Original Article

Institution of
MECHANICAL
ENGINEERS

Journal of
RAIL AND
RAPID TRANSIT

Use of a 3D model to improve the performance of laser-based railway track inspection

Jiaqi Ye, Edward Stewart and Clive Roberts

Abstract

In recent decades, 3D reconstruction techniques have been applied in an increasing number of areas such as virtual reality, robot navigation, medical imaging and architectural restoration of cultural relics. Most of the inspection techniques used in railway systems are, however, still implemented on a 2D basis. This is particularly true of track inspection due to its linear nature. Benefiting from the development of sensor technology and constantly improving processors, higher quality 3D model reconstructions are becoming possible which push the technology into more challenging areas. One such advancement is the use of 3D perceptual techniques in railway systems. This paper presents a novel 3D perceptual system, based on a low-cost 2D laser sensor, which has been developed for the detection and characterisation of physical surface defects in railway tracks. An innovative prototype system has been developed to capture and correlate the laser scan data; dedicated 3D data processing procedures have then been developed in the form of three specific defect-detection algorithms (depth gradient, face normal and face-normal gradient) which are applied to the 3D model. The system has been tested with rail samples in the laboratory and at the Long Marston Railway Test Track. The 3D models developed represent the external surface of the samples both laterally (2D slices) and longitudinally (3D model), and common surface defects can be detected and represented in 3D. The results demonstrate the feasibility of applying 3D reconstruction-based inspection techniques to railway systems.

Introduction

With increasing train speed, axle load and traffic density, rail defects are becoming more serious than ever. According to the European Railway Safety Agency,[1] there was, on average, a derailment or collision at least every second day in the European Union in the period 2010–2011, most of which were caused by defects in the track. In the UK, the rail network open to traffic is 15,753 km (9788 miles) and this is constantly increasing with the growth of passenger and freight demand, which means that railway track inspection is a significant concern to the industry.[2]

In the last few decades, an increasing number of non-contact techniques have been used for rail inspection. These techniques are generally based on measuring technologies such as ultrasound, eddy currents and lasers. A comprehensive survey of the existing rail inspection techniques by Papaelias et al. indicated that laser-based inspection is the most applicable technology for detection of surface defects.[3] Compared with conventional rail inspection methods,

laser-based inspection methods are truly non-contact, with the laser equipment up to 100 mm above the rail line. As a form of machine vision technology, numerous studies about laser-based approaches to assist rail inspection have been carried out both by research groups at universities and industrial railway organisations. For example, Network Rail's New Measurement Train (NMT) is equipped with laser track scanners to get a precise 2D rail-head profile; MERMEC developed a series of laser-based measurement systems that can ensure a 2D full-rail profile and evaluate the rail surface conditions by comparison with the standard profile. 2D imaging has its advantages, such as high resolution and high efficiency with

Centre for Railway Research and Education, University of Birmingham, Birmingham, UK

Corresponding author:
Jiaqi Ye, Centre for Railway Research and Education, University of Birmingham, Gisbert Kapp Building, Pritchatts Road, Edgbaston, Birmingham B15 2TT, UK.
Email: railCM@contacts.bham.ac.uk

the capability to inspect at train speed; however, the geometrical characteristics of some defects like the particular cracks, squats and partial deformation of crossing noses are longitudinal in nature, and thus difficult to detect using 2D techniques. This is significant for defect evaluation such as maintenance suggestions and remaining life prediction of components suffering from fatigue damage.

In recent years, the quality of 3D imaging techniques has been substantially improved with the development of sensor technology and processing capabilities. Applications involve many areas such as virtual reality, medical imaging and architectural restoration of cultural relics.[4] Accordingly, using 3D techniques in railway systems is considered to be the next logical improvement. The proposed laser-based system described in this paper explores the feasibility of applying 3D techniques into the area of rail inspection. The main goal is that the system can not only measure the 2D transverse profile but can also allow common longitudinal surface defects with different geometrical characteristics to be detected and characterised more comprehensively than using 2D techniques.

System description

Dedicated rail components

The 'rail defects' mentioned above offer a general identification. To answer questions like 'What kinds of rail defects are we concerned with?' and 'What methods could we use?', a systematic understanding of the geometry and some terminology of the rail is indispensable. In this section, dedicated rail components are divided into two parts, plain track and crossing noses.

Plain track. The sectional profile in Figure 1 shows that a UIC 60 rail profile consists of rail foot, rail web and rail head.

As has been summarised by Cannon et al.,[5] rail defects can be divided into three categories: (i) rail manufacturing defects, (ii) defects resulting from improper handling, use and installation and (iii) defects as a result of rolling contact fatigue (RCF) crack growth. Defects belonging to the first two categories have been greatly reduced globally with the improvement of materials and rail-making industries.[5] However, defects in the third category are difficult to control because they normally originate from the cyclical loading and long-term impact from rolling stock.

In railway systems, RCF is used to describe a range of defects caused by the development of excessive shear stresses at the rail contact interface. It is known that the gauge corner region, running surface and field corner region are the areas that make contact with the wheel, and thus defects in the third category that endanger the movement of the train are mostly centred on these regions which are the major concern to the railway industry.[6]
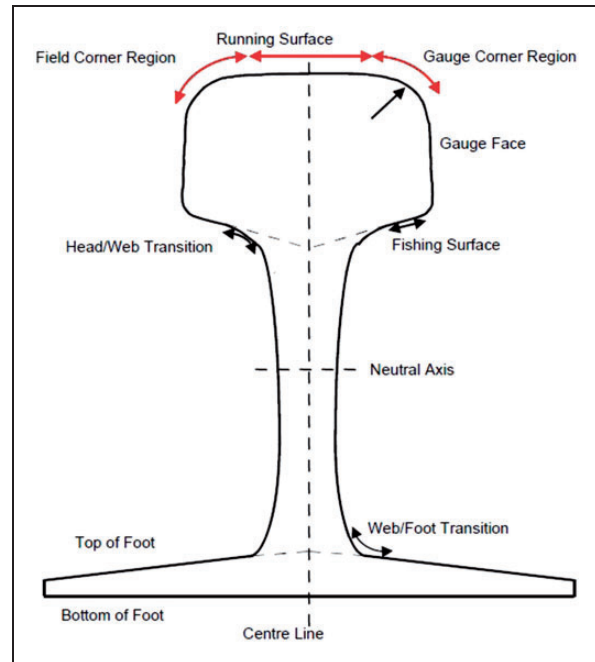

Figure 1. Transverse profile of plain track.

Crossing noses. As one of the most important components of railway infrastructure, switches and crossings (S&C) are used to guide trains from one track to another and enable lines to cross paths. The crossing nose is one of the key components of S&C; it can be either fabricated from two machined rails joined together or cast as a single unit.

Modern crossings are now cast from manganese steel, which is an advanced alloy that gets harder with use. However, an increase in axle loads and train speed creates larger lateral forces as they change course, and these forces can cause wear, RCF and deformation.[7] Flaws in crossings may eventually lead to grave consequences such as train derailments. For example, due to poor maintenance of S&C, a train derailed at Potters Bar in the UK on 10 May 2002, and seven people died.[8] Accordingly, the safe and reliable operation of crossing noses must be assured by high levels of routine inspection and maintenance.

Motivation and approaches

On the basis of an initial literature review, the system was proposed to be laser-based, and the dedicated rail components are the RCF defects in the rail contact interfaces of plain track and crossing noses. As a non-contact technology, the optical property of the laser and analysis of the existing systems both in the railway industry and other areas motivated the approaches and objectives of the proposed system.

Laser equipment. A laser generally refers to a device that emits light by means of optical amplification based on stimulated emission.[9] The property of optical coherence allows a laser light to be focused to a tight spot, and also

makes a laser beam stay highly collimated even after long-distance transmission.[10] According to the geometrical characteristics (laser dot, laser line or laser arrays) of the laser light, laser scanners can be categorised as 1D, 2D or 3D.[11] 2D rail profiles can be captured due to the linear nature of the tracks, so 2D laser scanners are generally the most common equipment for track measurement. Faulty track components such as surface defects, missing clips and deformation can be detected with a high-precision 2D digital profile.[3]

Existing laser-based systems. In practice, laser-based inspection methods often incorporate vision-based inspection. For example, profile data are analysed using image processing methods, or combine what the camera sees with laser inspection to improve accuracy. Lorente et al. developed a system combining 2D laser profile data with 2D depth images.[1] By applying derivative of Gaussian (DOG) edge detection and RANSAC-based line fitting, the track gauge can be calculated (Figure 2(a)). The system can also detect missing clips with 2D depth images using 2D template-matching algorithms. Research by Zhou et al. presented a rail profile alignment algorithm which compares the 2D transverse profile of the rail extracted from a 3D model with the 2D laser profile so that rail wear can be detected (Figure 2(b)).[12]

Generally, most of the existing inspection methods are implemented on a 2D basis. GRAW developed an optical system called SCORPION, for 3D measurement of rails and turnouts, but the measurement is actually implemented in 2D, with the 3D model discrete in the longitudinal direction There exist, therefore, some opportunities to improve the performance of laser-based rail inspection.

(a) Increase the dimension of inspection from 2D to 3D;
(b) Represent geometrical characteristics of rails and rail surface defects;
(c) Detect and characterise rail surface defects in 3D.

Design of the 3D perceptual system

To achieve 3D perception, the first step is to acquire 3D data points, known as a 3D point cloud. Commercial solutions are generally based on 3D laser scanners developed by a few companies (Cyra Technologies, Zoller & Frölich, Callidus Precision Systems, among others), but the robustness of these devices is unacceptable for most applications, and the cost is still high (from £20,000 to £80,000).[13] Therefore, many existing studies have been done to develop low-cost 3D scanners using different methodologies and technologies, such as the 1D optical distance sensor-based measuring systems developed by Arnulfo León Reyes et al.[14] Paulo Dias developed a low-cost 3D range scanner for 3D reconstruction of real-world scenes.[13] A mechanically stable tilt unit was designed to rotate the 2D laser range finder.

This project exploits similar methods to align high-resolution profiles from a low-cost 2D laser. The diagram in Figure 3 illustrates the basic structure of the system. The 2D profile data stream from the laser is sampled at a specified frequency. To obtain information in the third dimension, to gather a 3D data set, a pair of 1D draw-wire sensors is used to increase the capability of the 2D laser. And thus, the stream of 2D profiles can be aligned in a global coordinate system according to the position of the laser at the time of scanning, with which a 3D point cloud representing the surface of the target object can then be generated.

System integration

The most essential imaging unit used in this project is a scanCONTROL 2900-100/BL from Micro-Epsilon. Because of their compact size, high profile frequency (up to 200 Hz) and high resolution (1280 points/profile), scanCONTROL lasers are ideally suited to both static and dynamic measurements, such as profile measurement and defect recognition. This series of 2D laser scanners (profile sensors) uses the laser triangulation principle. As can be seen in Figure 4(b),
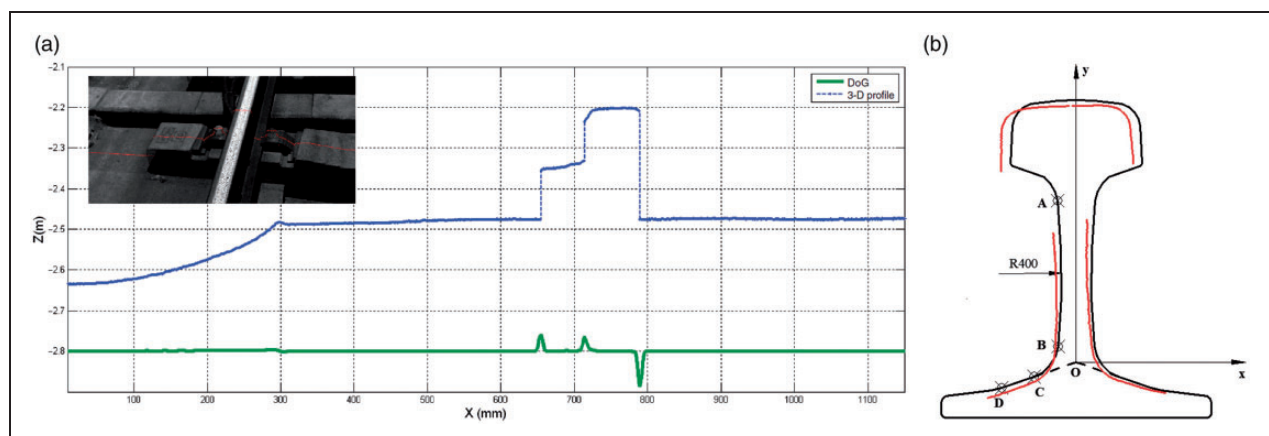
Figure 2. Laser-based rail inspection methods: (a) rail gauge and rail fastener detection; (b) standard rail profile and extracted profile.
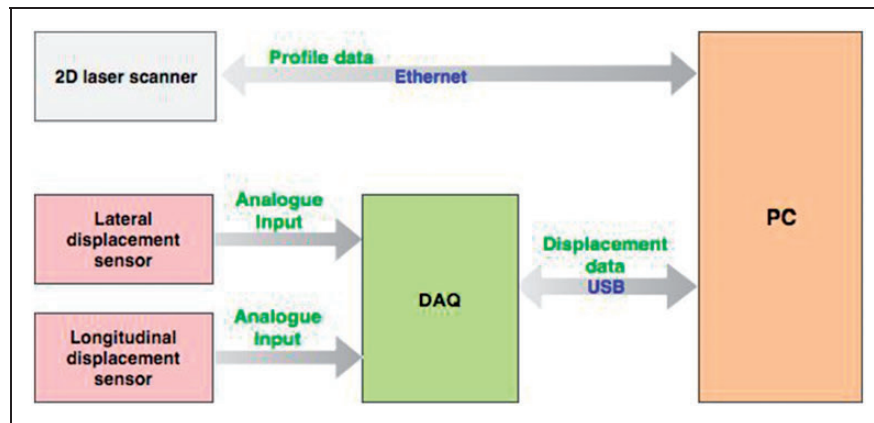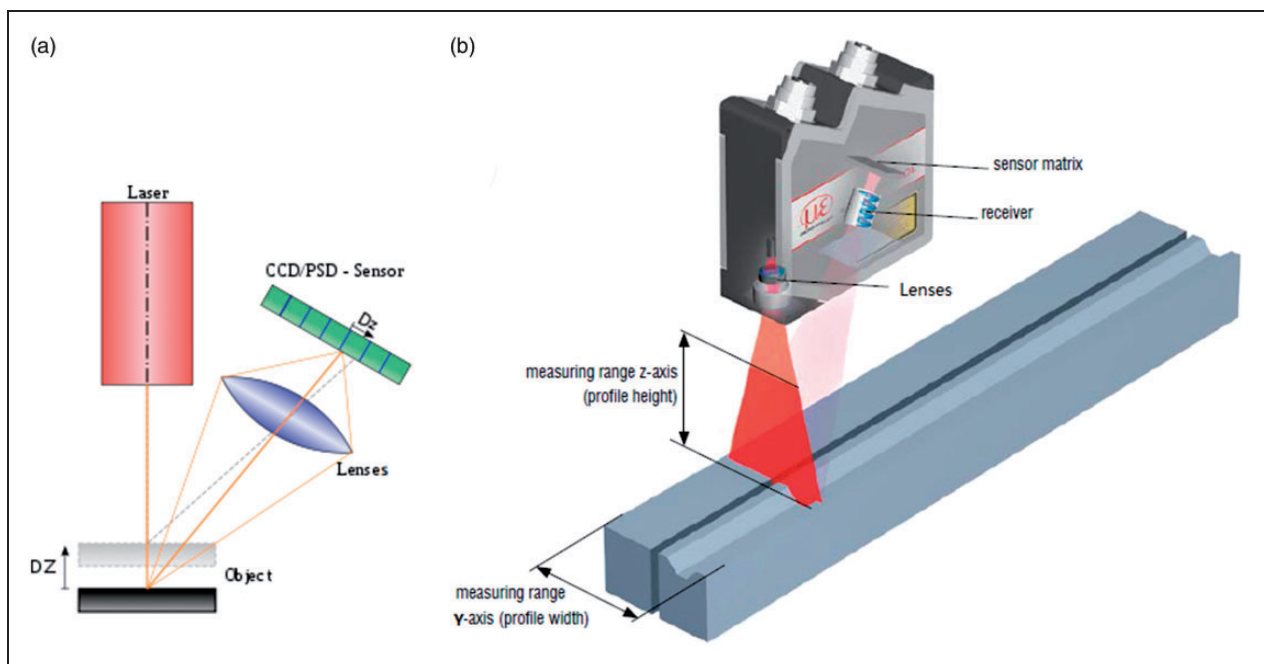
Figure 3. System mechanism.

Figure 4. Measuring principle of the laser: (a) laser triangulation; (b) scanCONTROL 2D laser scanner.

the distance information (Z axis) and the position alongside the laser line (Y axis) can be calculated from the position of the laser spot on the sensor matrix and the camera image.[15] These measured data are then output in a 2D coordinate system which is fixed with reference to the laser. Therefore, it is possible to do 3D measurement with this laser when moving the object or sensor.

Due to a high data rate, Gigabit Ethernet is the standard connection between the laser and the PC. For the analogue output from draw-wire sensors, NI USB-6210 from National Instruments is used. The analogue input is converted to a digital signal and transmitted to the PC. Given that the laser has a built-in timer, the only requirement for data synchronisation is to sample the sensors during laser scanning and provide the profile with the corresponding spatial location. So, software-timed acquisition

was chosen for this project. The sampling commands are sent by the C++ program, which do not increase the complexity of the hardware system.

Figure 5 shows the mechanical frame designed for the system. For periodic measurements, a simple four-footed frame with slide rails in both lateral and longitudinal directions was built to be the main operating platform. The laser is held by a mechanical arm with six degrees of freedom (6DoF). Although the laser just needs movement in two dimensions, X and Y, for the current project, 6DoF allows improvement of the system in the future. The length and width of the frame are 1.5 m and 0.5 m, respectively, which is enough to cover a rail section or crossing nose. The height of the frame is 0.5 m, which allows adaption of the location of the laser and makes transition of the measurement head over different-sized objects possible. Figure 5(b) shows the completed system
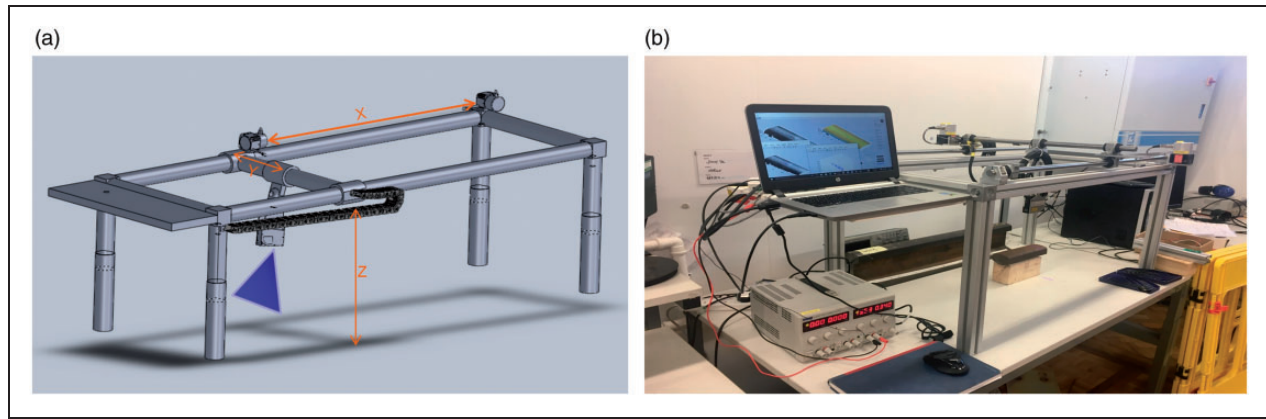
Figure 5. Mechanical frame: (a) 3D model of the frame in SolidWorks; (b) completed system in the lab.

in the lab. All equipment is powered by a DC power supply with a rated value of 24 V (last maximal 500 mA). Movement of the laser is manually controlled according to the real-time LCD display. The 3D point cloud is generated along with movement of the laser in longitudinal and lateral directions, which is then saved in.ply or.pcd format for further processing.

System resolution

Resolution is a significant property of any measurement system, which determines the quality of output data directly and influences the selection of data processing techniques. The resolution of the laser used is up to 1280 points in the Y axis, and the outputs of the displacement sensors are analogue signals. So, the resolution of the system in the X axis is limited by the 16-bit ADC converter in the sampling device (NI USB-6210). The input signal of each AI channel is an analogue voltage (0–10 V) representing movement of the laser in the X axis (0–1500 ms), and thus the resolution of the system in the X axis is

$$\frac{10\,V}{2^{16}} = 152.5\mu V \tag{1}$$

$$\frac{1500\,mm}{2^{16}} \approx 0.023\,mm \tag{2}$$

So, each 2D profile consists of 1280 points, and the minimum interval between each profile is 0.023 mm in theory and 1 mm in practice, due to the manually controlled movement of the laser and the filtering process of the point cloud (see later discussion on errors). This resolution of the integrated system is considered to be accurate enough.

Point cloud reconstruction

The point cloud derived from the 3D perceptual system discussed above includes a tremendous number of scattered points representing the surface of the objects. The next step is to reconstruct the surface with these scattered points, known as point cloud reconstruction. It is the process of applying a computer-based description of the surface of the reconstructed object, from which the geometrical characteristics of objects can be generated. The common procedures of point cloud reconstruction are as follows[16]:

- *Data registration*: Aligning several scans into a global coordinate system to generate a coherent point cloud.
- *Data integration*: Interpolating the points from the measured samples with a surface representation, which is helpful for filling holes and gaps and making the points structured in space.
- *Model conversion*: Representing the external surface of the objects using surface mesh with scattered data points.

The flowchart in Figure 6 illustrates the workflow of the 3D point cloud reconstruction for this project and is discussed with a rail sample in the following sections.

Filtering and down-sampling. There are always some noisy points caused by uncontrollable factors during data acquisition, such as surface roughness and mechanical vibrations, which will influence the accuracy of the model and increase the computational time of the system. To remove these noisy and redundant points, a series of filters are applied.

Figure 7 shows a point cloud sampled from a rail sample. Areas where data are not returned are indicated as 0 mm and can be omitted. According to the transverse profile of the rail, a filter based on the depth (Z) of each reflected point is applied. The points with depths that exceed the specified threshold will be trimmed from the point cloud data.

The filtered point cloud (Figure 8(a)) shows the shape of the rail sample clearly, but these 2D profiles are still distributed non-uniformly in the longitudinal direction, with an excessive point density in some regions due to the manually controlled movement of
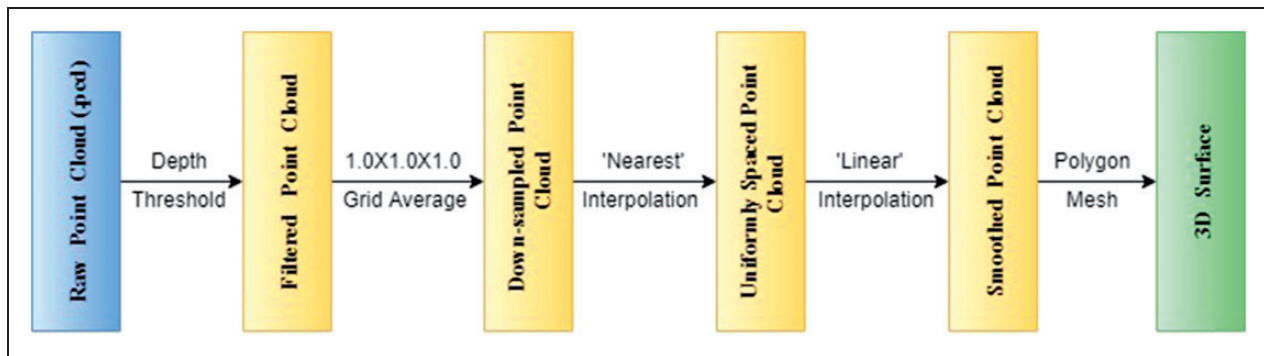
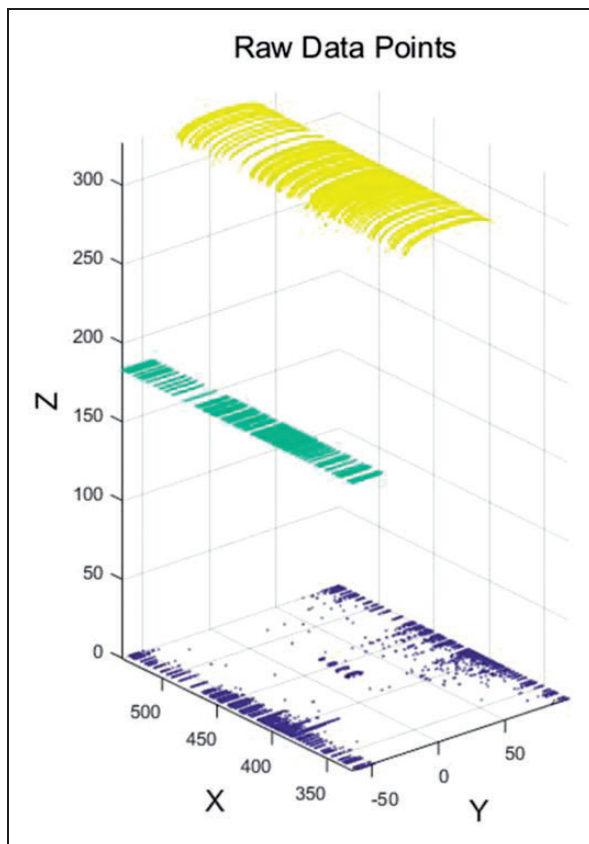**Figure 6.** Workflow of 3D point cloud processing.

Figure 7. Raw data points.

the laser. These excessive points will increase the computational time in further processing and have no positive effect on the accuracy of the model. To remove these points, the point cloud data need to be down-sampled. The method utilised in this project is grid average down-sampling, which computes the axis-aligned bounding box for the entire point cloud.[17] The bounding box is divided into grid boxes of a size specified ($1.0 \times 1.0 \times 1.0$ mm for this project). Points within each grid box are merged to a single point, by averaging their locations and normals. The graphs in Figure 8 show the filtered point cloud and the down-sampled point cloud. With this sample, the raw point cloud contains 256,000 points, and the

down-sampled point cloud retains the most representative 8604 points.

Interpolation. In numerical analysis, interpolation refers to computing new data points within the range of a discrete set of known data points. This concept is applied in the area of 3D reconstruction to generate an interpolating surface that encloses the raw data points in space. This surface is helpful for computing a set of interpolated points which retain the geometrical features of raw data points and are also structured in space.

There are various approaches to generate an interpolating surface. One common approach, applied in this project, is Delaunay triangulation.

Figure 9 illustrates the principles of Delaunay triangulation-based interpolation. The interpolating surface is represented in the form of an interpolant F with the following syntax

$$Z = F(X, Y) \tag{3}$$

where vectors X and Y specify the (X, Y) coordinates of the raw data points, and vector Z contains the depth associated with the points (X, Y).

Once the interpolating surface is generated, the interpolated points are calculated (non-connected points in the X–Y plane in Figure 9). To preserve the geometrical features of raw data, the nearest-neighbour interpolation method is used. Interpolated points are distributed uniformly in the projection area (connected points in the X–Y plane in Figure 9) of the raw data points with a specified density, based on the geometrical complexity of the model. For example, a planar surface needs few points to be defined. However, more points are required for reliable surface reconstruction of rough features.[18] Next, the depth Zq of the interpolated points is derived from the nearest raw data points. Figure 10(a) is the interpolation result of the down-sampled point cloud in Figure 8(b). It shows that the interpolated points are distributed uniformly with a controllable density (220 (X) × 70 (Y) in this example). Figure 10(c) shows the merged result of the raw data and interpolated points. The partial enlarged view in
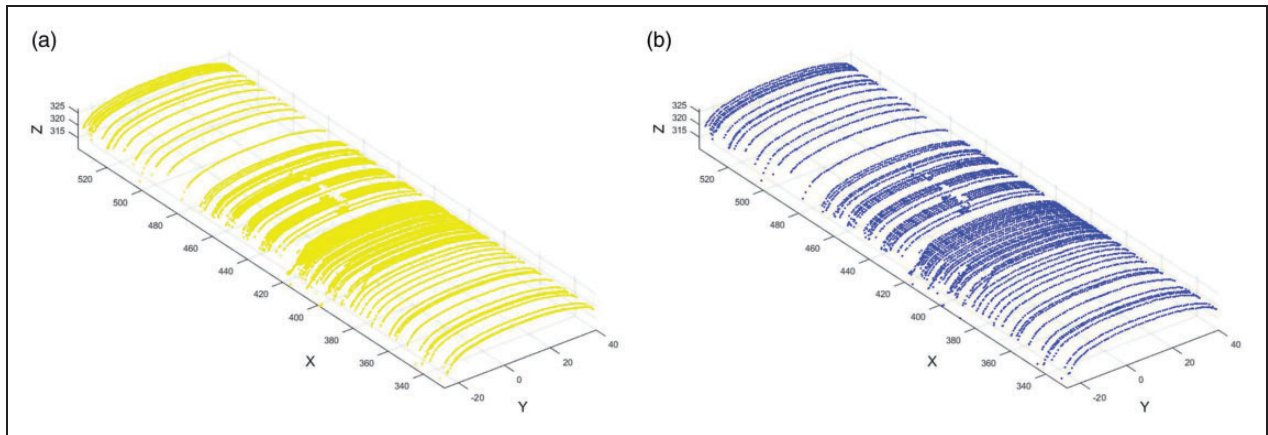
Figure 8. (a) Filtered point cloud; (b) down-sampled point cloud.
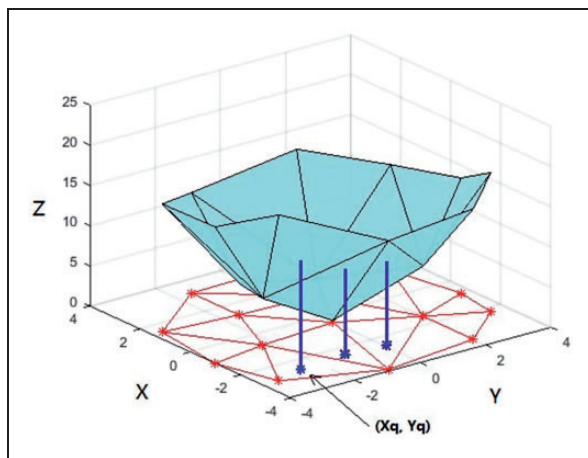
Figure 9. The principles of Delaunay triangulation.

Figure 10(d) illustrates that the interpolated points preserve the geometrical features of the raw data.

Surface polygon mesh. After a coherent point cloud is generated, the subsequent step is known as surface polygon mesh. It is known that three points (not distributed in a line) identify a planar face in 3D space. The polygon mesh is a collection of convex polygons with which the scattered points in a point cloud are connected to build the external surface of the reconstructed object.[19] The polygon mesh gives the data geometrical properties, e.g. surface angles and surface normal vectors, which is indispensable for further processing and analysis such as geometry description and defect detection.

As usual, the polygons used for meshing are triangles and quadrilaterals, which are each known as a 'face' in a 3D model. Different polygons are used depending on the geometrical complexity of the reconstructed object and the distribution of points in the point cloud. In this project, the interpolated points are distributed uniformly in the X–Y plane, so quadrilaterals were chosen to construct the surface. MATLAB graphics define the 'surf' function to build

the surface by connecting neighbouring matrix elements to form a mesh of quadrilaterals.[20] Figure 11 shows the reconstructed surface of the railway track model from the point cloud in Figure 10. The partially enlarged view in Figure 11 illustrates that the surface is built with thousands of quadrilaterals, and the geometrical properties of the surface defects are captured, which is essential for surface defect detection in the following section.

Surface defect detection

To explore the feasibility of applying 3D techniques to rail inspection, the key point is to replace conventional 2D-based defect detection with 3D, namely surface defect detection. It is a process of calculating and analysing the geometrical characteristics of the 3D model, and finding abnormal data caused by the flaws.[21] Before exploring suitable approaches for rail surface defect detection, the system was tested with different rail samples suffering from common defects. From a defect detection perspective, defects can be classified into three groups based on their geometrical characteristics, namely surface cracks, partially concave hulls with vertices lower than the original surface, and partially convex hulls with vertices higher than the original surface. Three defect-detection algorithms have been developed for these, and are described below.

Depth gradient-based detection algorithm

Surface cracks, such as gauge corner cracks, usually have an interspace from 0.5 to 10 mm. The long-term friction between wheel and rail surface can lead to surface cracks. Most occur at an angle of 15–30° to the traffic direction.

Figure 12(a) shows the laser scanning result of a rail sample with a surface crack, while Figure 12(b) illustrates the actual cross-section of the sample. It can be seen that the laser cannot capture the data behind the steep edge (circled and marked in red)
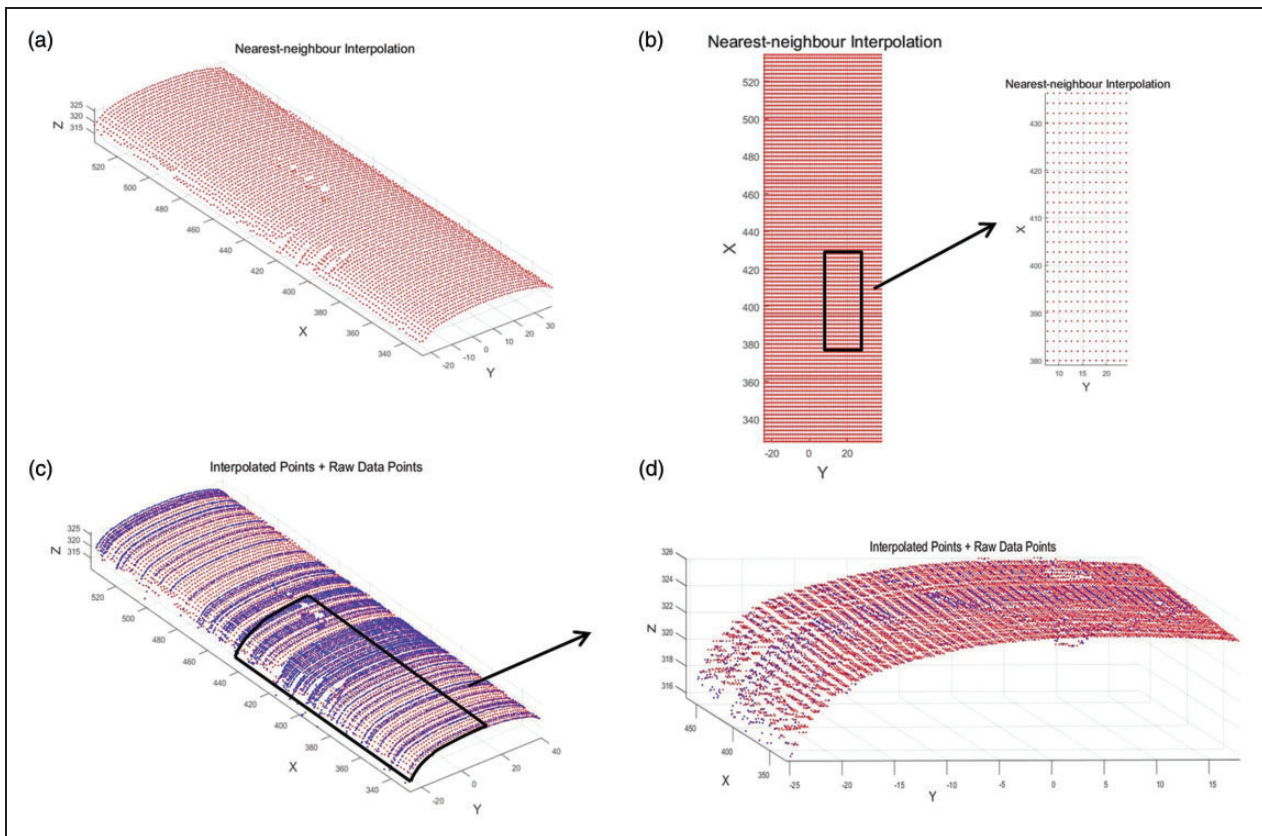
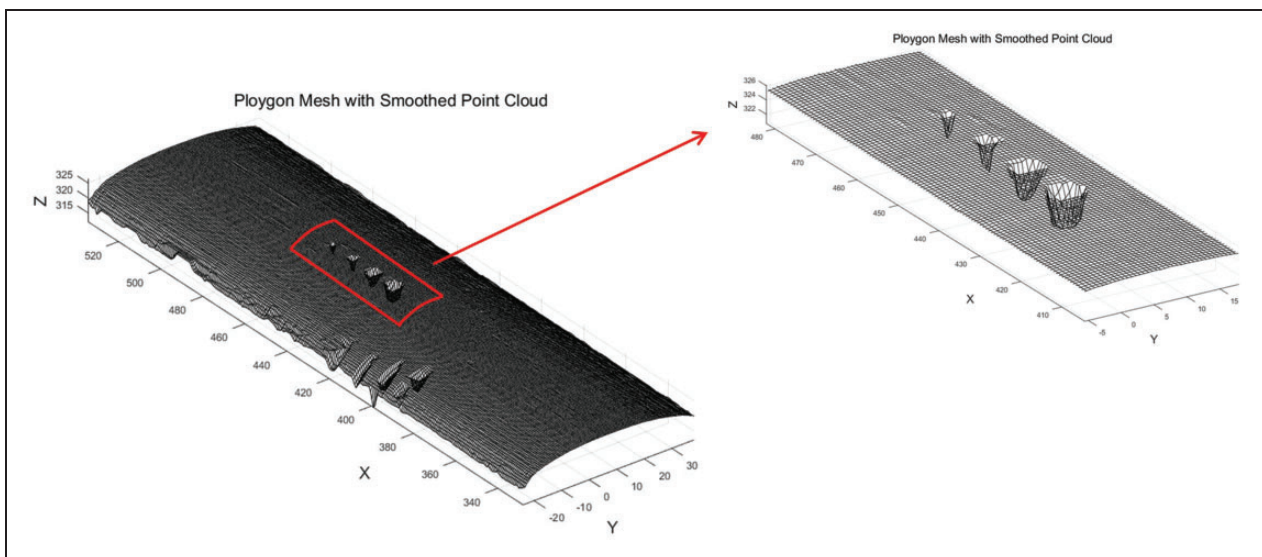**Figure 10.** Nearest-neighbour interpolation.

Figure 11. Polygon mesh of smoothed point cloud.

due to shadowing effects of the laser, which means that the laser line will disappear completely or partially behind the steep edge. Accordingly, the inner structure cannot be built accurately.

This type of defect can only be detected by recognising the points out of plane caused by surface

cracks, more specifically by detecting surface points with sharply changing depth.

To detect surface points with sharply changing depth, the concept of gradient is used. In mathematics, the gradient of a scalar field refers to the direction of a point in which the scalar field has the greatest rate

of increase.[22] The gradient of a scalar field is a vector field that can be represented as

$$\left|\nabla f\right| = \max_{|v|=1}[\nabla_v f] \tag{4}$$

where f represents a scalar field, $|\nabla f|$ denotes the absolute value of the gradient of the function, and $\nabla_v f$ is the directional derivative of the function.

In Figure 10(b), the rail-head surface is built with interpolated points distributed uniformly in the X–Y plane. If we define the function $H(x, y)$ representing the depth (z) of the point (x, y), the depth gradient of this scalar field is as follows:

$$\left|\nabla H(x, y)\right| = \max_{|v|=1}[H(x, y)] \tag{5}$$

Taking the simulated point cloud in Figure 13 as an example, the rectangular window (arrowed) is set to traverse the point cloud every n neighbouring points at a time, in which n must be a multiple of 4 to make sure that at least one mesh element is covered by the window. Data points in each window can be regarded as a subset of the whole scalar field.

Equation (5) can be simplified as

$$\left|\nabla H(x_i, y_i)\right| = \max(H(x_i, y_i)) - \min(H(x_i, y_i)) \tag{6}$$

where $H(x_i, y_i)$ represents the depth of different points in each subset, and max and min are the value of maximum depth and minimum depth, respectively.

Depending on the geometrical complexity of the 3D model and the estimated depth of the defects, the size of the rectangular window and threshold of depth gradient can be configured. If the depth
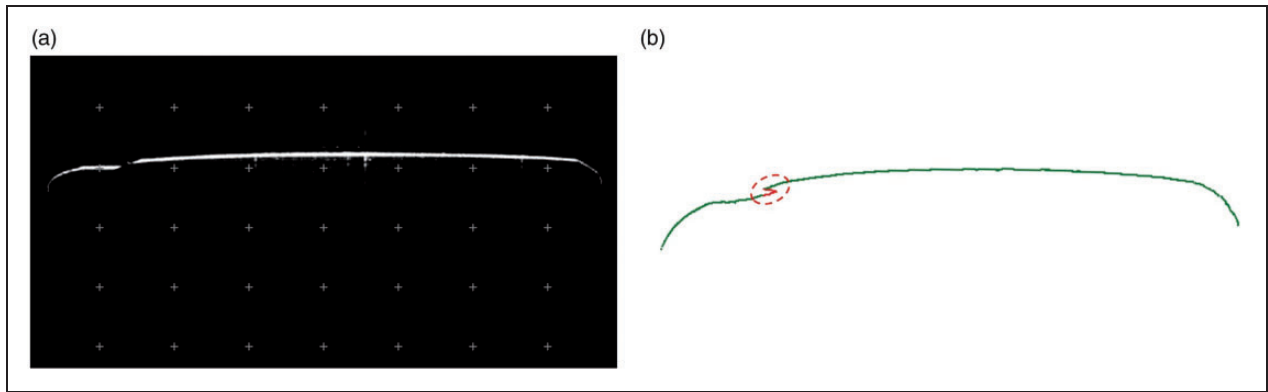

Figure 12. Cross-section of rail track: (a) laser scanning result of a rail sample suffering from a surface crack; (b) actual cross-section of the rail sample.
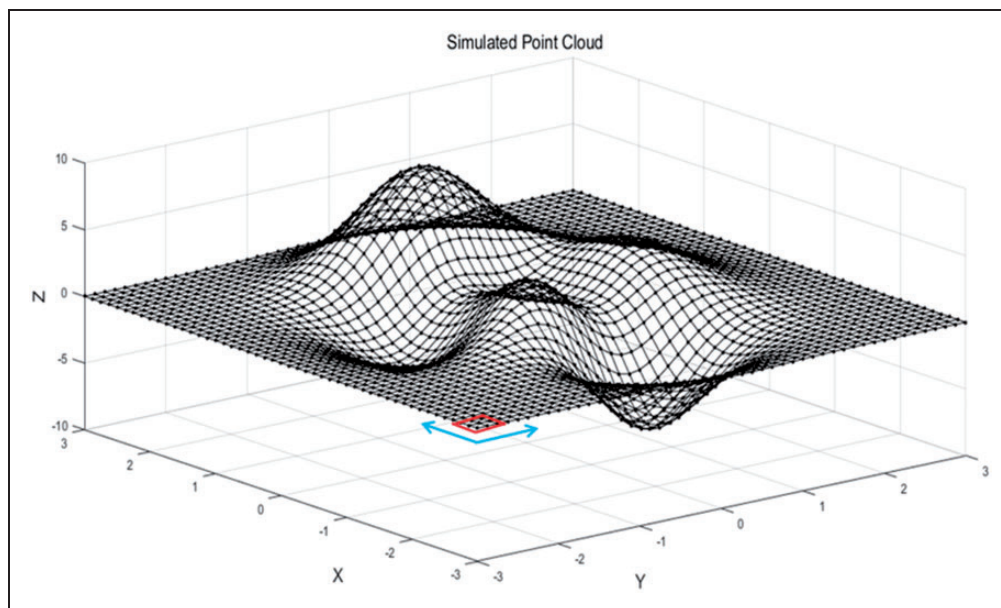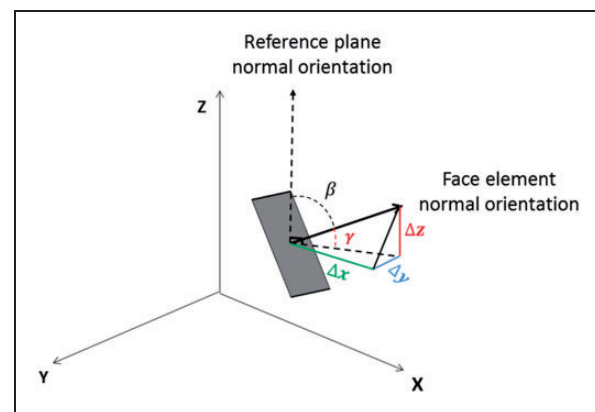

Figure 13. Traversal of the point cloud for depth gradient-based defect detection.
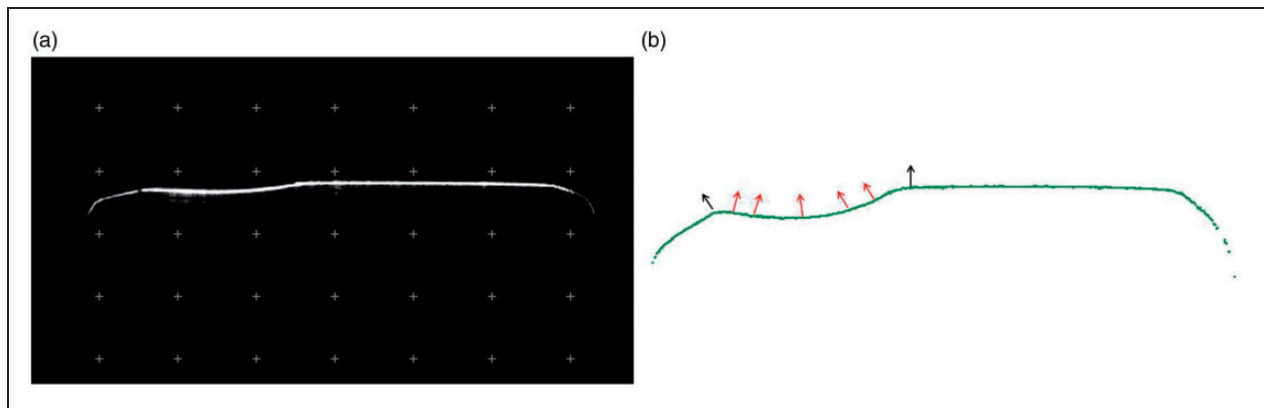
Figure 14. Cross-section of rail track: (a) laser scanning result of a rail sample suffering from squats; (b) actual cross-section of the rail sample.
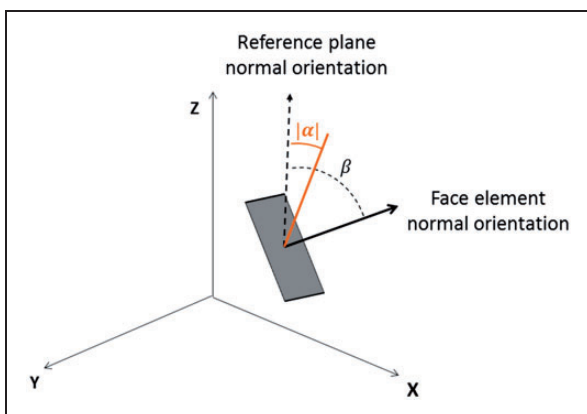
Figure 15. The principles of face normal-based detection.

Figure 16. Calculation of the reference angle of surface normal.

gradient is larger than the threshold, the corresponding area in the 3D model will be recognised as a defect (surface crack). In theory, the detectable depth variation for each mesh element is from 0 mm (when the mesh element is horizontal) to the side-length of each mesh element (when the mesh element is vertical). Because this approach is derived from the definition of the gradient, it is called a depth gradient-based detection (DGD) algorithm.

Face normal-based detection

Partially concave hulls are defects that grow into the rail surface such as shells, squats and broken crossing noses. Figure 14(a) shows the laser scanning result of a rail sample with such a defect. Figure 14(b) illustrates the cross-section of the rail sample. The arrows show the distribution of normal vectors of the concave hull caused by the defects. It indicates that the surface normal vectors of the mesh elements in this area are not in alignment with those on the original rail surface.

Through the geometrical analysis above, an algorithm was used to enable detection and

characterisation of the defects, called face normal-based detection (FND) as the detection is based on surface normal vectors. This approach was initially discussed by Torok et al.[21] By examining the surface normal vector of each mesh element, its orientation relative to a horizontal reference plane can be calculated. Mesh elements constituting the surface of the defects with surface normal vectors beyond a specified threshold can be extracted, and thus the defects can be detected. The diagram below shows the basic principles of this approach.

In Figure 15, the quadrilateral represents a mesh element. The angle β is the difference between the normal vector of a horizontal reference plane and the surface normal vector of the mesh element. The absolute value of α is set as a tolerance for slight deviation of mesh elements caused by surface roughness of the rail and system errors. In MATLAB, the function 'surfnorm' can be used for computing and displaying 3D surface normal vectors.[23] In Figure 16, assuming that γ represents the angle between the horizontal reference plane and the surface normal vector of the mesh element, it can be calculated from the difference
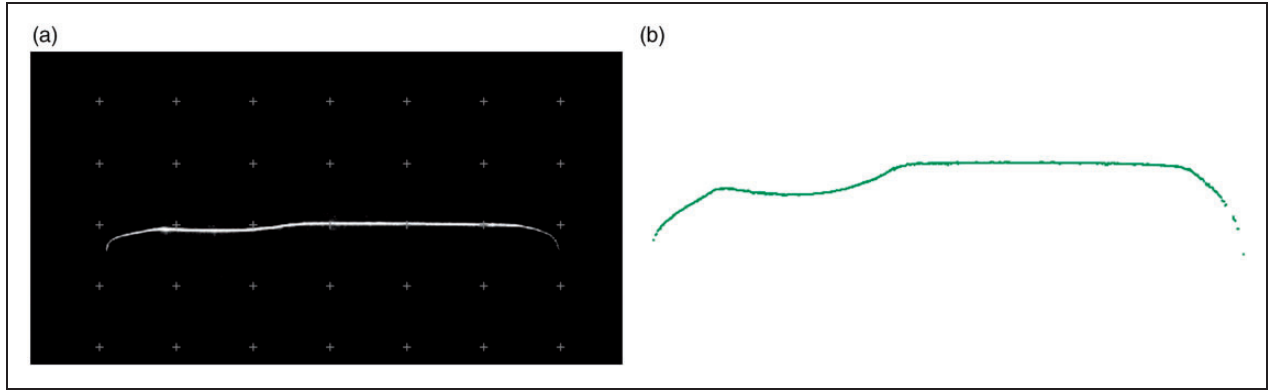
Figure 17. Cross-section of rail track: (a) laser scanning result of a rail sample suffering from squats; (b) actual cross-section of the rail sample.
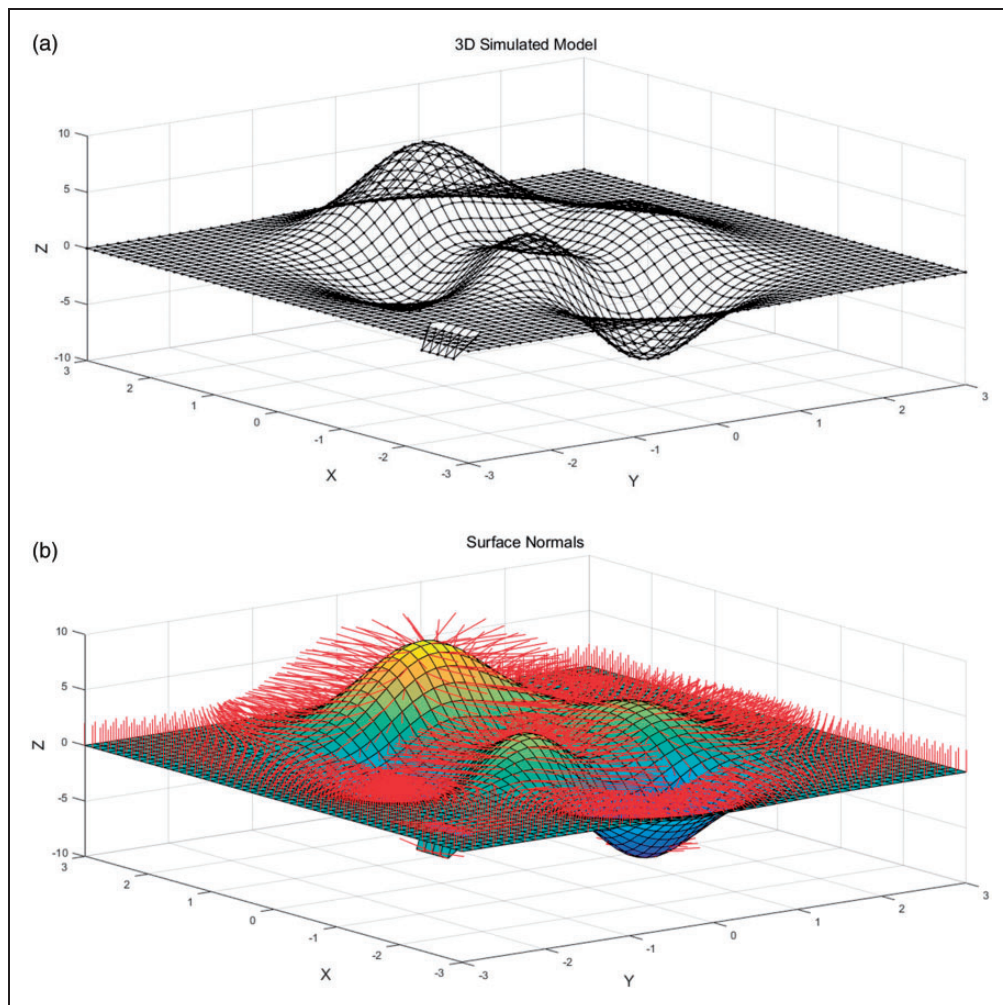
Figure 18. Simulated model: (a) surface polygon mesh; (b) surface normals.

between two vertices of the surface normal vector based on Pythagorean theorem in which the tangent of γ is

$$\tan\gamma = \frac{\Delta z}{\sqrt{(\Delta x^2 + \Delta y^2)}} \qquad (7)$$

where x, y and z are the differences between two vertices of the surface normal vector in each axis in a 3D coordinate system, and thus angle γ is

$$\gamma = \tan^{-1}\left(\frac{\Delta z}{\sqrt{(\Delta x^2 + \Delta y^2)}}\right) \qquad (8)$$

where $\tan^{-1}$ is the arc-tangent function. Accordingly, the angle β is equal to $(90° - \gamma)$. If β is greater than the tolerance angle, it indicates that the mesh element is not properly in alignment with the original rail surface, and thus is considered as part of the surface defects.[21]

Face normal gradient-based detection

A partially convex hull (Figure 17) usually occurs at the edge of a concave defect, or the long-term collision between wheel and crossing nose may cause deformation of the crossing nose with a partially convex surface. These kinds of defects are difficult to detect using FND. Because the orientations of surface normal vectors in this region are diverse, it is impossible to set a specified threshold region to cover all conditions. Meanwhile, the depths of surface points in this region change slowly in most cases, so DGD cannot work properly either.

The geometrical characteristics of these types of defect can be analysed using commonly used edge detection algorithms, such as brightness gradient-based detection in digital images which was first developed by J. Canny.[24] Similarly, a new approach that combines the principles of DGD and FND is applied because the

angles of surface normal vectors will have a sharp change at a convex surface or at the edge of concave defects with fluctuation in the surface. By comparing the normal vectors of adjacent mesh elements, the sharp change in adjacent surface normal vectors caused by these kinds of defect can be detected. When the difference reaches a certain threshold, the corresponding mesh elements will be recognised as part of a surface defect. Similar to DGD, this approach is called face normal gradient-based detection (FNGD). The face normal gradient can be calculated as follows:
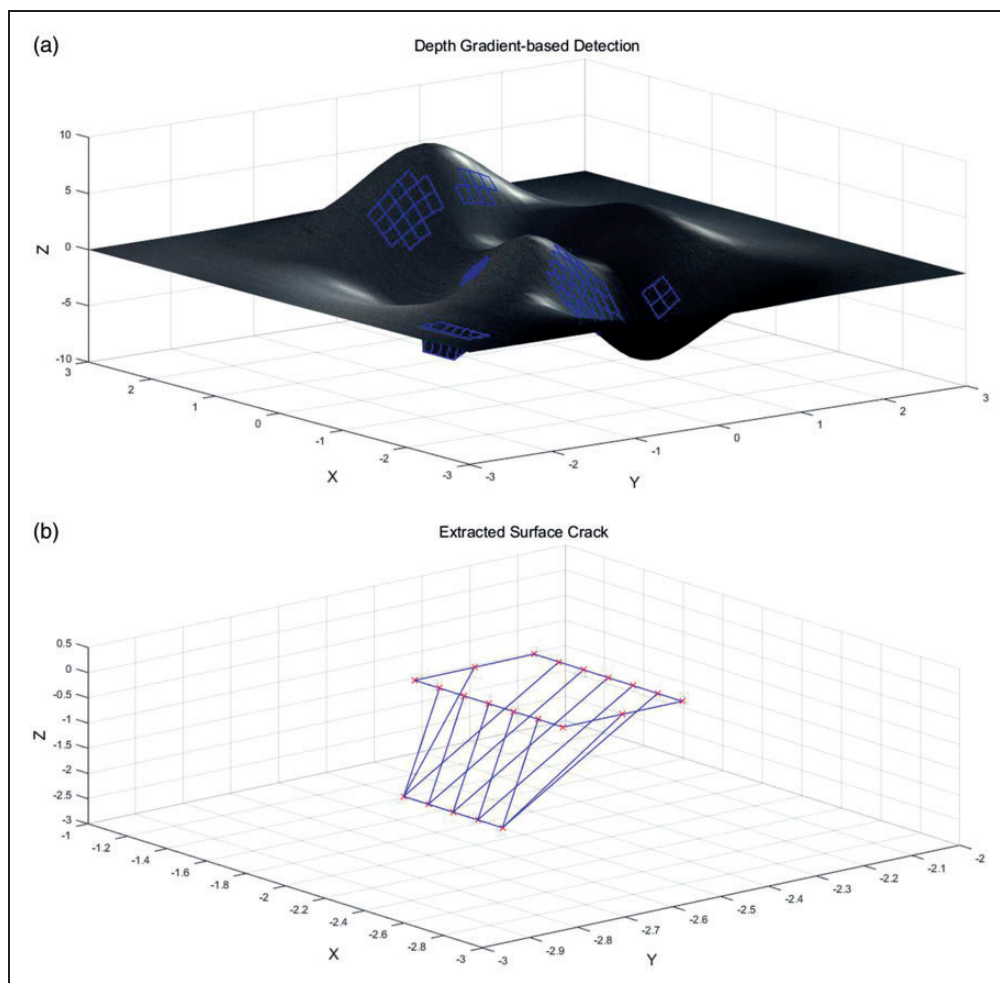
$$\left|\nabla N_j\right| = \max_{|v|=1}\left[\nabla_v N_j\right] = \max\left(N_j\right) - \min\left(N_j\right) \quad (9)$$

where N_j represents the angles of surface normal vectors of mesh elements in each subset, and max and min are the values of maximum angles and minimum depth, respectively.
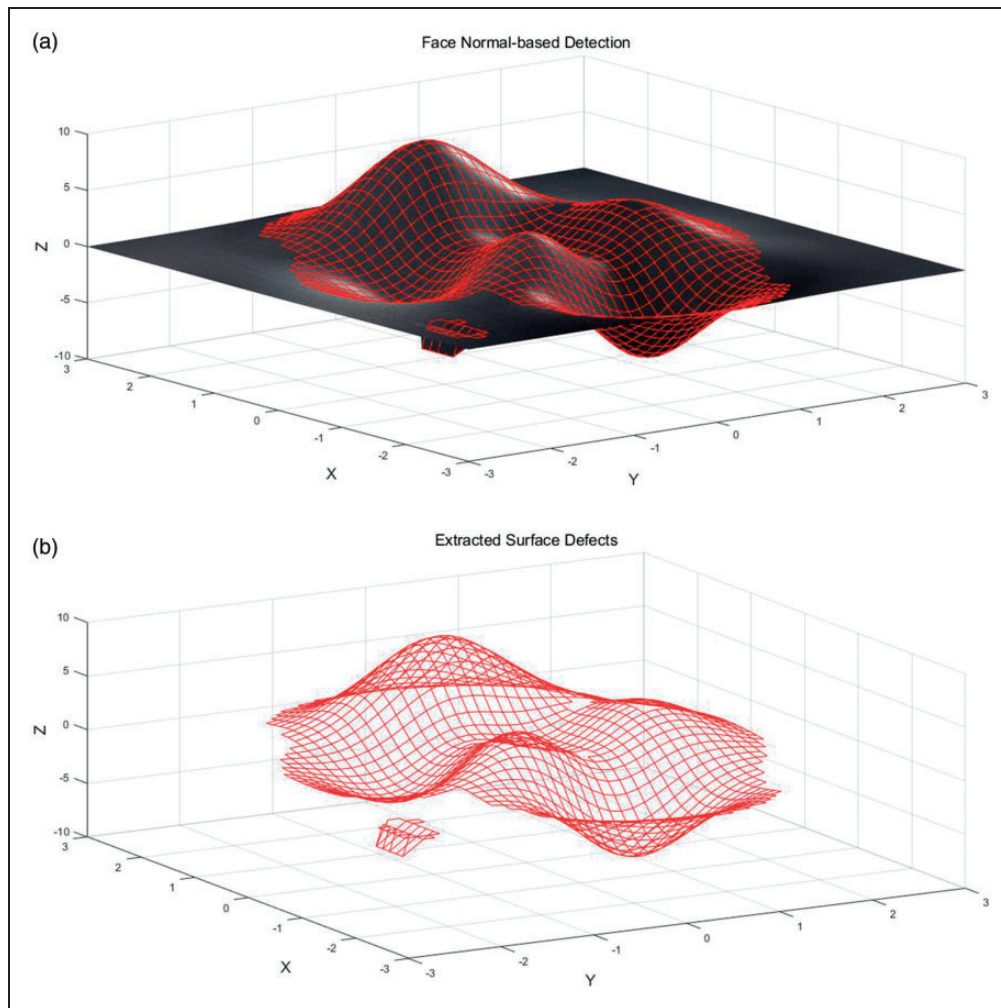
Experimental results

Simulation results

All the defect-detection algorithms were tested with a 3D surface (Figure 18) simulated in MATLAB for

Figure 19. Depth gradient-based detection: (a) results of detection; (b) extracted surface crack.

Figure 20. Face normal-based detection: (a) results of detection; (b) extracted surface defects.

feasibility analysis and optimisation before being integrated into the system. The 3D model is a plane with concave and convex hulls as well as a ditch simulating a surface crack.

Depth gradient-based detection. The first detection algorithm applied was DGD. In this case, the rectangular window was set to cover four neighbouring points. Figure 19(a) shows the results of the detection (defects indicated in blue), and (b) shows the geometrical structure of the simulated surface crack. Some concave and convex hulls experiencing a sharp change in depth are recognised; however, these can be ignored as they have no effect on the assessment of the surface crack.

Face normal-based detection. Figure 20 illustrates the results of FND. With the tolerance angle set to $30°$, the whole concave and convex hull will be covered. Figure 20(b) is the extracted frame showing the geometrical structure of the simulated defects. With this ideal model, the simulated surface crack is detectable. In practice, the inner structure of surface cracks cannot be reached by the laser line, which means

data for the corresponding area are unavailable, so the surface crack detected here should be ignored.

Face normal gradient-based detection. The third algorithm tested was FNGD. The testing results (Figure 21) demonstrate that the peak, bottom and edge of the simulated defects where the surface normal vectors have significant variation as a consequence of fluctuations in the surface can be detected when the maximum gradient is set to $12°$.
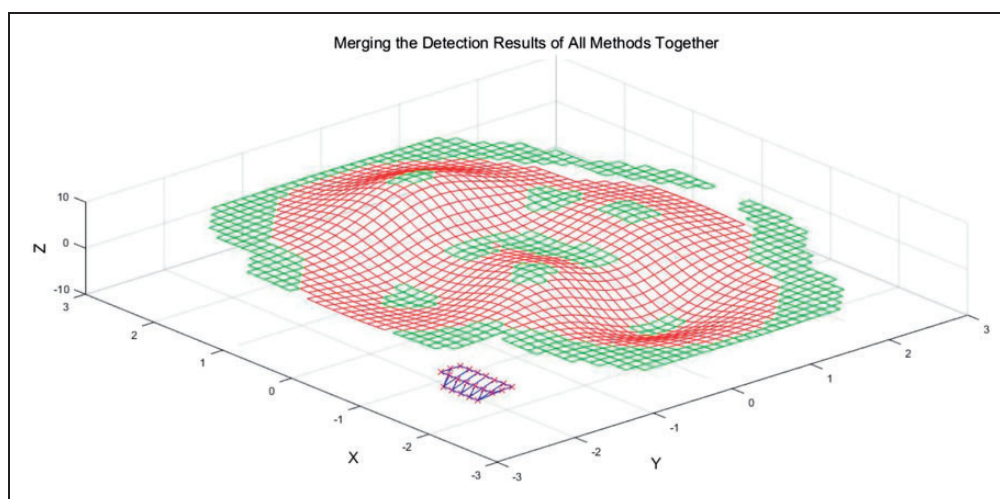
To demonstrate the feasibility of the detection algorithms further, the results of different detection algorithms were merged (Figure 22). From an approximate viewpoint, it can be seen that the areas covered by different detection algorithms are complementary to each other. The whole defect can be detected and recovered using the detection algorithms.

Laboratory tests

The experimental results were encouraging, to demonstrate the practical value of the approach presented. The system was tested with samples in both the
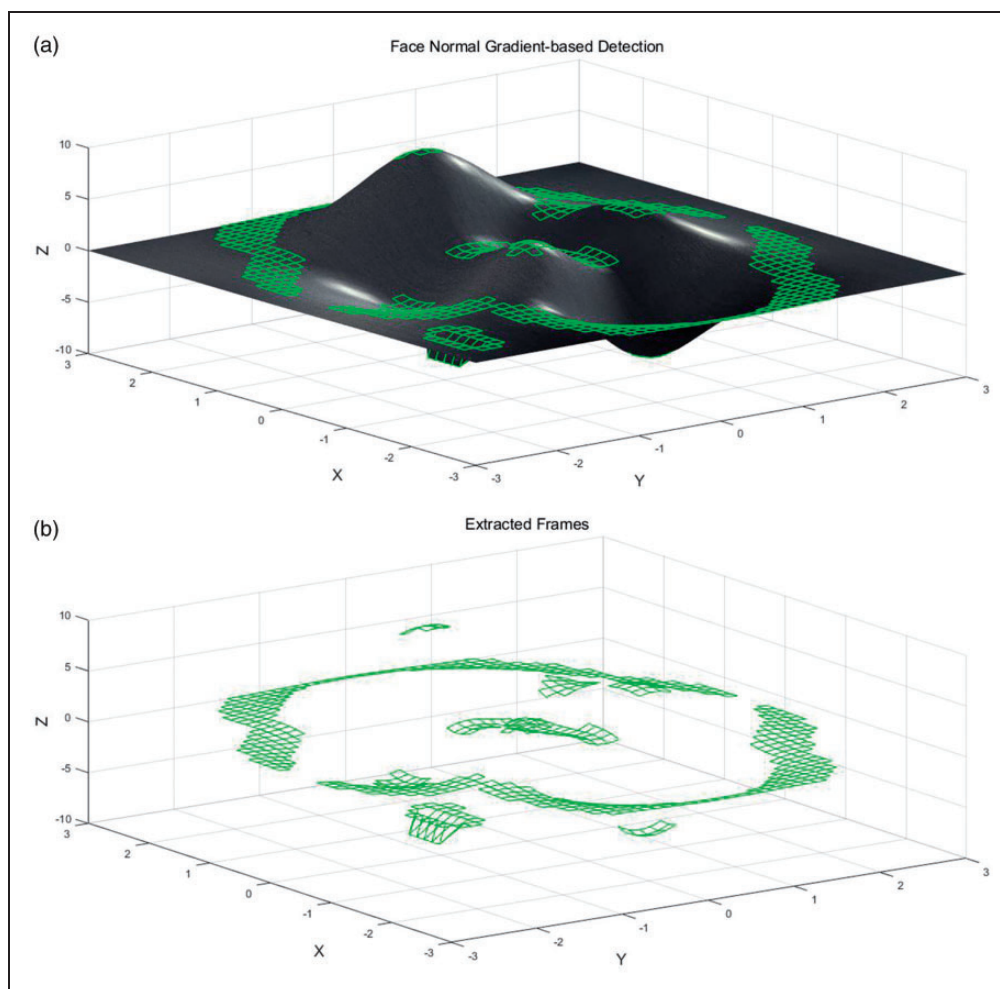
Figure 21. Face normal gradient-based detection: (a) results of detection; (b) extracted frames.

Figure 22. Merging the detection results of different algorithms.

laboratory and in the real world, which is more cogent for giving the advantages as follows:

1. To demonstrate the feasibility of performing rail measurement with the selected instruments;

2. To prove the ability to build a 3D model of the rail and characterise defects using 3D reconstruction techniques;

3. The more complicated geometrical structures of the rail defects demonstrate the

performance of the defect-detection algorithms developed.

The experimental procedures and evaluation mechanism referenced the approaches used by Rowshandel[25] and Torok et al.[21] The main strategy was to evaluate the accuracy of the 3D reconstruction and the defect-detection algorithms by deriving the relative errors from the following equation

$$\delta = \frac{Actual\ value - Measurement}{Actual\ Value} \times 100\% \quad (10)$$

Testing results. From a large number of experiments conducted, four representative cases have been chosen. Figure 23(a) shows a $118.0(L) \times 65.0(W) \times 151.0(H)$ mm rail sample in the laboratory, which has three artificially induced cracks in the gauge corner with a cutting angle of 45°, mimicking surface cracks. Figure 23(b) shows a $164.0(L) \times 60.0(W) \times 128.0(H)$ mm rail sample, which has two artificially

induced squats in the running surface with a depth of around 2 mm, mimicking surface squats. Figure 23(c) shows a section of the in-service rail at the Long Marston Railway Test Track. The width of the rail is 62 mm. The scanned section is 210 mm long and contains the joint of two plain rails, and a rail defect at the gauge corner. Figure 23(d) shows the crossing nose inspected at Long Marston. The scanned section is $705.0 (L) \times 98.5$ mm (W), and the joint between the two plain tracks and the crossing nose is 285.0 mm from the vertex of the crossing nose. The quantitative information for each target and the defects are listed in Table 1.

For visual ease and understanding of the approach, the procedures starting from surface polygon mesh to detailed surface defect detection have been summarised in Figure 24. Procedures before surface polygon mesh follow the method introduced in section 'Point cloud reconstruction'. The density of interpolation is set to make the size of each mesh element small enough to capture details of the rail surface. For example, the length and width of the rail
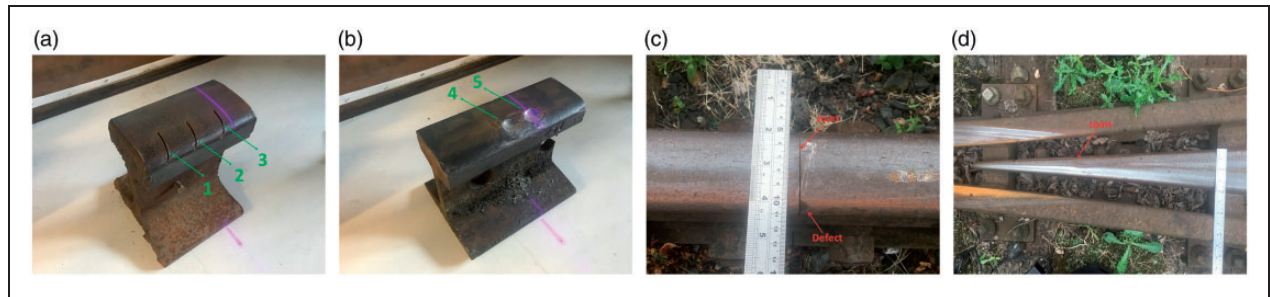
Figure 23. Targets of laboratory and field tests.

Table 1. Quantitative information of the defects.

Defect	Surface length (mm)[a]	Maximum depth (mm)[b]	Surface angle (°)[c]/ width (mm)[d]	Cross-section view (approx. geometry)
1	31.20	4.00	45 (2.00 mm)	
2	28.30	3.00	45 (2.00 mm)	
3	27.00	4.33	45 (2.00 mm)	
4	16.60	1.90	12	
5	19.50	2.42	12	
6	10.34	6.85	3.12	

[a]Surface length refers to the size of the defect in the longitudinal direction of the rail.
[b]Maximum depth is the maximal distance between the bottom of the defect and the original surface of the rail.
[c]Surface angle is the cutting edge when making the artificial defects.
[d]Surface width refers to the size of the defect in the lateral direction of the rail.
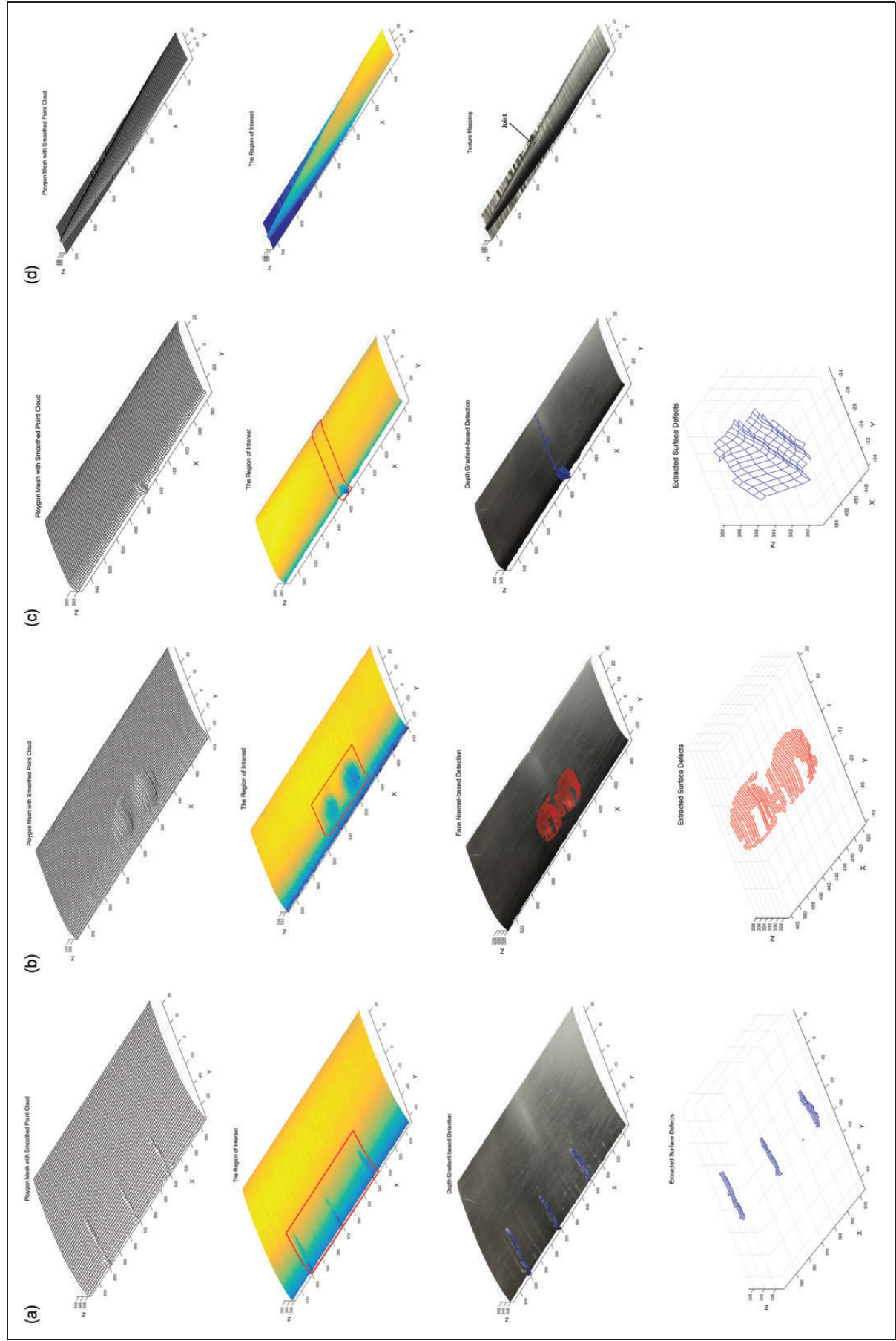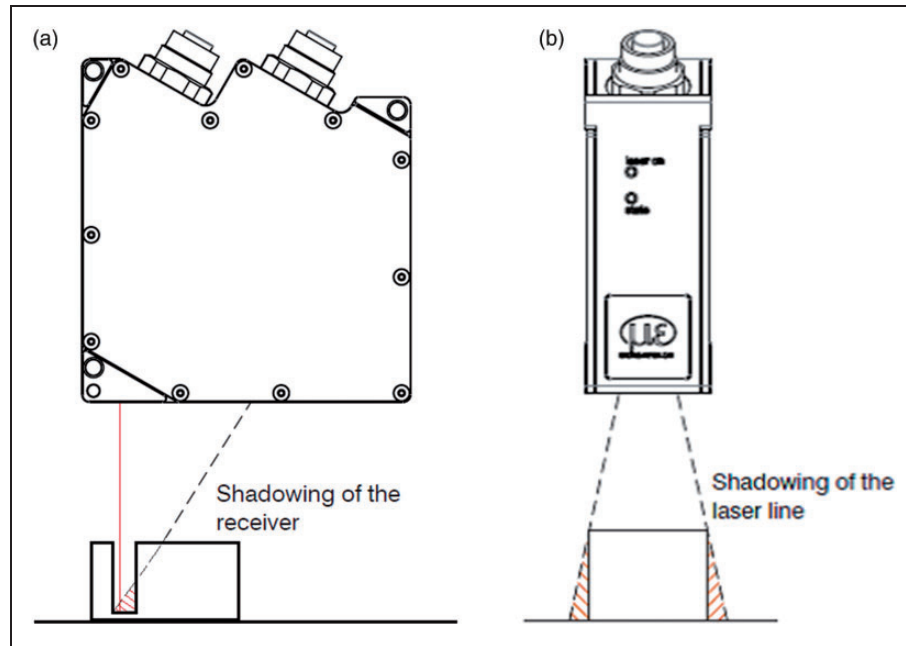
Figure 24. Results of the proposed approach for the detection and characterisation of physical surface defects in rails and crossing noses.

Figure 25. Shadow effects of the laser line.

Table 2. Comparison between the rail sample and the 3D model.

Rail sample	Length (mm)		Relative error (%)	Width (mm)		Relative error (%)
	Mea.	Actual		Mea.	Actual	
1	116.0	118.0	1.69	63.5	65.0	2.31
2	164.0	164.0	0.00	59.0	60.0	1.67
3	211.5	210.0	0.71	61.0	62.0	1.61
4	703.0	705.0	0.28	96.0	98.5	2.54

sample shown in Figure 24(a) are 118.0 mm and 65.0 mm, respectively. The density of interpolation was set to $120(X) \times 65(Y)$, so the area of each mesh element is about $1 \text{mm}^2 \left(\frac{118}{120-1} \times \frac{65}{65-1} \right)$ which is small enough to capture details of the rail surface.

The first row is the 3D surface polygon meshes of the four rail samples in Figure 24. It can be seen that all the geometrical details can be captured clearly. The graphs in the second row are the depth-coloured 3D models in which the colour of each mesh element is determined by the depth of the corresponding vertices. As discussed in section 'Surface defect detection', surface defects lead to the surrounding points being out of plane, which in turn are displayed as colour changes. Accordingly, the region of interest (ROI) can be extracted (in the rectangular box). Applying the defect-detection algorithms to the ROI rather than the entire data set can reduce the computational load, improving the efficiency of detection.[26]

The third row shows the results of surface defect detection. Because laser-based measurement does not give colour information for the rail surface, to make the texture of the 3D models more clear, digital photographs containing different textures of metal

were mapped using a method called texture mapping, more details of which can be found in the thesis by Paulo Dias.[27] The fourth row is the extracted surface defects, from which the geometrical characteristics of the defects can be evaluated.

Evaluation. To evaluate the accuracy of the 3D reconstruction and defect-detection algorithms, the relative errors are summarised in Tables 2 and 3. It can be seen that the relative errors of 3D reconstruction are very minor. As discussed in section 'System resolution', the resolution of the system in the X axis is 0.023 mm in theory, so the relative errors between the 3D model and the rail sample are mainly caused by the minimal human intervention in moving the laser, which is reasonable and could be reduced by improving the measuring mechanism of the system in future work.

The size of the extracted defects and the actual size of the corresponding defects are listed in Table 3. Most of the defects can be detected and characterised with a relative error lower than 10%, which demonstrates the accuracy and feasibility of the defect-detection algorithms developed. The relative errors greater than 10% are from measurement of the three artificial

Table 3. Comparison between measurements and the actual size of the defects.

Defect	Length (mm) Mea.	Actual	Relative error (%)	Width (mm) Mea.	Actual	Relative error (%)	Depth (mm) Mea.	Actual	Relative error (%)
1	25.97	31.20	16.76	2.30	2.00	15.00	3.30	4.00	17.50
2	21.07	28.30	25.55	2.30	2.00	15.00	2.10	3.00	30.00
3	21.61	27.00	19.96	2.20	2.00	10.00	2.90	4.33	33.03
4	17.00	16.60	2.41	–	–	–	1.80	1.90	5.26
5	19.10	19.50	2.05	–	–	–	2.40	2.42	0.83
6	9.50	10.34	8.12	6.17	6.85	9.93	2.80	3.12	10.26

cracks. According to the analysis in section 'Depth gradient-based detection algorithm', measurement is influenced by the shadowing effects of the triangulation laser. The laser line will disappear partially or completely behind the steep edge. As shown in Figure 25,[28] the laser cannot 'see' the red areas. That is the main reason why the relative errors increase when performing surface crack measurement.

There are also some other factors that may influence the accuracy of surface defect detection, such as surface roughness, mechanical vibrations and approximation-based interpolation. Most of them correspond to systematic errors that are caused by the instruments involved, or random errors which are the inherent property of the measurement process. Although the approximate surface for interpolation may cause errors, these can be compensated for by the advantages of interpolation, such as filling gaps and distributing the points uniformly. So, it can be concluded that the approach itself is reliable, and the errors could be suppressed by improving the instruments.

Conclusions

This paper has presented an innovative prototype and processing to explore the feasibility of applying 3D techniques in railway track and crossing-nose inspection to improve the capability of existing inspection methods which are usually implemented on a 2D basis. To ensure the research supports the requirements of the railway industry, the current state of rail and crossing-nose inspection has also been reviewed.

The methodology has been discussed in three phases: development of a low-cost 3D perceptual system, generation of 3D models, and development of defect-detection algorithms to be applied to these models. A 3D perceptual system has been developed to achieve 3D measurement at a cost much lower than that of commercial 3D scanners. The integrated system manages the interaction between different devices and synchronisation of the different types of signal. All the devices are mounted on a mechanical frame which is easily transportable and ensures stable operation of the device.

For data processing, a series of image processing techniques have been applied to generate a 3D model of the rail or crossing-nose surface. Three defect-detection algorithms have been developed for use specifically with the model, rather than just the 2D laser outputs. These techniques allow common surface defects with different geometrical characteristics to be detected and characterised more comprehensively than algorithms working in the 2D space alone. The results of laboratory and field tests demonstrate that applying 3D techniques and using 3D model-based analysis can identify a range of realistic defects, both transversal and longitudinal, and hence that 3D model based rail and crossing-nose inspection is feasible.

ORCID iD

Jiaqi Ye ⓘD http://orcid.org/0000-0002-9593-8995

References

1. Lorente A, Llorca D, Velasco M, et al. Detection of range-based rail gauge and missing rail fasteners. *Transp Res Record* 2014; 2448: 125–132.
2. Rockman S. Network Rail's new measurement train. *The Register*, 10 November 2014.
3. Papaelias MP, Roberts C and Davis CL. A review on non-destructive evaluation of rails: State-of-the-art and future development. *Proc IMechE, Part F: J Rail and Rapid Transit* 2008; 222: 367–384.

4. Ni H, Lin X, Ning X, et al. Edge detection and feature line tracing in 3D-point clouds by analyzing geometric properties of neighborhoods. *Rem Sens* 2016; 8: 710.

5. Cannon DF, Edel K-O, Grassie S, et al. Rail defects: An overview. *Fatigue Fract Eng Mater Struct* 2003; 26: 865–886.

6. Grassie SL. Rolling contact fatigue on the British railway system: Treatment. *Wear* 2005; 258: 1310–1318.

7. MainLine. *Development of new technologies for S&C replacement*. Deliverable. Report no. SST.2011.5.2-6, 2014.

8. Hassankiadeh SJ. *Failure analysis of railway switches and crossings for the purpose of preventive maintenance*. Master Degree Project. Report no. ISSN 1650-867X, 2011. Royal Institute of Technology.

9. Franken PA and Sands RH. The Ann Arbor conference on optical pumping: The University of Michigan, 15–18 June 1959, http://catalog.hathitrust.org/api/volumes/oclc/2460155.html (accessed 30 October 2015).

10. Townes CH. *How the laser happened: Adventures of a scientist*. Oxford, UK: Oxford University Press, 1999.

11. LMI Technologies. *A simple guide to understanding 3D scanning technologies*. 1st ed. Burnaby, Canada: LMI Technologies, 2017.

12. Zhou Y, Zhang J, Jiang J, et al. Study and implementation on rail profile alignment algorithm based on 3D scanning data. In: *CICTP*, Beijing, China, 25-27 July 2015, pp.1942–1951. USA: American Society of Civil Engineering.

13. Dias P, Matos M and Santos V. *3D reconstruction of real world scenes using a low-cost 3D range scanner*. USA: Computer-Aided civil and infrastructure engineering, 2006, pp.486–497.

14. Reyes AL, Cervantes JM and Gutiérrez NC. Low cost 3D scanner by means of a 1D optical distance sensor. *Proc Technol* 2013; 7: 223–230.

15. MICRO-EPSILON. *scanCONTROL 2D/3D laser scanner (laser profile sensors)*. U.S.: MICRO-EPSILON, 2015.

16. Bernardini F, Mittleman J, Rushmeier H, et al. The ball-pivoting algorithm for surface reconstruction. *IEEE Trans Vis Comput Graph* 1999; 5: 349–359.

17. MathWorks. pcdownsample, https://uk.mathworks.com (2015, accessed 5 September 2016).

18. Lin Y-J, Benziger RR and Habib A. Planar-based adaptive down-sampling of point clouds. *Photogramm Eng Rem S* 2016; 82: 955–966.

19. Marton ZC, Rusu RB and Beetz M. On fast surface reconstruction methods for large and noisy point clouds. In: *2009 IEEE international conference on robotics and automation*, 12–17 May 2009, pp.3218–3223.

20. MathWorks. surf, https://uk.mathworks.com/help/matlab/ref/surf.html (2016, accessed 8 May 2017).

21. Torok MM, Fard MG and Kochersberger KB. Post-disaster robotic building assessment: Automated 3D crack detection from image-based reconstructions. *Comput Civil Eng* 2012; 2012: 397–404. DOI: 10.1061/9780784412343.0050.

22. Dubrovin BA, Fomenko AT and Novikov SP. *Modern geometry—Methods and applications: Part I: The geometry of surfaces, transformation groups, and fields*. Moscow, Russia: Springer, 1991.

23. MathWorks. surfnorm, https://uk.mathworks.com/help/matlab/ref/surfnorm.html (2017, accessed 24 May 2017).

24. Canny J. A computational approach to edge detection. *IEEE Trans Pattern Anal Mach Intell* 1986; 8: 679–698.

25. Rowshandel H. *The development of an autonomous robotic inspection system to detect and characterise rolling contact fatigue cracks in railway track*. UK: University of Birmingham, 2014.

26. Brinkmann RON. Chapter five – Basic image compositing. *The art and science of digital compositing*. 2nd ed. Boston: Morgan Kaufmann, 2008, pp.149–188.

27. Dias P. *Three dimensional reconstruction of real world scenes using laser and intensity data*. Portugal: University of Aveiro, 2003.

28. MICRO-EPSILON. *Instruction manual scanCONTROL 29xx/BL*. Germany: MICRO-EPSILON, 2008.