUp, ArchiCAD, Descartes and EdgeWise.Yet, the automation provided by these software packages is tailored only to generic or predefined geometries; it is still far from being fully automatic (Agapaki and Brilakis, 2018). These packages can realise a certain degree of automation as the EUR 1 & 2 can be partially automated. The 2nd category: alignment-centred modelling tools (AMTs) such as Civil 3D and Power Rail Track has been developed for alignment-based assets such as roads and railways. For instance, OpenRail Designer has a certain degree of automation as EURs 1 & 2 have been partially automated by combining survey, design rules, and operational requirements to generate optimal geometry of the track on a 2D plane (Bentley Systems, 2018). However, AMTs' shape-creation method focuses only on continuous structures belonging to the alignment. The combination of BATs and AMTs (i.e. Civil 3D with Revit) does not work properly because there is no total integration between the two (Kwon et al., 2020). This lack of interoperability (EUR 6) between the existing software makes the modelling process challenging (Kenley et al., 2016). Other commercial applications cannot fully automate any one of the EURs. In addition, modellers need to enrich the resulting GDT with other explicit and implicit information, such as component's taxonomy (EUR 3), connectivity and aggregation (EUR 4), and defects (EUR 5) to meet the EURs. Then, all EURs need to be exported in IFC format (EUR 6).

The authors investigate the current railway twinning process using existing software packages (Ariyachandra and Brilakis, 2019). Up until the end of the manual operation, only EURs 1, 2, 3, and 6 are satisfied. 74% of of the labour hours are needed for the manual shape customisation and fitting process. The bottlenecks of digital twinning using current software applications are listed as follows:

- Existing software packages can semi-automatically extract standardised shapes in PCD. Yet, they cannot automatically extract non-canonical shapes, which are frequently required for the generation of DTIs.
- EUR 2 is manually achieved. In addition, the presence of occlusions and sparse data slows down the workflow and adds hours of adjustments.
- EURs 1, 3, & 6 can only be manually achieved, and EURs 4 & 5 are unavailable within existing software.
- There is no single software that can offer a one-stop DT generation solution. Modellers have to shuttle intermediate results in different formats back and forth between different software packages during the modelling process, giving rise to the possibility of information loss. In addition, this requires a substantial amount of manual modelling time due to the conversion needed for the hundreds of elements stretches over kilometres on the ground.

The next section provides a detailed review of the current state of research of GDT generation related to EURs 1, 2, 3, & 6, i.e. EURs required to generate railway GDTs. EURs 4 & 5 are beyond the scope of this paper.

STATE-OF-RESEARCH

The authors review the existing research methods by dividing them into two parts, namely, (1) object segmentation in PCDs (EURs 1 & 3) and (2) 3D solid model generation of the segmented point clusters (EURs 1 & 2).

Object Segmentation in Point Cloud Datasets

Point cluster segmentation can be achieved using different strategies, namely: (1) object detection, (2) class segmentation and (3) instance segmentation. The point cluster segmentation step within this study delivers labelled point clusters corresponding to railway elements (i.e. mast #1, trackbed #2). The following sections discuss the current state of research in current railway point cluster segmentation methods.

Mast Segmentation

The geometrical shape of the mast and other pole-like objects in railway PCDs (i.e. signal poles, traffic sign poles) are quite similar. Hence, the authors considered both masts and other pole-like object segmentation methods to derive the knowledge gaps. Readers can refer to the authors' previous work Ariyachandra and Brilakis, (2020a), for a comprehensive literature review of each of these methods.

Overhead Line Element (OLE) Segmentation

Methods for cable segmentation include: (1) Statistical analysis of PCDs based on height, density or number of pulses, etc. (2) Hough transform, and clustering based on 2D image processing (3) Supervised classification based on metrical and distribution features between points. Only two methods exist that segment cantilevers from PCD. The authors elaborate on the state-of-the-art literature on OLE segmentation methods for both railways and roads. Readers can refer to authors' previous work Ariyachandra and Brilakis, (2020b), for a detailed literature review of each of these methods.

Railway Track Structure Segmentation

A great deal of research has been focused on the segmentation of linear elements in railway environments. Track bed segmentation is the foundation for many subsequent railway track element segmentation methods. Readers can refer to the authors' previous work Ariyachandra and Brilakis (2019), for a comprehensive literature review of each of these methods.

3D Solid Model Generation

Recent advances in scanning technologies and large-scale 3D repositories have widened opportunities for 3D geometric data processing. Still, most of the resulting scanned data and the models in these databases are in unstructured formats such as PCDs. This low-level representation limits our ability to geometrically manipulate them due to the lack of structural information aligned with these formats. 3D modelling process has been introduced to convert the low-level digitised 3D data such as point clusters into high-level information-rich 3D structural formats (Li *et al.*, 2019). This is achieved by generating the 3D shape of the point cluster to describe and illustrate its shape, utilising computer graphic techniques. This high-level information-rich 3D representation is required to generate a meaningful DT of a real-world asset containing various attributes, including geometry, materials, and defects, among others. Only the shape and size have been considered in this research to describe the GDT of railway elements.

Building and infrastructure objects within the same PCD may significantly differ in terms of density, complexity and diversity. 3D shape reconstruction methods in the state of research are mainly two-fold. The 1st method is the 3D primitive arrangements which represent the scene as a combination of 3D primitives. These primitives include planes, cubes, lines and cylinders, among other shapes. 3D primitive model fitting to labelled point clusters and consequently generate DTPs and DTIs at the as-built stage is a solved problem. These techniques include Implicit Model fitting (Limberger and Oliveira, 2015; Schnabel et al., 2007), B-Rep fitting (Oesau et al., 2014; Valero and Cerrada, 2012), Constructive Solid Geometry fitting (Patil et al., 2017; Rusu et al., 2008; Xiao and Furukawa, 2012) and Swept Solid fitting (Budroni and Boehm, 2010; Laefer and Truong-hong, 2017; Lu and Brilakis, 2020). These 3D primitive descriptions remain a simplistic representation and often fail to model fine details and irregular shapes found in the various DTIs of the built environment. Yet, different variants and extended versions of these techniques can generate DTIs required for the maintenance stage (Lu and Brilakis, 2020).

The 2nd method of 3D shape reconstruction generates meshes for more irregular shapes often found in the DTIs of different instances of the project life cycle. Meshes are more appropriate over the 3D primitive arrangement for reconstructing 3D models of real-world objects. Yet, the generation of DTIs for different instances of infrastructure and built environment still remains a longstanding problem in computer vision and graphics (Zollhöfer et al., 2014). However, these two 3D shape representation methods have complementary advantages. The 1st method is advantageous as it leverages semantic knowledge and model compaction, while the 2nd method exploits detailed modelling and non-restricted usage for various shapes. Hence, researchers have considered merging both techniques to obtain a more rational 3D solid shape representation of the built environment scenes (Lafarge et al., 2010).

Gaps in Knowledge and Objectives

The problem of segmenting railway elements in the form of labelled point clusters from PCDs has yet to be solved (Ariyachandra and Brilakis, 2019, 2020a, 2020b). Likewise, the 3D solid model generation of segmented railway element point clusters to represent their geometry is still in its inception. Thus, the objective of this research is to devise, implement, and benchmark a framework that automates the generation of existing railway GDTs in IFC format.

PROPOSED FRAMEWORK

The proposed framework (Figure 2) uses railway topology knowledge as a guide to automatically generate GDTs of railway elements with no prior information. Railways are not perfectly straight or flat, and they usually contain varying horizontal and vertical elevations. Nevertheless, railways are a linear asset type; their geometric relations remain roughly unchanged, often over very long distances. Close inspection of railway PCDs validates this effect, with repeating geometrical features (Network Rail, 2018) such as:

- The geometric relationships among railway elements remain fairly unchanged along the railway corridor (Network Rail, 2018).
- The connections between railway masts and cables are placed in regular intervals (60 m intervals on average).
- The main axis of the railway masts (Z-axis) is roughly perpendicular to the rail track direction (X-axis) [error tolerance is 11° (Network Rail, 2018)].
- Masts are positioned as pairs throughout the rail track.

The authors leverage these four geometric features as railway topological relationships and use as assumptions when developing the proposed framework. The framework can deal with railway PCDs consisting of varying track slopes and curvatures and is effective in handling challenges inherited in PCDs such as occlusions, data gaps, and point diversity. This enables considerably improved large-scale object segmentation and modelling often required over kilometres without forfeiting precision and manual cost.

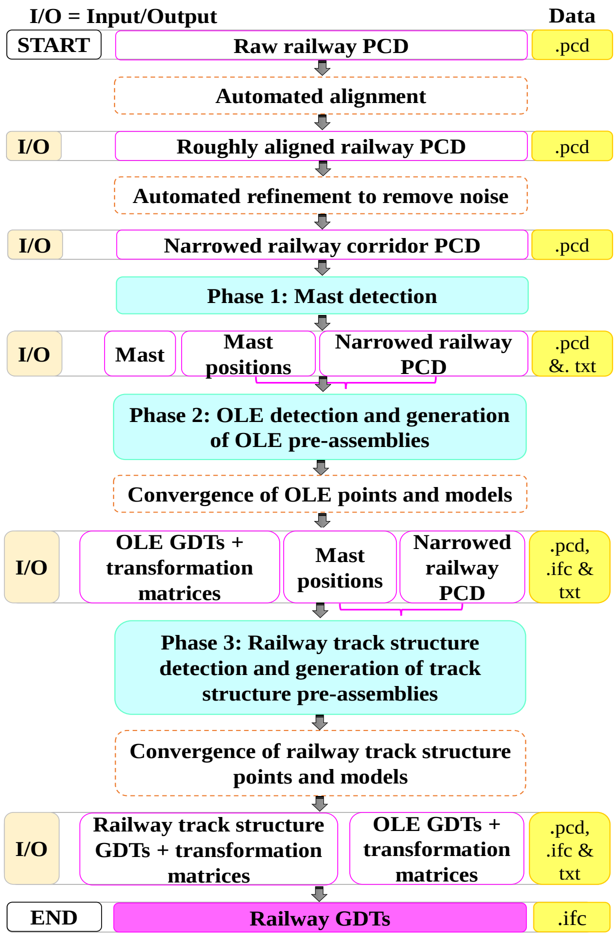

Figure 2: The proposed framework of this research

The framework is designed to twin only the typical double-track railways because they make up 70% of the existing railway network in the UK and Europe (Eurostat, 2019). The framework consists of three major phases which aim to meet EUR 1, 2, 3 & 6. The authors tested and validated this framework with three approximately 6 km (total 18 km) long PCDs (Dataset A, B, and C) obtained from the railway track located between 's-Hertogenbosch and Nijmegen in the Netherlands.

Phase 1: Mast Segmentation

Phase 1 segments masts in the form of point clusters. The input is raw railway PCD. The outputs are narrowed aligned railway PCD, labelled point clusters of masts (.pcd) and mast position coordinates (.txt).

The method initially uses an automated segmentation technique to align X, Y and Z axes of datasets parallel to the global reference system. This enables easy exploitation of the PCD features using various feature extraction algorithms because all features to be extracted in further steps lie in the global coordinate system. The result of this step yields a set of sub-bounding boxes (*SBB*)s which contains near-straight pieces of the railway PCD. Once the axes of *SBB*s are parallel to the global axes, the method gauges the centreline of each *SBB* $(CtrL_{SBB})$ and removes the vegetation and other noise from the railway PCD using $CtrL_{SBB}$. The method then segments masts as lines parallel to the global Z-axis using the Random Sample Consensus (RANSAC) algorithm with two refinement algorithms that differentiate masts from other pole-like objects.

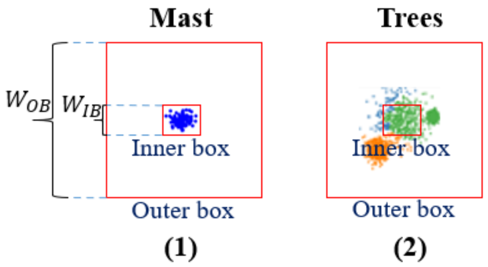

Figure 3: First refinement algorithm

The 1st refinement creates an inner box (*IB*) and an outer box (*OB*) around the XY projection of the detected lines on a ground plane removed PCD (Figure 3). The *OB* might contain other points surrounding the pole, usually caused by tree leaves, bushes, walls, etc. Next, the algorithm automatically calculates the point density ratio between *IB* and *OB* (D_{IB}/D_{OB}). This ratio is compared against a pre-defined threshold (R_D) which satisfies $0 < R_D < 1$, to filter masts from tree trunks. Using R_D and D_{IB}/D_{OB}, the method filters masts from other pole-like objects. Yet, when tree trunks, walls and rail bridges satisfy R_D this method recognises other pole-like objects as masts. To remedy the resulting outcome, the authors used a 2nd refinement (Figure 4). This algorithm takes railway geometric observations into account and limits the region of search to a certain radius from the first pair of masts. The 2nd refinement algorithm starts from the left side of the track and repeats over the spans between masts on the right side of the track. The details of the 2nd refinement algorithm can be found in the authors' previous work, Ariyachandra and Brilakis, (2020a). This step gives the segmented point clusters of the masts, along with the position coordinates (RM_{Cor}) and heights of individual clusters.

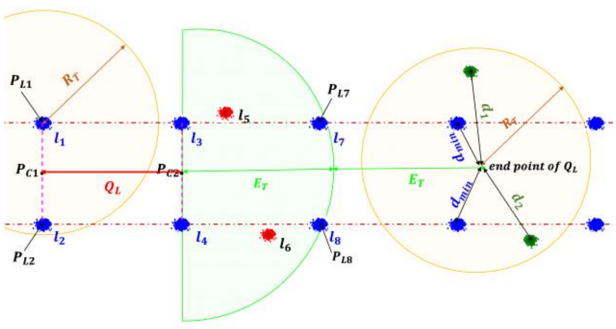

Figure 4: Second refinement algorithm

Phase 2: OLE Digital Twin Generation

This phase generates GDTs of OLE elements using the outputs of the previous step as the inputs. The outputs are OLE GDTs (.ifc) and transformation matrices (.txt).

OLE Element Segmentation

The method first extracts point clusters of the other OLE elements (except cables). This unit is hereafter known as 'C section'. The method uses a crop box filter (CBF_{rs}) to extract point clusters of C sections, which automatically extracts all points inside a given box. The limits [eq. (1) & (2)] of CBF_{rs} are defined relative to $RM_{Cor}(x, y, z,)$. This finally gives the resulting point segments of C sections (H_m – mast height, X_C, Y_C, Z_C CBF_{rs} coordinates).

$$X_{Cmin} = 0 - \frac{xRange}{2};\ Y_{Cmin} = 0 - \frac{yRange}{2};\ Z_{Cmin} = 0.23 \quad (1)$$

$$X_{Cmax} = \frac{xRange}{2};\ Y_{Cmax} = \frac{yRange}{2};\ Z_{Cmax} = Z_{Cmin} + H_m \quad (2)$$

The method then uses the RANSAC algorithm to extract point clusters of cables. The method initially computes bounding boxes (BB_{rs}) using RM_{Cor} along the track to crop the input PCD such that the resulting pieces are relatively straight enough for any further processing. The general RANSAC could not detect cables as lines due to few or no points on the cables. Hence, the method up-sampled the points on cables along the track direction to improve line detection. To determine the track direction, the method calculates the range between minimum and maximum of X and Y values of each BB_{rs} and sorts the general track direction along the X-axis if the X range > Y range and vice versa. Next, the method gets the XY projection of the cloud. This allows projecting the cables' catenary shapes into straight lines so that RANSAC can detect those cables despite their curved shapes. The method then segments cables as lines using RANSAC and classify cables based on the heights of the lines relative to the track structure. The segmented cables along with the previously extracted C sections, are hereafter known as 'Model A' (Figure 5 left).

Generate Dynamic IFC Models of the OLE System

The OLE elements are typically built with a larger wear allowance (Keenor, 2018) as they are either made of highly weathered metal or wood (Briginshaw, 2013). Hence, the elementwise damage and defects such as cracks, corrosion, cavities and potholes are seldom on OLE elements (Abubeker and Tilahun, 2016; Wu *et al.*, 2018) relative to generic building/infrastructure components such as walls, columns, piers etc. Thus, the geometric deviations between the pre-defined models for OLE elements versus as-is captured data become smaller for OLE elements. Hence, the method designs a parametric OLE system model, hereafter known as 'Model B', using standard railway electrification guidelines (Network Rail, 2018) to represent the geometry of the OLE elements. This model preserves the geometrical properties of the elements, such as angles between and relative distances compared to each element of the system. The orientation of Model B constantly changes from left to right along the track due to the stagger occur in the OLE system. Hence, the authors have created 10 variations of Model B, compatible with the left and right versions of the 5 types of OLE configurations that are widely used in the UK and European railways (Network Rail, 2018). Figure 5 (right) illustrates only one of those configurations. Note that on the actual model, two of these configurations are connected with cables.

The method defines each of the OLE elements using extruded area solid definition in IFC format with the cross-sectional dimensions given on Network Rail (2018) to define the 2D area profile for each element. The extruded area solid defined the extrusion of a 2D area; here defined as the section profile, by two attributes. One is the extruded direction, defining the direction in which the profile is to be swept. The other attribute is the distance over which the profile is to be swept. For each OLE element, the method defines these distances using either mean height (for masts) or length (for every other OLE element) obtained from the segmented point clusters. The extruded direction and relative angles are derived considering the position and the orientation of each element relative to RM_{Cor}.

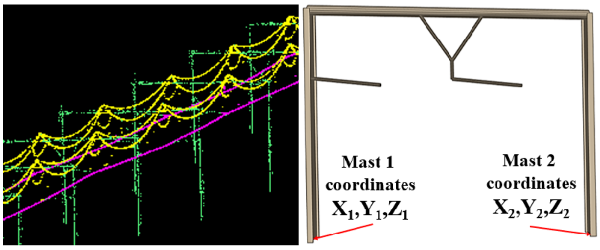

Figure 5: Left: A set of 'Model A's. Green - C section, Purple - Contact cables, Yellow - Other cables, Right: Model B

Convergence of Model A and Model B

The method uses Iterative closest point (ICP) algorithm to converge Model B to Model A automatically. The method set Model A as the reference cloud (R_C); is kept fixed while the left and right orientations of Model B are source clouds (S_C). The method first converts Model B into .pcd files, and then these S_C are transformed to find the best match with the R_C by minimising the distance ($RMSD$) between the two (3).

$$RMSD\ (T\ (S_C), \mu\ (R_C)) = \frac{\sum_C dist\ (r_j, T(s_i))^2}{|C|}, s_i, r_j \in C \quad (3)$$

where T – transformation, for a set of pairs of points $C = (s_i,\ r_j)$, $s_i\ \in\ S_C$, $r_j\ \in\ R_C$. Hence, by using ICP, the method first sorts the correct orientation (left or right) of OLE configuration and then converges the sorted model on to the correct position and finally gives transformation matrix, which provides the corresponding translation vector and rotation matrix of the Model B (IFC model) relative to Model A (point cluster). This step gives the segmented point clusters of the C sections (.pcd), cables (.pcd), pre-assemblies of OLE system elements (.ifc) along with their transformation matrices (.txt). These assemblies and transformation matrices are used to get the final railway superstructure GDT.

Phase 3: Railway Track Structure Digital Twin Generation

This phase generates GDTs of railway track structure element GDTs using the outputs of phase 1. The outputs of phase 3 are railway GDTs in .ifc format.

Rail and Track Bed Segmentation

Initially, the method uses the RANSAC plane detection algorithm to extract point clusters of the horizontal and quasi-horizontal ground planes. A pre-processing step is used before the RANSAC algorithm that divides the PCD dataset into sub boxes, approximately 60 m long (equals to the average span between two pairs of masts), using a crop box filter (CBF_{rt}) and RM_{Cor}. This allows segmenting rails and track beds points despite the track's varying horizontal and vertical elevations. Next, the method applies the RANSAC plane detection for each CBF_{rt} to detect points representing track structure elements.

The authors hypothesise that the only linear element on CBF_{rt} PCD now represents rails, while the rest of the CBF_{rt} PCD represents track bed. The previously calculated CBF_{rt} are now aligned along the track direction; yet, it is difficult to segment rail tracks parallel to the track direction, if there is a curvature occurred within any CBF_{rt}. Thus, the method automatically segments each CBF_{rt} such that the resulting pieces (SB_{rt}) are relatively straight enough to segment linear elements parallel to track direction. This step delivers 8 SB_{rt} s for every two pairs of masts (Figure 6). Then the method obtains the ground projection for each of these CBF_{rt} to improve RANSAC's robustness in segmenting linear elements.

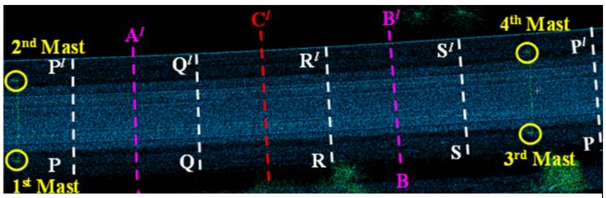

Figure 6: Obtaining SB_{rt}

Next, the method detects rails using RANSAC as lines parallel to track direction with an additional radius neighbour search to include any missing points during RANSAC detection. However, segmented linear elements at this stage represent both rails and other linear elements along the rail track direction in railway PCDs. The method uses a point-based calculation method to differentiate point clusters of rails from other linear elements. The authors experimentally define two thresholds: (1) D_{r1} - by calculating the ratio between the number of points per other linear elements such as walls and fences over the number of points per rail point cluster and (2) D_{r2} - by calculating the ratio between the number of points per rail point cluster over the number of points per other linear elements such as lines on the trackbed and ground along the track direction. Rails are now filtered from other linear elements using D_{r1} and D_{r2}. The use of ratios over point density provides the robustness required for the method; therefore, it will work for any input datasets despite their density. Once the method removes the segmented rail point clusters from the CBF_{rt} PCD, the CBF_{rt} PCD now contains track bed points only in line with the authors' initial hypothesis. Following the same notation as in OLE element segmentation, the segmented rails and track beds are hereinafter known as 'Model A'.

Dynamic IFC Models of the Railway Track Structure and Convergence

The method generates parametric models of different rail profiles and track bed profiles that exist in the UK and Europe railways (Network Rail, 2018) (Model B), following the same procedure explained in OLE IFC model generation. Next, the method uses the same convergence procedure explained previously to automatically select the optimum rail/track bed profile and converge Model B to Model A. The method then moves the .ifc format of the Model Bs (resulting railway superstructure and substructure elements) to the correct position using the resulting transformation matrices. Finally, it merges all units (all railway elements) into one file to get the final IFC model of the railway GDT (Figure 7). (The authors have not illustrated the graphs representing calculations for the parameters used in the framework due to limited space).

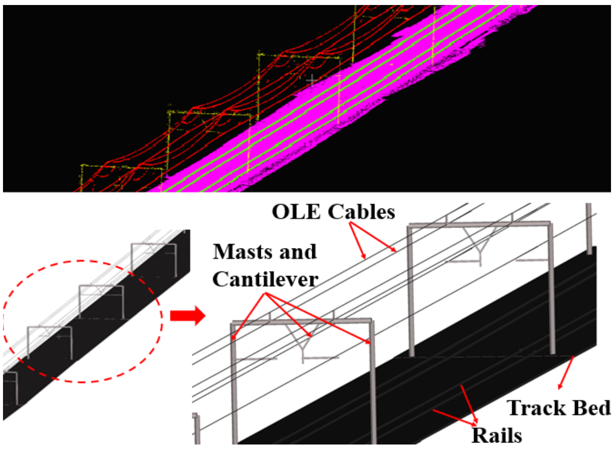

Figure 7: Top: Detected point clusters of railway elements, Bottom: Final railway GDTs

EXPERIMENTS AND EVALUATION

Ground Truth Data

The authors manually generated two sets of Ground Truth (GT) datasets consist of three sub-datasets each per one railway PCD;

- GT A: Manually extracted point clusters of railway elements from raw railway PCD. They are used to compare against the automatically segmented point clusters of railway elements.
- GT B: Manually created railway GDTs and used to compare against automated railway element GDTs.

The authors implemented the solution with the point cloud library (PCL) version 1.8.0 using C++ on Visual Studio 2017, on a laptop (Intel Core i7-8550U 1.8GHz CPU, 16 GB RAM, Samsung 256GB SSD).

Evaluation of Object Segmentation

The authors gauged the segmentation accuracy using performance metrics; precision (Pr), recall (R) and F1 score (F1) as (4), (5) and (6). (TP – correctly segmented railway elements, FN – railway elements were not segmented, FP – other objects were segmented as railway elements). The average segmentation accuracies are given in Table 1.

$$Pr = TP/(TP + FP) \quad (4)$$

$$R = TP/(TP + FN) \quad (5)$$

$$F1 = 2 * (Pr * R)/(Pr + R) \quad (6)$$

Table 1: Performance metrics for object detection

Railway element	F1 scores for datasets			
	A	B	C	Avg.
Mast	92.0%	88.8%	88.6%	90.1%
Cables	84.5%	84.0%	65.6%	78.6%
Other OLE	91.1%	86.0%	86.7%	88.6%
Rails	89.5%	81.5%	83.8%	85.5%
Track bed	90.1%	87.0%	86.7%	88.4%

Evaluation of Railway GDTs

The authors used cloud-to-cloud distance evaluation to detect changes between GT B and the automated ones. Initially, the authors converted the GT B and the automated GDTs into .pcd files. The evaluation method computed the Root Mean Square Error (RMSE) between each unit of automated GDT of railway elements and the corresponding GT B model. The average model distance between the two for all 18 km was 3.82 cm RMSE for railway superstructure, 3.38 cm RMSE for rails and 2.72 cm RMSE for track beds. The proposed twinning framework reduces manual twinning time by 90.2%. This implies the proposed method outperforms the manual operation.

CONCLUSIONS

This paper presents a framework for automated GDT generation of existing railways using airborne PCD to meet EURs 1,2,3, and 6. It is the first to use railway topology knowledge as a guide to automatically generate railway GDTs in IFC-INFRA format with no prior information. Experiments on the 18 km railway dataset demonstrate that this framework is more consistent, less liable to human errors and a more robust solution for railways that stretches over kilometres on the ground. Hence, this framework outperforms the current state-of-the-art methods for generating the geometry of DTIs as it does not forfeit precision and manual cost. This is a huge leap over the current railway digital twinning practice and allows rapid adoption of GDTs for railways.

ACKNOWLEDGEMENTS

The authors express their gratitude for Fugro NL Land B.V., who provided data for evaluation. This research is funded by Cambridge Commonwealth, European & International Trust and Bentley Systems UK Plc. The authors gratefully acknowledge their support. Any opinions, findings and conclusions expressed in this paper are those of the authors and do not necessarily reflect the views of the institutes mentioned above.

REFERENCES

Abubeker, S. and Tilahun, D. (2016), "Sliding wear and corrosion resistance of train overhead contact wire", *EEA*, Vol. 34 No. June.

Agapaki, E. and Brilakis, I. (2018), "State-of-practice on As-Is Modelling of Industrial Facilities", in Smith, I. and Domer, B. (Eds.), *Advanced Computing Strategies for Engineering. EG-ICE 2018. Lecture Notes in Computer Science*, Springer, Lausanne, Switzerland, available at:https://doi.org/10.1007/978-3-319-91635-4_6.

Alexopoulos, K., Nikolakis, N. and Chryssolouris, G. (2020), "Digital twin-driven supervised machine learning for the development of artificial intelligence applications in manufacturing", *International Journal of Computer Integrated Manufacturing*, Taylor & Francis, Vol. 33 No. 5, pp. 429–439.

Ariyachandra, M.R.M.F. and Brilakis, I. (2019), "Understanding the challenge of digitally twinning the geometry of existing rail infrastructure", *12th FARU International Research Conference (Faculty of Architecture Research Unit)*, Faculty of Architecture Research Unit, Univeristy of Moratuwa, Colombo, Sri Lanka, pp. 25–32.

Ariyachandra, M.R.M.F. and Brilakis, I. (2020a), "Detection of Railway Masts in Air-Borne LiDAR Data", *Journal of Construction Engineering and Management*, Vol. 146 No. 9, available at:https://doi.org/10.1061/(ASCE)CO.1943-7862.0001894.

Ariyachandra, M.R.M.F. and Brilakis, I. (2020b), "Digital Twinning of Railway Overhead Line Equipment from Airborne LiDAR Data", *Proceedings of the 37th International Symposium on Automation and Robotics in Construction (ISARC)*, pp. 1270–1277.

Ashtari Talkhestani, B., Jung, T., Lindemann, B., Sahlab, N., Jazdi, N., Schloegl, W. and Weyrich, M. (2019), "An architecture of an Intelligent Digital Twin in a Cyber-Physical Production System", *At-Automatisierungstechnik*, Vol. 67 No. 9, pp. 762–782.

Bentley Systems. (2018), *Bentley Systems ' New OpenRail Is First to Advance BIM for the Full Rail and Transit Lifecycle*.

Bolender, T., Bürvenich, G., Dalibor, M., Rumpe, B. and Wortmann, A. (2021), "Self-Adaptive Manufacturing with Digital Twins", available at: http://arxiv.org/abs/2103.11941.

Borrmann, A. and Berkhahn, V. (2018), *Principles of Geometric Modeling*, edited by Borrmann A., König M., Koch C., B.J., Springer International Publishing AG, available at:https://doi.org/https://doi.org/10.1007/978-3-319-92862-3_2.

Briginshaw, D. (2013), "Weathering steel shows promise for OLE structures", *International Railway Journal*, available at: https://www.railjournal.com/in_depth/weathering-steel-shows-promise-for-ole-structures (accessed 15 March 2020).

Budroni, A. and Boehm, J. (2010), "Automated 3D Reconstruction of Interiors from Point Clouds", *International Journal of Architectural Computing*, Vol. 8 No. 1, pp. 55–73.

Davis, M. (2019), "Realizing the Potential for Digital Twins in Rail", *Railway Age*, available at: https://www.railwayage.com/analytics/realizing-the-potential-for-digital-twins-in-rail/ (accessed 9 September 2020).

Dvořák, Z., Sventeková, E., Řehák, D. and Čekerevac, Z. (2017), "Assessment of Critical Infrastructure Elements in Transport", *Procedia Engineering*, Vol. 187 No. June, pp. 548–555.

Erkoyuncu, J.A., del Amo, I.F., Ariansyah, D., Bulka, D., Vrabič, R. and Roy, R. (2020), "A design framework for adaptive digital twins", *CIRP Annals*, Vol. 69 No. 1, pp. 145–148.

European Railway Agency. (2020), *Report on Railway Safety and Interoperability in the EU*, available at:https://doi.org/10.2821/30980.

Eurostat. (2019), "Railway transport - length of lines, by number of tracks", available at: https://data.europa.eu/euodp/en/data/dataset/zibSXeaTsrM57MTXruQVvw (accessed 2 February 2018).

Gardner, P., Dal Borgo, M., Ruffini, V., Hughes, A.J., Zhu, Y. and Wagg, D.J. (2020), "Towards the Development of an Operational Digital Twin", *Vibration*, Vol. 3 No. 3, pp. 235–265.

Gerbert, P., Castagnino, S., Rothballer, C., Renz, A. and Filitz, R. (2016), *Digital in Engineering and*

Construction: The Transformative Power of Building Information Modeling, available at: http://futureofconstruction.org/content/uploads/2016/09/BCG-Digital-in-Engineering-and-Construction-Mar-2016.pdf.

Grieves, M. and Vickers, J. (2016), "Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems", *Transdisciplinary Perspectives on Complex Systems: New Findings and Approaches*, No. August, pp. 85–113.

Hüthwohl, P., Brilakis, I., Borrmann, A. and Sacks, R. (2018), "Integrating RC Bridge Defect Information into BIM Models", *Journal of Computing in Civil Engineering*, Vol. 32 No. 3, p. 04018013.

Jones, D., Snider, C., Nassehi, A., Yon, J. and Hicks, B. (2020), "Characterising the Digital Twin: A systematic literature review", *CIRP Journal of Manufacturing Science and Technology*, CIRP, Vol. 29, pp. 36–52.

Kaewunruen, S. and Lian, Q. (2019), "Digital twin aided sustainability-based lifecycle management for railway turnout systems", *Journal of Cleaner Production*, Elsevier Ltd, Vol. 228, pp. 1537–1551.

Kaewunruen, S., Peng, S. and Phil-Ebosie, O. (2020), "Digital twin aided sustainability and vulnerability audit for subway stations", *Sustainability (Switzerland)*, Vol. 12 No. 19, pp. 1–17.

Keenor, G. (2018), "Overhead Line Electrification for Railways", *Garry Keenor*.

Kenley, R., Harfield, T. and Behnam, A. (2016), "BIM Interoperability Limitations: Australian and Malaysian Rail Projects", *MATEC Web of Conferences*, Vol. 66, p. 00102.

Koch, C. and Konig, M. (2018), "Data Modeling", *Building Information Modeling*, Springer International Publishing AG, part of Springer Nature 2018, pp. 43–62.

Kwon, T.H., Park, S.I., Jang, Y.H. and Lee, S.H. (2020), "Design of railway track model with three-dimensional alignment based on extended industry foundation classes", *Applied Sciences (Switzerland)*, Vol. 10 No. 10, available at:https://doi.org/10.3390/app10103649.

Laefer, D.F. and Truong-hong, L. (2017), "Toward automatic generation of 3D steel structures for building information modelling", *Automation in Construction*, No. November, available at:https://doi.org/10.1016/j.autcon.2016.11.011.

Lafarge, F., Keriven, R. and Brédif, M. (2010), "Insertion of 3-D-primitives in mesh-based representations: Towards compact models preserving the details", *IEEE Transactions on Image Processing*, Vol. 19 No. 7, pp. 1683–1694.

Lee, C.E. (1945), "The World's Oldest Railway", *Transactions of the Newcomen Society*, Vol. 25 No. 1, pp. 141–162.

Li, L., Sung, M., Dubrovina, A., Yi, L. and Guibas, L.J. (2019), "Supervised fitting of geometric primitives to 3D point clouds", *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2647–2655.

Limberger, F.A. and Oliveira, M.M. (2015), "Real-time detection of planar regions in unorganized point clouds", *Pattern Recognition*, Elsevier, Vol. 48 No. 6, pp. 2043–2053.

Lu, R. and Brilakis, I. (2020), "A benchmarked framework for geometric digital twinning of slab and beam-and-slab bridges", Vol. 172 No. 2019, pp. 3–18.

Madni, A., Madni, C. and Lucero, S. (2019), "Leveraging Digital Twin Technology in Model-Based Systems Engineering", *Systems*, Vol. 7 No. 1, p. 7.

Network Rail. (2018), "Catalogue of Network Rail Standards", available at: https://www.networkrail.co.uk/wp-content/uploads/2018/04/Network-Rail-Standards-Catalogue.pdf (accessed 5 January 2019).

Oesau, S., Lafarge, F. and Alliez, P. (2014), "Indoor scene reconstruction using feature sensitive primitive extraction and graph-cut", *ISPRS Journal of Photogrammetry and Remote Sensing*, International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS), Vol. 90, pp. 68–82.

Office of Rail and Road. (2020), "Office of Rail and Road (ORR)", available at: https://dataportal.orr.gov.uk/ (accessed 7 October 2020).

Patil, A.K., Holi, P., Lee, S.K. and Chai, Y.H. (2017), "An adaptive approach for the reconstruction and modeling of as-built 3D pipelines from point clouds Automation in Construction An adaptive approach for the reconstruction and modeling of as-built 3D pipelines from point clouds", *Automation in Construction*, Elsevier B.V., Vol. 75 No. November 2018, pp. 65–78.

Rail Delivery Group. (2014), *What Is the Contribution of Rail to the UK Economy?*, available at: https://www.oxera.com/wp-content/uploads/2018/07/Contribution-of-rail-to-the-UK-economy-140714.pdf.pdf.

RailEngineer. (2020), *Why Rail Electrification Is Key?*

Rusu, R.B., Marton, Z.C., Blodow, N., Dolha, M. and Beetz, M. (2008), "Towards 3D Point cloud based object maps for household environments", *Robotics and Autonomous Systems*, Elsevier B.V., Vol. 56 No. 11, pp. 927–941.

Sacks, R., Eastman, C., Lee, G. and Teicholz, P. (2018), *BIM Handbook*, John Wiley & Sons, Inc., available at:https://doi.org/10.1002/9781119287568.

Sacks, R., Ma, L., Yosef, R., Borrmann, A., Daum, S. and

Kattel, U. (2017), "Semantic Enrichment for Building Information Modeling: Procedure for Compiling Inference Rules and Operators for Complex Geometry", *Journal of Computing in Civil Engineering*, Vol. 31 No. 6, p. 04017062.

Schnabel, R., Wahl, R. and Klein, R. (2007), "Efficient RANSAC for Point-Cloud Shape Detection", *Computer Graphics Forum*, Vol. 26 No. 2, pp. 214–226.

Soni, A. (2016), *Non-Contact Monitoring of Railway Infrastructure with Terrestrial Laser Scanning and Photogrammetry at Network Rail*, available at: https://discovery.ucl.ac.uk/id/eprint/1477352/.

Tao, F., Cheng, J., Qi, Q., Zhang, M., Zhang, H. and Sui, F. (2017), "Digital twin-driven product design, manufacturing and service with big data", *International Journal of Advanced Manufacturing Technology*, The International Journal of Advanced Manufacturing Technology, Vol. 94 No. 9–12, pp. 3563–3576.

Tomar, R., Pisek J, Sprengel, H., Isleyen, E., Duzgun, S. and Rostami, J. (2020), "Digital twin of tunnel construction for safety and efficiency", *Tunnels and Underground Cities: Engineering and Innovation Meet Archaeology, Architecture and Art*, available at:https://doi.org/10.4324/9781003031635-71.

Urbancová, Z. and Sventeková, E. (2019), "Assessing vulnerability of key elements of railway infrastructure", *Transportation Research Procedia*, Elsevier B.V., Vol. 40, pp. 1597–1603.

Valero, E. and Cerrada, C. (2012), "Automatic Method for Building Indoor Boundary Models from Dense Point Clouds Collected by Laser Scanners", *Sensors*, Vol. 12, pp. 16099–16115.

Wilson, J.R., Farrington-Darby, T., Cox, G., Bye, R. and Hockey, G.R.J. (2007), "The railway as a socio-technical system: Human factors at the heart of successful rail engineering", *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, Vol. 221 No. 1, pp. 101–115.

Wu, G., Wu, J., Wei, W., Zhou, Y., Yang, Z. and Gao, G. (2018), "Characteristics of the sliding electric contact of pantograph/contact wire systems in electric railways", *Energies*, Vol. 11 No. 1, available at:https://doi.org/10.3390/en11010017.

Xiao, J. and Furukawa, Y. (2012), "Reconstructing the World ' s Museums", *International Journal of Computer Vision*, Vol. 110 No. 3, pp. 668–681.

Zio, E. (2018), "The future of risk assessment", *Reliability Engineering and System Safety*, Elsevier Ltd, Vol. 177 No. June 2017, pp. 176–190.

Zollhöfer, M., Nießner, M., Izadi, S., Rehmann, C., Zach, C., Fisher, M., Wu, C., *et al.* (2014), "Real-time non-rigid reconstruction using an RGB-D camera", *ACM Transactions on Graphics*, Vol. 33 No. 4, available at:https://doi.org/10.1145/2601097.2601165.